Developing Pedagogical Content Knowledge For Stem Integration Through Data Literacy: A Case Study Of High School Science Teachers

Katherine M. Miller
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/edissertations

Part of the Science and Mathematics Education Commons, and the Teacher Education and Professional Development Commons

Recommended Citation
https://repository.upenn.edu/edissertations/4956

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/4956
For more information, please contact repository@pobox.upenn.edu.
Developing Pedagogical Content Knowledge For Stem Integration Through Data Literacy: A Case Study Of High School Science Teachers

Abstract
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. Data literacy is inherently STEM integration, especially when working with real-world scenarios, as it requires bringing math and technology into the science classroom to interpret and engage with authentic data. Since teachers are often trained in one specific subject, they need additional support to accomplish STEM integration. In addition to subject matter knowledge of data literacy, this support must focus on the pedagogical content knowledge (PCK) required to implement and support student learning. This study sought to add an extra treatment to an existing STEM integration PD which focused specifically on teachers’ PCK for data literacy. There is a major dearth of research on PCK for data literacy. Therefore, this study built on the limited research done on PCK for STEM integration and for statistics education to begin to develop an understanding of PCK for data literacy, an under-explored concept. This was an exploratory research multiple case study of four secondary school science teachers in an urban school district, who participated in the existing bioinformatics summer PD, then engaged in approximately 20 hours of workshop sessions focused on PCK development throughout the school year as they taught the bioinformatics unit. The initial findings show that there are unique components of PCK for data literacy that are different from those for science education. The participating teachers were able to surface their knowledge of student understanding of data and strategies for teaching with and about data to define a number of components of PCK that begin to build a framework for what PCK for data literacy might look like in a generalized form that could be used to support additional teachers seeking to growth their ability to teach with complex data. Additionally, strategies used in the extension PD, such as CoRes and video reflections were shown to offer support for teachers in their implementation of authentic data literacy in their classrooms.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Education

First Advisor
Susan Yoon

Keywords
Data Literacy, Pedagogical Content Knowledge, Professional Development, STEM Integration

Subject Categories
Science and Mathematics Education | Teacher Education and Professional Development

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/4956
DEVELOPING PEDAGOGICAL CONTENT KNOWLEDGE FOR STEM INTEGRATION THROUGH DATA LITERACY: A CASE STUDY OF HIGH SCHOOL SCIENCE TEACHERS

Katherine M. Miller

A DISSERTATION

in

Education

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

2022

Supervisor of Dissertation:
Susan A. Yoon, Professor of Education

Graduate Group Chairperson:
J. Matthew Hartley, Professor of Education

Dissertation Committee:
Susan A. Yoon, Professor of Education (Chair)
Janine Remillard, Professor of Education
Joseph Polman, Associate Dean for Research, University of Colorado Boulder
To the many teachers throughout my life, both formal and informal, who believed in me, inspired me, and helped shape me into the person I am today.
ACKNOWLEDGEMENTS

First and foremost, I would like to thank Dr. Susan Yoon for being an incredible guide, mentor, thought-partner, teacher, and supporter throughout this process. I have learned so much from her over the years and her continued faith in me was a flame on who’s light I could always rely. Additionally, I am grateful to Dr. Janine Remillard and Dr. Joe Polman for giving me their time and advice. Dr. Remillard stoked my interest in working with and learning from teachers which led me to this research and Dr. Polman’s insightful questions made this research better.

Thank you to my incredible group of teacher participants. I am so lucky that you agreed to muddle through this fledgling idea with me and turn it into something worthwhile. I learned so much from each of you. Keep making the world better.

I was so incredibly lucky to start this journey with an amazing group of strong, brilliant, inspirational women who have been a source of comfort, energy, and insight as we navigated this process together. To Chenelle, Gayithri, Mia, Sarah, Shamya, and Taylor: thank you for going on this crazy ride with me.

I realize I must also thank Dr. Yoon for building an incredible team of students and researchers with whom I was able to learn, stumble, and get back up again. To Jooeun, Thomas, and Noora: thank you for being my work family, for holding my hand, for helping to carry the load, and for giving never ending feedback and advice.

Eternal love to my sisters, their partners, my perfect nephews, and my parents for being the best family a girl could ask for.

To all of my lifelong friends spread out across the world, Masha, Cari, Posen, Claire, Naina, Jackie, Whitney, Alex, Will, thank you for still being friends with me despite my disappearance from social life to work on this research and for always being there when I needed you. And especially to Scott for being my rock and my better half. I love you 3000.

And finally, I send my gratitude to a small town in Vermont and especially to Mrs. Harriet Worrell, who instilled in me an insatiable sense of wonder.
ABSTRACT

DEVELOPING PEDAGOGICAL CONTENT KNOWLEDGE FOR STEM INTEGRATION THROUGH DATA LITERACY: A CASE STUDY OF HIGH SCHOOL SCIENCE TEACHERS

Katherine M. Miller
Susan A. Yoon

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. Data literacy is inherently STEM integration, especially when working with real-world scenarios, as it requires bringing math and technology into the science classroom to interpret and engage with authentic data. Since teachers are often trained in one specific subject, they need additional support to accomplish STEM integration. In addition to subject matter knowledge of data literacy, this support must focus on the pedagogical content knowledge (PCK) required to implement and support student learning. This study sought to add an extra treatment to an existing STEM integration PD which focused specifically on teachers’ PCK for data literacy. There is a major dearth of research on PCK for data literacy. Therefore, this study built on the limited research done on PCK for STEM integration and for statistics education to begin to develop an understanding of PCK for data literacy, an under-explored concept. This was an exploratory research multiple case study of four secondary school science teachers in an urban school district, who participated in the existing bioinformatics summer PD, then engaged in approximately 20 hours of workshop sessions focused on PCK development throughout the school year as
they taught the bioinformatics unit. The initial findings show that there are unique components of PCK for data literacy that are different from those for science education. The participating teachers were able to surface their knowledge of student understanding of data and strategies for teaching with and about data to define a number of components of PCK that begin to build a framework for what PCK for data literacy might look like in a generalized form that could be used to support additional teachers seeking to growth their ability to teach with complex data. Additionally, strategies used in the extension PD, such as CoRes and video reflections were shown to offer support for teachers in their implementation of authentic data literacy in their classrooms.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS .................................................................................................................. III

ABSTRACT ....................................................................................................................................... IV

LIST OF TABLES ................................................................................................................................. X

LIST OF ILLUSTRATIONS ................................................................................................................ XI

CHAPTER 1. INTRODUCTION ........................................................................................................ 1

1.1 Research Questions .................................................................................................................. 10

CHAPTER 2. BACKGROUND LITERATURE .................................................................................. 12

2.1 STEM Integration ...................................................................................................................... 12
   2.1.1 Data Literacy and its Role in STEM Integration ................................................................. 15
   2.1.2 Teaching and Learning for Data Literacy ......................................................................... 22

2.2 Pedagogical Content Knowledge ............................................................................................ 28
   2.2.1 PCK of STEM Integration ............................................................................................... 31
   2.2.2 PCK of Data Literacy ...................................................................................................... 43

2.3 Characteristics of High Quality PD for Data Literacy in STEM Integration ......... 46
   2.3.1 Characteristics of High-Quality PD .................................................................................. 46
   2.3.2 PD for STEM Integration ................................................................................................. 47
   2.3.3 PD for PCK ....................................................................................................................... 53

CHAPTER 3. CONCEPTUAL FRAMEWORK ................................................................................. 63

3.1 The Bioinformatics PD ............................................................................................................ 64
   3.1.1 PD for STEM and Data Literacy ...................................................................................... 64
   3.1.2 PD for PCK ....................................................................................................................... 66

3.2 Subject Matter Knowledge of Data Literacy ......................................................................... 67
   3.2.1 Data in Context ................................................................................................................ 68
   3.2.2 Variability in Data .......................................................................................................... 69
   3.2.3 Data in Aggregate .......................................................................................................... 70
   3.2.4 Data Visualization .......................................................................................................... 71
   3.2.5 Inference with Data ....................................................................................................... 72
   3.2.6 Complexity of Data Sets ............................................................................................... 73

3.3 PCK for Data Literacy .............................................................................................................. 73
3.3.1 Orientations to Teaching ................................................................. 74
3.3.2 Knowledge of Students’ Understanding ............................................... 76
3.3.3 Knowledge of Instructional Strategies .................................................. 77

CHAPTER 4. METHODS ........................................................................... 80

4.1 Context ......................................................................................... 81
4.1.1 The Bioinformatics PBL Unit .............................................................. 83
4.1.2 The Existing Summer PD ................................................................. 84
4.1.3 Bioinformatics PCK Extension Workshop Series ................................. 89

4.2 Participants ................................................................................... 96
4.2.1 Hallie ......................................................................................... 98
4.2.2 Mary .......................................................................................... 99
4.2.3 Manisha .................................................................................... 101
4.2.4 Will .......................................................................................... 102

4.3 Data Sources ............................................................................... 105
4.3.1 Surveys ..................................................................................... 105
4.3.2 Discussion Forums ..................................................................... 106
4.3.3 Synchronous Virtual Workshop Sessions ....................................... 107
4.3.4 CoRes ...................................................................................... 107
4.3.5 Written Reflections .................................................................... 110
4.3.6 Concept Map ............................................................................ 113
4.3.7 Interviews ............................................................................... 113
4.3.8 Classroom Observations ............................................................. 114

4.4 Data Analysis .............................................................................. 115
4.4.1 Memos .................................................................................... 115
4.4.2 Coding .................................................................................... 116
4.4.3 Secondary Sources ................................................................ 119
4.4.4 Building Cases ........................................................................ 119
4.4.5 Researcher Positionality ............................................................... 120

CHAPTER 5. FINDINGS ON RQ1: THE COMPONENTS OF PCK FOR DATA LITERACY SURFACED BY TEACHERS ................................................................. 123

5.1 Orientation to Teaching With and About Data ..................................... 126
5.1.1 Conceptions of Teaching and Learning Data ...................................... 127
5.1.2 Conceptions of the Nature of Data ................................................. 137
5.1.3 Conceptions of the Goals and Functions of Teaching With and About Data ... 144

5.2 Knowledge of Students’ Understanding of Data ................................... 153
5.2.1 Students’ Interest in and Motivation to Learn Data Literacy ............... 153
5.2.2 Students’ Orientation to Data .......................................................... 157
### 5.3 Knowledge of Strategies for Teaching With and About Data

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3.1 Making Connections to Students’ Lives</td>
<td>172</td>
</tr>
<tr>
<td>5.3.2 Making Data Tangible</td>
<td>175</td>
</tr>
<tr>
<td>5.3.3 Framing Scientific Argumentation as Storytelling in Order to Build Inferences</td>
<td>178</td>
</tr>
<tr>
<td>5.3.4 Interrogating Data by Questioning Assumptions</td>
<td>183</td>
</tr>
<tr>
<td>5.3.5 Scaffolding the Tools Through Modeling</td>
<td>186</td>
</tr>
</tbody>
</table>

### 5.4 Summarizing the Components of PCK for Data Literacy

### CHAPTER 6. FINDINGS ON RQ2: HOW THE PD COMPONENTS SUPPORTED CHANGES IN TEACHERS’ KNOWLEDGE

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Video Reflections as Affirming Anchors</td>
<td>193</td>
</tr>
<tr>
<td>6.2 Content Representations as Guidelines for Discussion and Reflection</td>
<td>198</td>
</tr>
<tr>
<td>6.2.1 Developing Big Ideas</td>
<td>201</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.3 Building a Community of Support</td>
<td>203</td>
</tr>
</tbody>
</table>

### CHAPTER 7. FINDINGS ON RQ3: THE SUBJECT MATTER KNOWLEDGE AND PCK FOR DATA LITERACY DEVELOPED BY TEACHERS

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>7.1 Growth in Teacher Collective Subject Matter Knowledge</td>
<td>207</td>
</tr>
<tr>
<td>7.2 Growth in Teachers’ Individual and Personal PCK and Subject Matter Knowledge</td>
<td>210</td>
</tr>
<tr>
<td>7.2.1 Will</td>
<td>213</td>
</tr>
<tr>
<td>7.2.2 Hallie</td>
<td>224</td>
</tr>
<tr>
<td>7.2.3 Manisha</td>
<td>232</td>
</tr>
<tr>
<td>7.2.4 Changes in PCK for Data Literacy</td>
<td>241</td>
</tr>
</tbody>
</table>

### CHAPTER 8. DISCUSSION

<table>
<thead>
<tr>
<th>Subsection</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1 The Components of PCK for Data Literacy Surfaced and Identified during the PD</td>
<td>242</td>
</tr>
<tr>
<td>8.1.1 Orientation to Teaching with and About Data</td>
<td>243</td>
</tr>
<tr>
<td>8.1.2 Students’ Understanding of Data</td>
<td>244</td>
</tr>
<tr>
<td>8.2.3 Strategies for Teaching With and About Data</td>
<td>249</td>
</tr>
<tr>
<td>8.2 How the PD Components Supported Teachers’ Knowledge Development</td>
<td>252</td>
</tr>
<tr>
<td>8.3 Changes in Teacher Knowledge for Data Literacy</td>
<td>256</td>
</tr>
<tr>
<td>8.3.1 Understanding Data in Context Primarily as Consumers not Producers</td>
<td>259</td>
</tr>
<tr>
<td>8.3.2 Variability in Data Deemed a Low Priority by Teachers</td>
<td>260</td>
</tr>
<tr>
<td>8.3.3 Data in Aggregate, the What and the So What</td>
<td>263</td>
</tr>
<tr>
<td>8.3.4 Changes in Teacher Knowledge for Data Literacy</td>
<td>264</td>
</tr>
</tbody>
</table>
8.3.4 Mastery of Data Visualization ................................................................. 265
8.3.5 Inference with Data as Conceptual Rather the Procedural .................. 267
8.3.6 Complexity of Data Sets Requires Knowledge of Tools to Master ........ 268
8.3.7 Developing PCK for Data Literacy.......................................................... 270

8.4 Limitations ........................................................................................................ 271
  8.4.1 Challenges with the context of the 2021-2022 school year .................. 272
  8.4.2 Challenges with the Bioinformatics PBL Unit ........................................ 275
  8.3.3 Developing PCK for Data Literacy vs Understanding PCK for Data Literacy . 278
  8.3.4 Making ‘Data Literacy as a Form of STEM Integration’ Explicit .......... 280

8.5 Implications for Practice ...................................................................................... 280
8.6 Implications for Future Research ...................................................................... 282
8.7 Conclusion ......................................................................................................... 284

APPENDICES ......................................................................................................... 286
Appendix A - Content Representation (CoRe) Template .................................. 286
Appendix B – Video Reflection Framework for PCK for Data Literacy .......... 287
Appendix C - Surveys ............................................................................................ 288
Appendix D - Interviews ......................................................................................... 290
Appendix E - Classroom Observation Protocol Template ................................ 292
Appendix F – Final Coding Manual ...................................................................... 296

REFERENCES ....................................................................................................... 305
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>The Six Data Literacy Competencies in the Bioinformatics PBL Unit</td>
<td>85</td>
</tr>
<tr>
<td>Table 2</td>
<td>Core Data Literacy Features of the PD Course</td>
<td>87</td>
</tr>
<tr>
<td>Table 3</td>
<td>Workshop Session Outline for Extra Treatment on PCK Development</td>
<td>90</td>
</tr>
<tr>
<td>Table 4</td>
<td>Demographics on Teachers and Schools</td>
<td>98</td>
</tr>
<tr>
<td>Table 5</td>
<td>The Big Ideas Explored in the CoRes</td>
<td>108</td>
</tr>
<tr>
<td>Table 6</td>
<td>Example Completed CoRe on Data Visualization</td>
<td>109</td>
</tr>
<tr>
<td>Table 7</td>
<td>The Big Ideas Identified During the Workshop Sessions</td>
<td>209</td>
</tr>
<tr>
<td>Table 8</td>
<td>Results of Likert-Scale Survey Questions on Data Literacy</td>
<td>212</td>
</tr>
<tr>
<td>Table 9</td>
<td>Results of Likert-Scale Survey Questions on STEM Integration</td>
<td>212</td>
</tr>
</tbody>
</table>
LIST OF ILLUSTRATIONS

Figure 1  Situating Data Literacy Within Other Overlapping Fields ..................17
Figure 2  The Refined Consensus Model (RCM) .............................................36
Figure 3  Theoretical Model for PCK of Data Literacy ..................................44
Figure 4  Example of a Completed CoRe for Genetics .................................54
Figure 5  Conceptual Model for PD for PCK for Data Literacy in STEM Integration ...............................................................63
Figure 6  The Bioinformatics PBL Unit: Asthma and Air-quality Problem ....83
Figure 7  Example Completed Video Reflection Framework ..........................112
Figure 8  Summary of the Components of PCK for Data Literacy .................124
Figure 9  Concept Map on PCK for Data Literacy .......................................126
Figure 10 Concept Map on PCK for Data Literacy: Relevance Section ..........131
Figure 11 Concept Map on PCK for Data Literacy: Google Sheets Section ......136
Figure 12 Concept Map on PCK for Data Literacy: Fear of Numbers Section ....160
Figure 13 Concept Map on PCK for Data Literacy: Data Conceptions Ideas .....163
Figure 14 Concept Map on PCK for Data Literacy: Interrogating Data Section ...168
Figure 15 Concept Map on PCK for Data Literacy: Modeling the Process Section..................................................................................187
Figure 16 Concept Map on PCK for Data Literacy - Will’s Posts ..................221
CHAPTER 1. INTRODUCTION

One of the main tenets of STEM (science, technology, engineering, and mathematics) literacy and a cross-cutting theme for STEM fields and careers is the ability to work with, understand, and interpret, real-world data (English, 2016; Gebre, 2018; Kelley & Knowles, 2016; Kjelvik & Schultheis, 2019). As computing power increases, data sets that store information on everything from air quality in locations across the globe, to detailed public health records get bigger and more complex with many more variables available to study (Dorsey & Finzer, 2017; Pentland, 2013). This increase in available data drives an increased need for those working across the STEM fields such as bioinformatics to be comfortable working with large datasets (Dorsey & Finzer, 2017; Wilkerson & Polman, 2020; Wolff et al., 2016). Additionally, data literacy is becoming increasingly necessary for navigating daily life in which almost every aspect is permeated by data. Data dictates many of the choices we make in our daily lives about decisions such as our health, which products we buy, and which route we take while driving. There are additional choices made for us by algorithms, computers, and the many devices we use such as which ads and news we are exposed to and what levels of certain chemicals or pollutants exist in our drinking water or air (Dorsey & Finzer, 2017; Erikson, 2020; Gould et al., 2016; Wolff et al., 2016). Being able to understand the way data is used by those around us, as well as learning to interpret and use it in turn is a powerful tool, both for surviving in the fast-paced world of technology, but also for thriving and wielding authority to help enact change in our communities through scientific action (Schultheis & Kjelvik, 2020; Van Wart, Lanouette, & Parikh, 2020; Wise, 2020).
While the field of data science has expanded over the last few decades to address the explosion of data and the way it permeates so many aspects of life, education, particularly in K-12, has not yet caught up, resulting in strong calls for greater focus on data literacy within education at all levels (Frank et al., 2016; Gebre, 2018; Gould et al., 2016; Wolff et al., 2016). Similar calls are being made for STEM education more broadly, with a particular focus on integrating technology and engineering skills into primary and secondary school curricula and on grounding learning in STEM fields in real-world contexts and problems (Kelley & Knowles, 2016; NRC, 2014; Nadelson & Seifert, 2017). Addressing real-world contexts and problems requires collecting, interpreting, and communicating real-world data, which is often big and messy, so in addition to a need for increased focus on data science education, there is also a need for new approaches to how data is integrated into STEM classrooms in order to prepare students for working with the big, messy data sets that imbue research, industry and society (Dorsey & Finzer, 2017; Lee & Wilkerson, 2018; Nadelson & Seifert, 2017).

Though it is interdisciplinary in nature, existing as a common component of education in all of the STEM fields, data literacy has suffered from being nebulously defined and not having a clear home amongst the STEM fields (Gebre, 2018; Gould et al., 2016; Wolff et al., 2016). Research has been conducted and progress has been made toward identifying components and tools for teaching with data separately in math education (e.g., Vahey et al., 2012), science education (e.g., Finzer et al., 2018), computer education (e.g., Hassan & Liu, 2019) and other fields around data literacy, mainly in out-of-school time and at the higher education level (Gould et al., 2016; Rubin, 2020). However, it is unclear how this research integrates into secondary school
classrooms, specifically for complex types of data such as large public data sets, in a way that supports connections and patterns across the STEM fields (Lee & Wilkerson, 2018). As Nadelson and Seifert (2017) highlight, as careers in STEM grow more integrated, there is a misalignment between how STEM is taught and how it is experienced in the real world. The siloed nature of K-12 education, especially at the secondary level, where even science and math are broken down into separate sub-disciplines is not how these topics are studied and engaged with in a world where fields such as bioinformatics, environmental chemistry engineering, and nanotechnology manufacturing that combine multiple STEM disciplines are becoming more prevalent (Aydin-Gunbatar, et al., 2020; Kelley & Knowles, 2016). Increasing the level of integration for STEM in secondary schools requires a shift in the way these subjects are currently taught. Data literacy is inherently STEM integration in that it naturally and necessarily combines technology through data analysis tools, mathematics through the statistics of data analysis and science through the context from which the data arises (Kjelvik & Schultheis, 2019). As such, much of the research conducted on teachers’ knowledge and skills for STEM integration can be applied to data literacy.

STEM integration has been shown to present a challenge to teachers who have received training in only one subject area (Aslam et al., 2018; Kelley et al., 2020). The hesitancy among teachers for STEM integration, shows a need for additional support for teachers to increase integration of STEM topics and bring data into their classrooms in an authentic way that prepares students for engaging with all types of data in the real world (Aydin-Gunbatar, et al., 2020; Gould et al., 2016; Kelley & Knowles, 2016; Wise, 2020). Though science teachers believe that data is an important part of science education, most
engagement with data in science classrooms comes in the form of labs or activities
designed to teach a known relationship or concept rather than engagement with real-
world problems that involve emergent types of data (Finzer et al., 2018; Hardy, Dixon, &
Hsi, 2020; Margot & Kettler, 2019). Though these activities sometimes engage
mathematics and technology, the engagement is often surface level and not made explicit
as the data is usually simple and neatly organized.

As the need for data literacy that uses messy data that spans across disciplines
grows, the way we seek to teach students to interact with data also needs to change to
promote a broader relationship with data (Kjelvik & Schultheis, 2019; Lee & Wilkerson,
2018; Wise, 2020). Teachers need additional support to develop knowledge and tools for
integrating that type of data into their classrooms, including both subject matter
knowledge about data and how to work with it, but also knowledge about how to teach
with it (Kelley & Knowles, 2016; NRC, 2014). This professional knowledge about
teaching a particular subject in a particular context is often called pedagogical content
knowledge (PCK). It is a concept first introduced by Shulman (1986) as a way to measure
the specific knowledge that teachers hold about how to engage their particular students
with different concepts, such as an elementary math teacher knowing that students often
forget to carry from the tens place when subtracting two larger numbers and then having
a strategy to help students overcome this particular barrier (Ball et al., 2008). The concept
of PCK will be unpacked further in later chapters as it is the primary knowledge focus of
this study. Currently very little is known about what PCK for teaching with and about
data for the purpose of developing data literacy in a classroom looks like and what sort of
experiences are needed to support it (Aydin-Gunbatar et al., 2020). Though some studies
have made suggestions for instructional strategies for teaching data literacy based on research about how students learn with data (e.g., Lee & Wilkerson, 2018; Wilkerson & Polman, 2020) there is a dearth of research exploring the knowledge that teachers might already have about PCK for data literacy and how they conceive of how students understand data and how to engage them with it. This study hopes to add to the research in this area.

Through an existing project designed to promote STEM integration, and engagement with cutting-edge real world science in existing secondary school science classrooms, our research team sought to design a bioinformatics unit and a corresponding professional development (PD) program to support teachers in implementing the unit in their classrooms (NSF Grant No. #1812738). Bioinformatics, described by our team of content experts as “an interdisciplinary field that combines aspects of computer science, mathematics, and statistics to collect, store, manage, analyze and interpret biological data” is a perfect example of a STEM-integrated field (Nadelson & Seifert, 2017) which requires data literacy and in which there is an increasing need to train future experts (Attwood et al., 2019; Cooper et al., 2017; Martins et al., 2020). Our PD program involved a three week summer PD which was designed based on characteristics of high-quality PD (Darling-Hammond et al., 2017). During the PD teachers engaged with content and skills around bioinformatics, data literacy, computational thinking, and mobile learning. This was followed up by continued support during the school year in the form of monthly PD sessions and in-person classroom support. While we found some success in our pilot year, with most teachers completing implementation of the
intervention, we ran into many of the same barriers and challenges proposed by previous research (e.g., Aslam et al., 2018; Gould et al., 2016; Kelley et al., 2020; NRC, 2014).

Due to the exploratory nature of the project, and its vast scope with goals across bioinformatics, culturally relevant pedagogy, data literacy, mobile learning, and scientific agency, success in implementation was varied across teachers and across the goals of the project. While the pilot year included some teachers who entered the program with a higher level of data literacy who had greater success implementing, results from the pilot year of implementation showed that some teachers came into the PD with hesitancy about data integration (Miller et al., 2021). Despite a specific focus on data literacy as one of the primary components of the curriculum and instruction framework, these teachers were still struggling with content knowledge for data integration during implementation (Miller et al., 2021; Yoon et al., in press). This is encapsulated well in the following quote from one of the pilot teachers during the post-implementation interview, “We were good in terms of the other science concepts that were there, like asthma and air quality particles. But as far as the statistics and relating that real research to our... and teaching our students that, I think I was a little bit under prepared.” This showed that while many of the other components of the bioinformatics unit including the biology content and mobile learning were implemented successfully, many of the teachers struggled with the component of data literacy (Miller et al., 2021; Yoon et al., in press). Additionally, some of the teachers also experienced barriers to implementing student-centered inquiry-based instruction (Noushad et al., 2021; Yoon et al., in press) and there was a perceived disconnect between teacher knowledge and implementation in the classroom (Miller et al., 2021; Shim et al., 2021).
In order to address teachers’ discomfort with data literacy concepts, we modified the PD for the second cohort and created additional resources for the curriculum, including teacher guides with notes for instruction (Yoon et al., in progress). As well as moving the entire PD experience online to increase anywhere anytime access (due to the COVID pivot), we also increased the amount of time spent explicitly discussing data literacy concepts in order to build their subject matter knowledge for data. We laid out common student misconceptions around engagement with data and built in a number of discussion prompts that asked teachers to reflect on data integration in their classroom in an effort to engage them with thinking about PCK (Miller et al., 2021). We also created better alignment between the PD and the student-facing curriculum and providing stronger student facing-materials that aligned with pedagogical supports teachers were familiar with (Yoon et al., in progress). Despite these efforts to increase support for learning and working with data literacy concepts, preliminary results from the second cohort of the PD showed that teachers were still experiencing barriers to implementation of data literacy.

Though some of the changes made to the PD were aimed at development of PCK, specifically the encouragement of reflection on practice through discussion prompts, research into PCK development has shown that teachers do not always reflect in ways that support their own knowledge growth (Monet & Etkina, 2008; van Driel & Berry, 2017). While research on PCK for data literacy is mostly lacking, research on PCK for STEM integration has shown that in order to successfully integrate STEM learning in their classrooms, teachers need to understand how students learn and apply STEM and be able to ground STEM concepts in the learning context and prior knowledge of their
students (El-Deghaidy et al., 2017; Vossen et al., 2020). PCK is knowledge about how to teach a specific topic within a specific context (Shulman, 2015). As such, pedagogical strategies for data literacy implementation, like strategies for STEM integration, are unique from teachers’ existing PCK for science education, and they must be explicitly taught to and modeled for teachers (Aydin-Gunbatar et al., 2020; Margot & Kettler, 2019).

More explicit discussion of the components of PCK, and tools for scaffolding development of PCK are needed to make PCK a focus of teacher learning in our PD. This is supported by an observation made by a number of the participating teachers in a focus group interview at the end of the PD stating that they felt there was still a gap between their content knowledge and pedagogical skills, especially when it came to the data literacy component of the unit (Miller et al., 2021). However, the lack of research on PCK for data literacy makes it difficult to determine how to support teachers in developing it. While one of the recommendations from the NRC report on STEM Integration (2014) was for an increase in PD opportunities for teachers to support them in developing effective approaches for teaching STEM-integrated curricula (NRC, 2014, Recommendation 8, p. 148) and there has been some research done on developing PCK for STEM integration since (Aydin-Gunbatar et al., 2020; El-Deghaidy et al., 2017; Vossen et al., 2020), there is a major dearth of research on PCK for data literacy as a key component of STEM integration. Most of the research on PCK for working with data has been conducted in the field of statistics education, though the research is limited there as well (Callingham et al., 2016; Ijeh & Onwu, 2013). Some research has been done on how students think about and learn with data, of which Lee and Wilkerson (2018) offer a
review, and they and (Wilkerson & Polman, 2020, Wise, 2020; Wolff et al., 2019) have made theoretical suggestions for instructional strategies for teaching data literacy. However, there is a lack of research into how well those strategies work in the classroom. Additionally, there has been little to no research done on PCK for data literacy, that is, what teachers actually know about student thinking on and instructional strategies for data literacy.

This project hopes to build on the limited research done on PCK for STEM integration and for statistics education to develop an understanding of PCK for data literacy, a yet unexplored concept. A number of additional strategies for specifically developing PCK were added to the existing PD for the 2021 summer including the use of content representations (CoRes, Appendix A) which challenge teachers to think deeply about implementation (Aydin-Gunbatar et al., 2020; Loughran et al., 2012; Walan, Nilsson, & Ewen, 2017) and an increased focus on high-level reflection for pedagogical strategies (Gess-Newsome, 2019; Daehler et al., 2015). CoRes are a strategy that has been used to measure and develop teachers’ PCK by many studies across different disciplines such as chemistry (Alvarado et al., 2014; Aydin et al., 2013), physics (Cooper et al., 2015), different topics within elementary school science (Nilsson & Loughran, 2012; Walen, Nilsson & Ewen, 2017), and technology (Williams & Lockley, 2012). A CoRe is a simple template that prompts teachers to identify a big idea about the subject they are planning to teach. It then prompts them to write responses to questions about not only how they plan to teach that big idea, but also why they are planning to teach it that way. It uses prompts such as Knowledge about students’ thinking which influences your teaching of this idea to make PCK visible so that it can then be discussed or reflected on.
As the CoRe template was used heavily as a tool for making teachers’ PCK visible during this study, it will be discussed in more detail later with multiple examples provided.

In order to determine what knowledge teachers have about how to teach data literacy and how that knowledge can be engaged, a cohort of four teachers from the larger summer 2021 Bioinformatics PD cohort were recruited to participate in a PCK Extension Workshop Series. The Workshop Series continued into the fall and winter and included additional work with CoRes, video reflections of their own and colleagues’ teaching, and extended reflections through both written assignments and interviews (Aydin-Gunbatar et al., 2020; Evens et al., 2015; Gess-Newsome, 2019; Wilson et al., 2018). Through analysis of the CoRes, transcripts of the sessions, and written reflections, combined with classroom observations and interviews, this project hopes to add to the literature about PCK for data literacy including the knowledge that teachers already held about teaching data literacy and how participating in a PD which explicitly addresses it changed that knowledge.

1.1 Research Questions

This study seeks to better understand how PD can support teachers in surfacing their existing knowledge of teaching data literacy in a high school science classroom and using that identified knowledge to develop PCK for data literacy further in order to enhance STEM integration. In order to do so, I ask the following questions:

1. What components of PCK for data literacy are surfaced by the participating teachers during the PD experience?
2. How the PD components appear to support teachers’ development of subject matter knowledge and PCK for data literacy?

3. What subject matter knowledge and PCK for data literacy was developed by teachers during PD experiences that have been designed to explicitly focus on data?
CHAPTER 2. BACKGROUND LITERATURE

In building a theoretical framework for this work, three primary areas of previous research will be considered: data literacy as a form of STEM integration, PCK, and characteristics that lead to development of high-quality PD. The first section discusses the need for STEM integration in more detail and the role that data literacy plays in successful implementation of STEM-integrated teaching, as well as the challenges that teachers have in working with real-world data in classroom settings. Next, I discuss research on PCK and how it has been conceptualized. I draw on research about PCK in STEM integration and statistics education, as no research yet exists for PCK for data literacy and propose a framework for how PCK for data literacy can be conceptualized. Finally, I lay out how PD can be designed to support development of subject matter knowledge and PCK for data literacy and what strategies have been successful at doing so.

2.1 STEM Integration

Data literacy is inherently STEM integration as it requires knowledge of mathematics, technology, and the science or engineering context from which the data draws (Kjelvik & Schultheis, 2019). So, in order to understand the central role that data literacy plays in the integration of STEM in the classroom, it is first necessary to understand the goals of STEM integration and how it has been addressed in the classroom. As the four STEM fields begin to play a bigger role in an increasingly interconnected and synthesized world, there has been more discussion of STEM literacy
as one of the desirable literacies that allow people to engage with society (NRC, 2014). Being STEM literate has been defined as having a good enough understanding of ideas and processes in the four STEM fields and how they affect society in order to apply knowledge in using common technologies, critically evaluating STEM concepts in the news, and engaging in everyday problem solving (Margot & Kettler, 2019; NRC, 2014; Sneider & Purzer, 2014). The Next Generation Science Standards (NGSS Lead States, 2013), the current science standards used by a majority of U.S. states, present a new devotion to bringing STEM literacy into the mainstream of science education as they include extensive references to technology and engineering topics and practices throughout, and connections to the Common Core State Standards in Mathematics (Koehler, et al., 2016; NRC, 2014; Sneider & Purzer, 2014). The U.S. federal government’s National Science and Technology Council recently commissioned a report on STEM education (NSTC, 2018) which listed “Engaging students where disciplines converge” as a primary pathway to achieving national goals in STEM education, with a focus on “encouraging transdisciplinary learning.” Despite these high-profile uses of the term STEM in relation to education, what counts as STEM education, and especially what counts as STEM-integrated education is still not well defined (English, 2016; Kelley & Knowles, 2016; Margot & Kettler, 2019; Nadelson & Seifert, 2017). The lack of definition complicates efforts to bring STEM-integrated education into secondary school classrooms, as teachers and those who support them may have different conceptions of what STEM integration means and how to accomplish it (Burrows & Slater, 2015; El-Deghaidy et al., 2017).
The term STEM education is used to describe educational scenarios that range from mostly disciplinary in nature to more fully interdisciplinary or transdisciplinary (English, 2016; Nadelson & Seifert, 2017, NRC, 2014). Definitions range from more constricted requirements for all four fields to be equally represented and grounded in engineering practices (e.g., Bryan et al., 2016) to more inclusive definitions that place any combination of at least two of the fields directed towards real-world problems in the realm of STEM integration (e.g., NRC, 2014). Given the wide range of existing conceptualizations for STEM integration and the newness of the field of STEM-integrated education, the NRC report on STEM education (2014) decided to encourage inclusion and define STEM integration broadly, as “working in the context of complex phenomena or situations on tasks that require students to use knowledge and skills from multiple disciplines” (NRC, 2014, p. 52). This highlights a common theme across all the literature on STEM integration, which is that it should be grounded in a discovery-based, real-world, transdisciplinary project or problem that allows for learning to be student-centered and encourages problem solving and collaboration (Bryan et al., 2016; Chamberlain & Pereia, 2017; Margot & Kettler, 2019; Nadelson & Seifert, 2017; NRC, 2014).

The existing bioinformatics project that our team has been building illustrates well this definition of STEM integration. Bioinformatics, described by our team of content experts as “an interdisciplinary field that combines aspects of computer science, mathematics, and statistics to collect, store, manage, analyze and interpret biological data” is a perfect example of a STEM-integrated field that requires data literacy (Cooper et al., 2017; Kovarik et al., 2013; Nadelson & Seifert, 2017). The field of bioinformatics
is rapidly expanding with increased technological capabilities and access to population
data and data storage, much of which is publicly available through government agencies
such as the U.S. National Institute of Health (NIH) (Attwood et al., 2019; Barker et al.,
2020; Martins et al., 2020). This makes it an ideal subject for grounding STEM
integration in real-world problems that are related to students’ lives through public health
and their own genetic and environmental influences. There are calls across the field to
increase training for bioinformatics by integrating it into science classes at all levels of
education (Attwood et al., 2019; Cooper et al., 2017). Though most research into
bioinformatics education has been at the post-secondary level (e.g., Duncan et al., 2016;
Magana et al., 2014), some recent studies have examined its potential impact in
secondary school classrooms (e.g., Barker et al., 2020; Cooper et al., 2017; Martins et al.,
2020). Working within a real-world context on a complex, transdisciplinary problem,
requires the use of data, but not the type of organized, neat data that is traditionally seen
in science and math classes. It requires messy, often unwieldy data, that relies on special
knowledge and skills to manipulate and reason with (Kjelvik, & Schultheis, 2019; Wise,
2020; Wolff et al., 2019). This is where data literacy plays a vital role in facilitating
STEM-integrated units like the existing bioinformatics unit on which this research builds.

2.1.1 Data Literacy and its Role in STEM Integration

Data literacy is an emerging concept without a clear definition and the overlap
between data science, data literacy, computational literacy, and statistical literacy is still
nebulous (Gould et al., 2016; Wolff et al., 2016). As with the disagreement over the
definition of STEM literacy, this lack of common definition can act as a barrier to
implementation and study. Wolff and colleagues (2016) noted this in a recent review of literature on data literacy stating, “without a clear definition of what data literacy is, it is both hard to teach and to assess the outcomes of teaching” (p. 10). They went on to attempt 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. It is based on core practical and creative skills, with the ability to extend knowledge of specialist data handling skills according to goals. These include the abilities to select, clean, analyze, visualize, critique and interpret data, as well as to communicate stories from data and to use data as part of a design process. (p. 23)

Though this definition focuses more on the inquiry-process rather than data literacy applications, it does highlight some of the important components of data literacy, such as the specific skills required to work with data. It also grounds data literacy in real-world contexts and a focus on using data rather than simply analyzing it, which is a commonly cited difference between data literacy and statistical literacy (Gebre, 2018; Gould et al., 2016; Rubin, 2020; Wolff, et al., 2016).

In an attempt to distinguish data literacy from some of the other fields, Kjelvik and Schultheis (2019) chose to define data literacy as existing in the overlap between quantitative reasoning and data science where the data is grounded in an authentic context (See Figure 1). This framework is useful in outlining some of the important components of data literacy, namely that it involves applying mathematical principles, working with computers and other technologies, and understanding the context of the data. It also demonstrates the way that data literacy is inherently STEM integration and a way of engaging STEM integration in classrooms. Quantitative reasoning, which requires
the application of mathematical principles, aligns with the discipline of mathematics and specifically statistics. Data science often goes hand in hand with technological knowledge as it requires computational thinking, and the use of digital tools. The authentic context requires knowledge of content from a field such as biology and so could be seen as part of scientific knowledge. An important component of both STEM integration and data literacy is that they are grounded in authentic, real-world problems that require authentic real-world data, which is often complex, messy, and unlike most of the data typically used in secondary school classrooms (Bowen & Bartley, 2014; Finzer et al., 2018; Hardy et al., 2020).

**Figure 1**
_Situating Data Literacy Within Other Overlapping Fields_

![Diagram](image)

*Note: From Kjelvik & Schultheis, 2019*

In order to be able to teach data literacy effectively, teachers must develop their own data literacy, which serves as the subject matter knowledge base on which PCK for
data literacy is built (Gould et al., 2016). A focus on data literacy as a component of teacher knowledge is a relatively new concept. In the last few decades an increased emphasis on data-based assessment and data-driven instruction has led to a push to ensure that teachers are prepared to work in a data-driven environment (e.g., Mandinach & Gummer, 2016). As a result, much of the research conducted on teachers’ data literacy has focused on how teachers use data for their own and their school’s growth and development and not on how they support their students’ data literacy knowledge and skills (Gummer & Mandinach, 2015; Wolff et al., 2019). Science teachers often do have some knowledge of data as a component of science knowledge and how students are expected to use it, as data is a vital part of most science curriculum and standards (NRC, 2014).

For example, the Next Generation Science Standards (NGSS Lead States, 2013) have as one of their three interwoven dimensions, science and engineering practices. In the practices, there is extensive mention of data within the components that students are expected to master by the end of each grade band, such as:

- plan and conduct an investigation individually and collaboratively to produce data to serve as the basis for evidence;
- select appropriate tools to collect, record, analyze, and evaluate data;
- consider limitations of data analysis (e.g., measurement error, sample selection) when analyzing and interpreting data; and
- apply scientific reasoning, theory, and/or models to link evidence to the claims to assess the extent to which the reasoning and data support the explanation or conclusion.
The problem is that for many science teachers, data is usually taught and applied in service of learning a specific topic within their discipline (Dorsey & Finzer, 2017; Gould et al., 2016; Hardy et al., 2020) and doesn’t allow for students to explore more complex datasets. Science teachers don’t always have the subject matter knowledge for teaching with and about complex data because they were not trained or given experience working with messy real-world data sets (Hardy et al., 2020). Research on statistics literacy has also shown that teachers in both science and mathematics don’t always have the content knowledge to engage with large complex data sets in a meaningful way (Chick & Pierce, 2012; Zieffler et al., 2018). In working with math teachers, Gould and colleagues (2016) found that the teachers didn’t have a background in asking problem-based statistical questions or working with messy data sets. In the existing bioinformatics project, our team attempted to address this misalignment between how data is typically used in a science classroom and the more robust approach to data literacy needed to engage with bioinformatics problems. In order to do this, we used two independently developed but significantly aligned frameworks for what constitutes subject matter knowledge in data literacy that are briefly reviewed next.

Lee and Wilkerson (2018) draw on literature from both science and statistics education, to outline four understandings for data literacy that tend to be common sources of struggle for secondary school students: a) measurement and sampling, which includes understanding how measurements are taken, how samples for measurement are chosen, and potential sources of error and consequent uncertainty; b) the characteristics of data, including measures of centrality, distribution, and variability which encourage aggregate or global views of data sets rather than case-based or local views; c) different types of
data representation and how they can be read and used; and d) making inferences from data, which involves creating a scientific argument using evidence and reasoning to support a claim. Separately in a commentary on a special issue of the *Journal of the Learning Sciences* devoted to data science in education, Rubin (2020) synthesized the articles and came up with similar understandings to Lee and Wilkerson (2018). While Rubin (2020) includes *data visualizations* and *inferences* she breaks down Lee and Wilkerson’s other two understandings in slightly different ways into three: a) context, the who, when, where, what, why and how of data collection and use, b) variability, understanding variability in data, its causes, and how to analyze it, and c) aggregate, which requires leading students from a case-based view of data to an aggregate view that considers frequency and patterns across the whole data set. These five data literacy content knowledge areas were explicitly outlined for teachers in the existing PD and time was created for them to investigate the concepts and reflect on both their own learning and how they could be implemented in the classroom. However, when implementing in the classroom, many teachers still ran into barriers with their own data literacy, specifically in the area of using tools to manipulate the messy data sets collected by the students (Miller et al., 2021). In order to address this barrier, an additional data literacy content area was added to the PD for summer 2021, which is the idea of data complexity.

Kjelvik and Schultheis (2019) lay out five features of data complexity that each exist on a scale from simple to complex: scope, selection, curation, size, and messiness. *Scope* refers to whether the dataset includes only the relevant variables to the students’ problem or question, or whether it includes additional variables that must be parsed through. *Selection* determines whether within a given dataset the relevant variables are
selected for the students or whether they need to determine which variables will define their dataset. *Curation* allows for a range from fully-curated datasets that have been summarized already and are ready for analysis to raw data sets that have not been curated at all and may require data manipulation and transformation. The *size* of a dataset is an obvious component of complexity and in the world of big data and increasingly accessible storage, the maximum size for datasets is growing exponentially to cover the entire history of humanity and the depths of the known universe. Finally, *messiness* is a measure of the extent to which individual data points have been accounted for and can include dealing with outliers and missing data. In order to build data literacy and an ability to work with authentic datasets, it is important for students to encounter data during their schooling that does not lie only at the simple end of the spectrum on these five components but builds into increasing complexity (Gould et al., 2016; Kjelvik & Schultheis, 2019; Wolff et al., 2019). While *complexity of datasets* was not a component of data literacy presented in the first two years’ PD, the revised PD and bioinformatics unit for year 3 included additional scaffolding to support development of this component of subject matter knowledge for data literacy.

The next section discusses in more detail some pedagogical strategies for teaching data literacy as a form of STEM integration and why teachers often struggle with them. These strategies are important to consider because PCK builds on both content knowledge and pedagogical knowledge to build a bridge between knowledge of content and domain and the actual practice of teaching in a classroom (Shulman, 1986; 1987).
2.1.2 Teaching and Learning for Data Literacy

As previously discussed, despite its importance to STEM integration and to living and thriving in a world awash in data, there has been limited research focused on pedagogical best practices for teaching data literacy, especially through complex, messy data (Gould et al., 2016; Erikson, 2020; Wolff et al., 2019). Wilkerson and Polman (2020) highlight what they call two “core pedagogical commitments” for data science: first, teachers must not only teach the technical skills associated with working with data, but also focus on developing flexibility in order to ensure that learning will transfer to new tools and methods as the capabilities of technology continue to advance, and second, data exploration should be grounded in real-world investigations that are consequential and relevant to students, in which they can engage with data through a whole process from collection to analysis to communication of findings. In a previous article, 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. Work by Erikson and colleagues (2019) on specific “data moves” necessary for working with messy data show that it is also important to ground the teaching of technical skills in the reasoning for doing them. Doing so can help develop the flexibility for transfer highlighted by Wilkerson and Polman (2020).

Wolff, Wermelinger, and Petre (2019) propose six design principles for supporting the development of data literacy: a) inquiry principle, data use should be scaffolded through a full inquiry cycle; b) expansion principle, rather than starting with
large complex datasets, begin with a smaller sample which reduces complexity and expand outward as comfort develops; c) context principle, the context for the data should be familiar and relatable to students; d) foundational competencies principle; rather than teaching skills specific to one dataset or question, make connections to larger ideas and transferable skills; e) STEAM principle, add in the A for arts to STEM exploration and encourage students’ creativity; and f) personal data collection principle, students should engage in data collection themselves, so that even if working with a large external dataset, they understand the process in which the data was collected. These six principles build on those highlighted previously (Lee & Wilkerson, 2018, Wilkerson & Polman, 2020) as well as echoing work done by Kjelvik and Schultheis (2019) about the need to scaffold complex data sets for students by starting with a simplified sample and gradually building toward complexity.

Furthermore, while developing subject matter knowledge for data is vital for teaching data literacy, when embedded in a STEM-integrated setting, there is additional content to attend to, the most important of which is the need to make the integration between the disciplines explicit for all students (English, 2016; Gardner & Tillotson, 2019; Kelley & Knowles, 2016; NRC, 2014). When learning content in a STEM-integrated context, students do not necessarily make the connections between the disciplines (Brown & Bogiages, 2019; Nathan et al., 2013), or between the specific context they are exploring and larger STEM principles and patterns (Bryan et al., 2016; NRC, 2014). In a study on how students understand connections between STEM disciplines, Nathan and colleagues (2013) found that students can be shown to make meaningful connections when teachers construct clear links between representations and
terms from different disciplines as well as making forward and backward connections to
text and ideas that were learned previously or will be addressed in the future. The
NRC STEM integration report (2014) built on this research and highlighted the need to
also make connections to larger patterns and principles, as students can become hyper
focused on the specific context and need guidance to see the bigger picture. A study by
English and King (2019) shows that strategically using epistemic reflection with students
participating in STEM integration can support their development of connections between
the different disciplines and skills. Technology is particularly difficult for both teachers
and students to integrate fully into classroom learning (El-Deghaidy et al., 2017; Gerard
et al., 2011; Mishra & Koehler, 2006; NRC, 2014). While teachers often struggle with
usage of technology and how to integrate technological tools in meaningful ways (Chai et
al., 2019; El-Deghaidy et al., 2017; Gerard et al., 2011), students often struggle to make a
connection between technology as a tool and the larger way in which the usage of
technology can impact society (Kelley & Knowles, 2016; Niess, 2017; NRC, 2014).

One framework for teaching for STEM integration that provides space for larger
connections to be made and is used by the existing bioinformatics project from which this
project is drawing, is problem-based learning (PBL). Though PBL is a framework that
can promote learning that supports making connections between the STEM disciplines in
a way that is organic and meaningful, it also presents challenges for teachers attempting
to implement it. It is therefore important to understand what is unique about teaching
with PBL as it will provide part of the learning context in which teachers in this study are
implementing.
2.1.2.1 PBL in STEM Integration. According to Savery (2006) in the inaugural issue of the *Interdisciplinary Journal of Problem-Based Learning* PBL is, “an instructional (and curricular) learner-centered approach that empowers learners to conduct research, integrate theory and practice, and apply knowledge and skills to develop a viable solution to a defined problem” (p. 12). Savery goes on to outline some essential characteristics of PBL which include that it should be student-centered, involve collaboration, grounded in authentic real-world situations, and integrate concepts and skills from a variety of disciplines and subject areas. What makes PBL particularly useful and relevant for STEM integration is that it centers the authentic problem scenario and places students in the position of authority to determine the best course of action (Asghar et al., 2012; Bryan et al., 2016; NRC, 2014). Additionally, it embeds the important STEM concepts in the real-world issues that, if designed well, are relevant to students’ lives and can inspire motivation and interest (Finzer et al., 2018; Krajcik, 2015). PBL can also encourage and guide the use of technological tools and complex data in order to build a strong argument for a particular solution to the given problem (Koehler et al., 2016; Niess, 2017).

The team at The Concord Consortium is specifically developing PBL units that engage students with large, complex datasets in order to support students in learning to work with datasets and data visualizations and develop the skills to choose which technologies and data are appropriate to address different problems (see Finzer et al., 2018). However, they and others have found that like other aspects of STEM integration, teachers often have not received training or experience in implementing PBL lessons and units and that an open-ended, student-centered learning environment is not something
they are comfortable with (Asghar, 2012; Aydin-Gunbatar et al., 2020; Kelley & Knowles, 2016; Nadelson & Seifert, 2017; Rosenberg et al., 2020). This discomfort with PBL is something that the existing bioinformatics PD attempts to dispel through extensive discussion of the affordances of PBL as a framework as well as modelling of the curriculum and how it supports student learning.

Working with data is a component of standard curricula for both math and science (NGA, 2010; NGSS Lead States, 2013) which means that many experienced teachers in these two subject areas have developed some PCK for working with data. However, most data analysis and visualization that happens in science classrooms serves to teach a known relationship or concept (Dorsey & Finzer, 2017; Finzer et al., 2018; Gould et al., 2016; Hardy et al., 2020) and teachers haven’t always been taught how to engage with messy, authentic data (Bowen & Bartley, 2014; Kjelvik & Schultheis, 2019). In arguing that data should be actively produced by students rather than passively collected, Hardy, Dixon, and Hsi (2020) outline three primary issues with the traditional way of teaching with data, which is to forefront the disciplinary ideas while sidelining the students and the tools used to collect the data as active agents in the process. These are, a) limiting the effects of material context and possible resistance; b) pre-prescribing tools and methods; and c) creating fixed and unnecessarily narrow goals and purposes for data collection. By engaging in these three practices in the classroom, teachers limit student ability to develop a deeper understanding of the nature of data and its relationship to the tools used to collect it, themselves, and the world around them. In order to address the issues outlined by Hardy and colleagues (2020), teachers need to not only change their curriculum and teaching strategies to more open-ended explorations, such as PBL, they
also need to develop specific PCK for providing students support in developing data literacy. Additionally, as outlined by Kjelvik and Schultheis (2019) teachers need to develop an understanding of the characteristics of complex data sets and how to properly scaffold these for students, which includes components of both knowledge of students understanding and knowledge of teaching strategies.

While the existing bioinformatics curriculum utilized many of these pedagogical strategies and attempts to highlight some of them in the teacher guides and power point notes in order to make them explicit for teachers, classroom observations and teacher interviews from previous cohorts showed that big picture connections are not being made by or for students in the classroom. When asked during a post implementation interview whether they thought students understood the relationship between the data skills they were learning and how they could be used to support action for improved health or environmental awareness, one teacher responded, “I don't know that they really did. I don't know that I did that well. I left it up to them to see how they would think about it, that was my mantra. So, I don't know that they're necessarily going to.” Another teacher, responding to the same question said simply, “I would definitely say no to that one.” These quotes show that despite the efforts of the existing bioinformatics curriculum and supporting PD, teachers were still not able to use real-world data in a STEM-integrated unit in a way that allowed students to develop data literacy that would transfer beyond the specific project and guide connections between the disciplines of math, science, and technology to support engagement in a transdisciplinary field such as bioinformatics.

For all of the pedagogical strategies discussed in this section, one of the primary barriers to implementation found in prior research is teachers’ experience and training or
lack thereof (Aydin-Gunbatar et al., 2020; El-Deghaidy et al., 2017; Gould et al., 2016; Kjelvik & Schultheis, 2019; Margot & Kettler, 2019;). Teachers have often not received training in data-centered lessons (Finzer, 2013; Kjelvik & Schultheis, 2019) or in STEM integration (Aydin-Gunbatar, 2020; Brand et al., 2020; El-Deghaidy et al., 2017; Vossen et al., 2020; Yang et al., 2020) which means that they often require additional resources and development to implement STEM-integrated lessons that use authentic messy data (Gould et al., 2016; Kelley & Knowles, 2016; Kjelvik & Schultheis, 2019). This additional support needs to address both content knowledge and pedagogical knowledge, but additionally, as this research seeks to explore, the amalgam of the two, often referred to as pedagogical content knowledge (PCK). In the next section I will define PCK and explore how it has been conceptualized by the different fields of science, technology, and mathematics education in order to find common components which can be developed through PD to better support teacher’s implementation of STEM-integrated curriculum and data literacy learning.

2.2 Pedagogical Content Knowledge

PCK was introduced by Shulman (1987) as that “special amalgam of content and pedagogy that is uniquely the province of teachers, their own special form of professional understanding” (p. 8). The concept was deeply compelling to many education researchers as a way to conceptualize teacher professional knowledge (Gess-Newsome et al., 2019). By its nature, PCK is subject specific as it is the knowledge of how to teach a specific topic in a specific way (Smith & Banilower, 2015). For example, an experienced high school biology teacher with high levels of PCK would have knowledge about the many
preconceptions that students might have about evolution, such as that many high school students think of evolution as occurring at an individual level rather than a species level. They would also have at their disposal knowledge and experience for which strategies could be successful at engaging those preconceptions and shifting students’ understanding (e.g., a specific computer simulation designed to model evolution occurring over multiple generations and which probing questions would trigger students to see and internalize this multigenerational process). This knowledge about evolution would be unique to this particular topic and would not help the teacher to teach calculus, newton’s laws, or even about the structure of cells.

As a result, similar to research on teaching, research on PCK has mostly developed along parallel strands in the fields of science, math, and technology education (Chan & Hume, 2019). However, there are some commonalities across the strands, primarily that PCK is a distinct form of knowledge, separate from subject matter knowledge, but that it is affected by subject matter knowledge which can constrain or contribute to a teacher’s development of PCK (Baumert et al., 2010; Gess-Newsome et al., 2019; Park et al., 2012). Furthermore, a focus on PCK within teacher education and PD across the different disciplines has shown that it is a dynamic form of knowledge that grows over time and can be developed through targeted support (e.g., Baumert et al., 2010; Gess-Newsome et al., 2019). This is a foundational finding for the proposed research, which seeks to explore PCK development through targeted PD.

However, also important to this research is the consensus across the domains of research on PCK, that it is not a monolith, but rather made up of a number of components and levels that are applied in different phases and contexts of teaching. Furthermore, each
of these different levels and components needs to be developed, measured, and assessed differently (Chan & Hume, 2019; Magnusson et al., 1999; Park & Chen, 2012). When Shulman first conceptualized the idea of PCK (1986; 1987) he described it as composed of both a special knowledge of students’ understanding and of how to organize and represent knowledge to connect to that understanding. Since then, a number of researchers have built on that to attempt to model a component breakdown of PCK and a general assembling has occurred around five components: a) orientation to teaching science; b) knowledge of instructional strategies for teaching science; c) knowledge of students’ understanding in science; d) knowledge of science curriculum; and e) knowledge of assessment of science learning (Grossman, 1990; Magnusson et al., 1999; Park & Oliver, 2008). Additionally, the PCK a teacher has about teaching is different than the PCK applied in the classroom while enacting a lesson in context, and while these are both personal knowledge bases, there is also a collective PCK held by groups of teachers and researchers of different sizes such as of teachers in a learning community, or at the same school, or across the field as a whole (Ball et al., 2008; Gess-Newsome, 2015). These different levels of PCK, enacted, personal, and collective, each break down into the five components of PCK, and each of these needs to be considered separately when designing for PCK development and assessing the outcomes (Park, 2019). The proposed research will attempt to address a subset of these, but in order to understand what they each look like for data literacy, it is important to first understand how they have been conceived and measured by others. As PCK for data literacy has not been explicitly studied, I turn to the limited literature on STEM integration more broadly as
well as literature on PCK of each of the disciplines encased in data literacy: science, mathematics, and technology all of which are discussed in the next section.

2.2.1 PCK of STEM Integration

Throughout the literature on STEM integration, there are implicit, and even explicit but vague (e.g., Bryan et al., 2016; El-Deghaidy et al., 2017; NRC, 2014), references to PCK, particularly the components of student understanding (e.g., Kelley & Knowles, 2016) and instructional strategies (e.g., Asghar et al., 2012). However explicit research on PCK for STEM integration is sparse (Aydin-Gunbatar, 2020; NRC, 2014). The NRC report on STEM Integration (2014) notes that “implementation of integrated STEM experiences will in many cases demand educator expertise beyond that needed to teach any of the STEM subjects individually. Thus, many educators will need additional content and PCK in disciplinary areas beyond their previous education or experience” (p. 129). This observation led the report to recommend an increased focus on opportunities for teachers to gain experience identifying connections between the STEM disciplines and determining how to make those connections explicit to their students. However, in the half decade since the report came out, though there has been an increase in research on professional development for STEM integration, little of it has specifically focused on PCK (Aydin-Gunbatar et al., 2020; Margot & Kettler, 2019). In a study that looked broadly at the contextual factors affecting STEM integration in schools through focus group interviews with teachers, El-Deghaidy and colleagues (2017) found that the need for additional PCK was one of the major factors affecting implementation of STEM-
integrated experiences. Two recent studies (Aydin-Gunbatar et al., 2020; Vossen et al., 2020) dug deeper into PCK for STEM integration.

Vossen and colleagues (2020) conducted a case study of six teachers engaging in a professional learning community (PLC) aimed at developing their PCK for teaching Research and Design courses that engage students in project-based learning to address problems in their school’s community. They focused primarily on the components of knowledge of instructional strategies and representations, knowledge of student understandings, and knowledge of learning goals, which falls within knowledge of curriculum. For the learning goals, they found that an important focus was the goal of making connections between different topics across the disciplines, which some but not all teachers were able to identify as a goal. Determining how to make those connections explicit for students was a primary focus of the PLC and a knowledge of instructional strategies that is unique to STEM-integrated teaching and an important piece of PCK for teachers to have when implementing. Vossen and colleagues found that in relation to knowledge of student understanding, there was some work that teachers needed to do around familiarization with student misconceptions for research and design, as well as shifting perceptions on student interest and motivation, as some teachers came into the PLC experience believing that students generally find research boring and unmotivating.

Vossen and colleagues engaged teachers with collective PCK development through the PLC where they were working collaboratively to surface and make visible student understanding and strategies for teaching, they also measured teachers’ personal PCK. While the collective PCK was examined by examining the artifacts teachers created together, personal PCK was also measured through interviews to determine what each
teacher took away from the collaborative PLC work as their own personal understanding of the strategies and perception of student understanding. While they found that the teachers’ personal PCK was affected by the experience, they did not conduct classroom observations and so enacted PCK was excluded from their findings and they were unable to determine whether the shifts in personal PCK was transferred to changes in classroom enactment.

While Vossen and colleagues (2020) worked with in-service teachers, Aydin-Gunbatar and colleagues (2020) worked with a group of pre-service teachers. They focused on the same three components of PCK as Vossen and colleagues but with the addition of knowledge for assessment, noting that STEM integration often requires forms of assessment that teachers have less experience with as such as rubrics for project presentations and group work, and that good assessments for STEM integration don’t yet exist, a hole that others have noted as well (e.g., Asghar et al., 2012; Nadelson & Siefert, 2017; NRC, 2014). Much of the PCK Aydin-Gunbatar and colleagues focused on was also directed towards making connections between the topics in different disciplines. Some examples of PCK they highlighted: a) setting goals that included more than one discipline; b) awareness of real-world problems related to the content; c) how to use mathematics and technology in design problems based on science content knowledge; and d) use of alternative assessments to assess student learning on design concepts as well as science content. Additionally, since the teachers in their study were not engaged in a practicum, the knowledge building for PCK was purely theoretical with no way to connect the personal PCK being developed to enacted PCK in the classroom. The connection between knowledge about teaching (personal PCK) and knowledge in
teaching (enacted PCK) is one that researchers of PCK across domains are grappling with (Alonzo et al., 2019).

These two very recent studies show that there is PCK specific to STEM integration that goes beyond PCK in each of the STEM fields, which mostly focuses on the connections between the different disciplines, what they are, what students’ misconceptions and motivations towards them are, how to make them explicit to students, and how to assess them. However, they are the only studies which explicitly examine PCK for STEM integration and they do not put any emphasis on the different levels of PCK and the interactions between them. In order to build a more robust understanding of PCK at all levels and across all components, I turn next to research within each of the individual fields of science, technology, and math education. The field of science education has developed a model for PCK that will help highlight how PCK, and its levels, fit into the larger scope of teacher knowledge (Figure 2, discussed in more detail below). The field of technology education has been doing significant research on how to develop PCK for integration of one of the STEM fields into existing secondary school classrooms and may have some lessons for PCK for more robust STEM integration. Finally, the field of mathematics education, and specifically statistics education has some lessons for how to conceptualize PCK for working with real-world data. These are all discussed further below.

2.2.1.1 PCK in Science: The Refined Consensus Model. In 2012, a group of science education researchers with a strong focus on PCK came together for a week-long summit to build consensus around a model for PCK. There they developed an operational definition of PCK as “the knowledge of, reasoning behind, and enactment of the teaching
of particular topics in a particular way with particular students for particular reasons for enhanced student outcomes” (Carlson et al., 2015, p. 24) and a since widely-cited consensus model (CM) of PCK (Gess-Newsome, 2015). This definition defines PCK as containing both knowledge and skills, sets the grain size at the topic level and grounds it firmly in context. The CM also attempts to address how PCK fits into the larger model of teacher professional knowledge but both the nature of PCK as personal or collective, and the components of PCK were left unaddressed. A few years later, in 2016, unsatisfied with the CM after using it in the field and finding that many who were citing it were doing so only superficially (Chan & Hume, 2019) a second summit was held and the Refined Consensus Model (RCM) of PCK in science education (Figure 2) was developed (Carlson & Daehler, 2019).

Though the authors (Carlson & Daehler, 2019) acknowledge its lack of detail on the components of PCK, others (e.g., Park, 2019) have begun the work of connecting the RCM to models of PCK components, and the model provides useful insight into the levels (collective, personal, and enacted) of PCK and the way in which knowledge flows between them. By placing enacted PCK at the center, this model is accentuating the importance of the practitioner and the actual teaching that occurs in classrooms. While enacting teaching in the classroom, a teacher draws on knowledge from all the outer circles of the model to apply it to practice. Additionally, it is the PCK that is enacted in the classroom that informs and grows all the forms of PCK in the circles around it. Importantly, the RCM also places student outcomes at the center of the model, with the student icons representing both the students themselves and the outcomes of the teacher’s interactions with them. Student outcomes are perhaps the most important way to
determine the effectiveness of a teacher’s practice, and there is a strong need for more research into how teachers’ PCK affects student outcomes (Baumert et al., 2010; Chan & Hume, 2019; Gess-Newsome, 2015; Kirschner et al., 2016; Roth et al., 2019; Shulman, 2015; Smith & Banilower, 2015). By placing student outcomes at the center, the RCM is calling on the field to increase focus on this important aspect.

Figure 2
The Refined Consensus Model (RCM)

Note: From the 2nd PCK Summit (Carlson & Daehler, 2019)

The RCM makes a distinction between the levels of personal PCK and enacted PCK in a way that helps to clarify conflicting definitions of PCK. Personal PCK is all the knowledge a particular teacher has about teaching a particular topic within a particular learning context, yet on any given day while enacting a specific lesson with a specific
group of students, they won’t use every pedagogical strategy they know, they will pick and choose which of the strategies they know will work best in the given situation. The greater a teacher’s personal PCK is, the larger the toolbox they have to draw from and the stronger their enacted PCK will be. Likewise, enacted PCK can inform and grow personal PCK as new strategies are applied in new scenarios, or old ones elicit different learning responses or outcomes in the classroom. Personal PCK also draws from conversations with colleagues, educational preparation, and other professional learning experiences which comprise different groupings holding different collective PCK.

Enacted and personal PCK both exist inside the learning context circle because all teaching and learning is situated within a context, a specific time and place, with specific learners (Cobb & Bowers, 1999; Matsko & Hammerness, 2014; Shulman, 2015). The learning context includes some things that may be within a teachers’ control but also many that are not, such as state policy, national testing standards, funding and resources, and individual student attributes. These compounding factors act as a filter between the collective PCK of the field and the personal PCK of an individual teacher.

The outer layers of the RCM represent collective knowledge. The collective PCK is the knowledge that is held by groups of people that the teacher may have access to whether that is a learning community they are a part of or the literature on PCK and shared by those in the field doing research and working in teacher education and development they may have encountered during their pre-service training or continuing PD (Carlson & Daehler, 2019). While collective PCK can be applied to any collective of researchers or practitioners those collectives will have different levels of access to each other, so the PCK held by a learning community of practitioners might not become part
of the literature unless a researcher is working with them, and the PCK studied by the research field might not be accessible to a group of practicing teachers. All of these levels of and access to different sources of collective knowledge will affect personal PCK differently. This is one reason why partnerships between researchers and practitioners that collectively build PCK, such as this study attempts to do, are important for building bridges between PCK held by teachers and that identified by researchers in the field.

Additionally, it is here in the RCM model that grain size is addressed, though the authors explain that grain size exists at every level and was only included once for clarity. The inclusion of discipline, topic, and concept as potentials for grain size acknowledges the wide variations in the grain size of existing PCK research. Others (Chan & Hume, 2019; Mavhunga, 2020) have noted that a variety of grain size may actually benefit the field’s understanding of PCK as long as researchers are clear about what grain size they are studying. Finally, the outermost circle of the RCM holds the other professional knowledge bases that inform and interact with PCK. Content knowledge is the largest of these as it has been shown to have a strong influence on PCK (Aydin et al., 2015; Baumert et al., 2010; Daehler et al., 2015; Gess-Newsome et al., 2019; Kind & Chan, 2019), but also included are some of the other knowledge bases mentioned by Shulman in his original conceptualization (1987) and shown over the years to interact with PCK (Grossman 1990; Loughran et al., 2006; Magnusson et al., 1999; Park & Oliver 2008).

The RCM presents an external view of PCK, how it fits into the larger picture of teacher professional knowledge, and how knowledge flows between the collective PCK, a teacher’s personal PCK, and what is enacted in the classroom. As a model, it is useful to this research project because the project is situated within science classrooms and the
participants in this research project have training and experience as science teachers. However, the RCM has some shortcomings, most noticeably that it doesn’t address the components of PCK, but also, as has been noted by some others, the knowledge bases in the outermost layer that are depicted as informing collective PCK also play a role in PCK at the personal and enacted levels (Alonzo et al., 2019). Additionally, though the RCM tries to be discipline agnostic, in the field of science education research on PCK, subject matter knowledge always refers to a science topic (Chan & Hume, 2019). In this current study, the subject matter knowledge on which PCK draws is content knowledge of data and how it connects the STEM fields. As such, it is important to draw on other fields to understand how additional disciplines can be integrated into existing secondary school classrooms. This is an area in which the field of technology education has produced some important insights.

2.2.1.2 PCK in Technology: Technological Pedagogical and Content Knowledge (TPACK). The research on what would become TPACK started at the beginning of the century as technology was becoming more ubiquitous in classrooms and labs, especially at the higher education level (Angeli & Valanides, 2015; Mishra & Koehler, 2006). Over the course of the next decade or so, TPACK evolved into a Venn diagram framework that combines three kinds of overlapping knowledge, technology, pedagogy, and content, with technological pedagogical content knowledge at the center (Koehler & Mishra, 2008). An important aspect of the TPACK model that is relevant to the conversation about PCK for data literacy and STEM integration is that it acknowledges the existence of both technological content knowledge and technological pedagogical knowledge; there is knowledge about the content of technology that teachers
must have as well as knowledge about how to use technology to teach in the classroom (Mishra & Koehler, 2006; Niess, 2015). The acknowledgement that there is both content and pedagogical knowledge in the other three disciplines has been more accepted for much longer, but TPACK adds to the STEM integration conversation by bringing technology up to an even playing field with the other three disciplines. It also draws parallels for thinking about data and how there is both content knowledge and PCK for data literacy.

There has even been some work done recently to merge research on TPACK with research on STEM integration (e.g., Chai et al., 2019; Rahman, Krishnan, & Kapila, 2017). Chai and colleagues (2019) highlight that both STEM integration and TPACK focus on the development of 21st century skills and the use authentic real-world problems to encourage critical thinking and problem solving and note that there are many overlapping teacher competencies between the two frameworks. Additionally, there has been some research done in the TPACK community on the components of PCK as they apply to technology (e.g., Koehler & Mishra, 2008; Neiss, 2015). Neiss (2015) takes a particular technology (spreadsheet software) and identifies what the PCK components would be for four of the five PCK components. Generalizing the specific description to a more general technology approach gives: Orientations to teaching as the belief that the technology can enhance learning of content; knowledge of student understandings, the technology can deepen students’ understanding of the content by engaging students and offering new representations of content; knowledge of curriculum involves teachers viewing the technology as integral to the curriculum rather than an add on, and rearranging the curriculum to take full advantage of it; and knowledge of instructional
strategies, scaffolding use of technology to minimize the impact of challenges, and using the technology to engage in high-level thinking with students. This breakdown has additional parallels to framing an understanding of PCK for data literacy, as an important component of data literacy is the use of technology in order to work with big, messy datasets (Dorsey & Finzer, 2017; Finzer, 2013; Hardy et al., 2020). However, previous research on framing, developing, and measuring TPACK has not focused on data but rather on the tools used to collect, analysis, visualize, and communicate it. To better understand how PCK of data literacy is conceptualized, I turn next to statistics education, the field on which much of the literature on data literacy is based.

2.2.1.3 PCK in Math: Lessons from Statistics. A lot of research has been conducted on PCK for mathematics, and similar to the field of science education, math education has developed their own models for PCK (Ball et al., 2008; Depaepe et al., 2013). However, for the purpose of this research, I will focus primarily on research conducted in statistics education which, though often considered its own specific domain at higher education levels, is usually contained within broader mathematics courses at the secondary level (Chick & Pierce, 2012; Zieffler et al., 2018). Research on PCK for statistic is still an emerging field with limited findings (Callingham et al., 2016; Depaepe et al., 2013). In a study that is particularly relevant to this research Ijeh and Onwu (2013) studied expert secondary school mathematics teachers to understand their PCK for teaching data representations. They focused on knowledge of students’ understanding, specifically student misconceptions, and knowledge of instructional strategies, specifically representations of content. They found that in order to teach effectively about data visualizations, the expert teachers used a mix of procedural and conceptual
approaches to teaching. The development of data visualizations was taught in a very procedural way by walking students step by step how to create a graph. However, the interpretation of visualizations was approached more holistically with a focus on the role of the data in creating an interpretation of the data and how that visualization can convey new information about the data set.

Though focused on creating and interpreting data visualizations, a primary component of data literacy, the teachers in the study by Ijeh and Onwu (2013) were focused primarily on the skill set of creating the visualizations accurately and interpreting the mathematical relationships they conveyed mostly devoid from meaningful context, so the visualizations were of data that was provided to them and had no significance or meaning for the students such as the age distribution on an unidentified netball team.

Though there is beginning to be some research into using real-world data in the context of statistics education (e.g., Chick & Pierce, 2012; Sole & Weinberg, 2017) this research has not yet intersected with research on PCK in statistics education. Much of the literature on data literacy has drawn from the fields of statistics literacy and statistics education (Rubin, 2020). The difference is that data literacy takes the concepts of statistics education and using the tools of technology education, applies them to real-world data sets that are often embedded in interdisciplinary science content such as that of bioinformatics. In the next section I present a model that uses PCK for science education as a base but pulls in technology and mathematics as well to frame the as yet unexplored concept of PCK for data literacy.
2.2.2 PCK of Data Literacy.

There is little to no existing research on PCK for data literacy. In an attempt to frame a conversation about PCK for data literacy in STEM classroom settings, I would like to present a theoretical model that builds on the RCM for PCK in science education by bringing in the role of data literacy as a form of STEM integration (See Figure 3). The RCM places subject matter knowledge in the outermost ring of PCK, acknowledging its role in influencing and framing all forms of PCK. In my adapted model, I have taken the subject matter knowledge arc and pulled it out to examine it further, showing that in order to engage students in developing data literacy skills, teachers need to have subject matter knowledge in multiple disciplines. Since, for this project, the STEM-integrated experience will take place in a science classroom, conducted by science teachers, scientific knowledge is still the largest component. However, teachers also need to have mathematical and technological subject matter knowledge beyond what is needed to teach science alone. The scientific, mathematical, and technological knowledge all interact and integrate to form data literacy, which is then pulled through every layer of the RCM, affecting PCK at each level, into the center where classroom enactment needs to draw on teachers’ PCK for data literacy. For example, in order to create a scientific argument about weather patterns from a large messy data set, one needs to have knowledge about the science of weather and what variables effect it, but one also needs to have knowledge of the technology used to collect the data and of a statistical analysis software. Additionally, one needs enough statistical knowledge and skills to conduct the data analysis, identify measures of aggregate and the effects of outliers. All of these different subject matter knowledge components then are combined with pedagogical
knowledge and PCK from those different domains to determine how to teach data literacy through building a scientific argument about weather patterns from a large messy data set. Engineering knowledge has been left out of this model on purpose because the particular intervention being studied in this paper uses a problem based learning approach that is not grounded in engineering principles but rather focuses on the integration of science, technology, and math.

**Figure 3**
*Theoretical Model for PCK of Data Literacy*

*Note:* Adapted from the Refined Consensus Model of PCK for Science Education (Carlson & Daehler, 2019)

This model serves to highlight the fact that most research on PCK has remained siloed into separate disciplines of science, technology and math education with content within those disciplines serving as the subject matter knowledge on which PCK was built.
(Chan & Hume, 2019). However, in the case of data literacy, which is a form of STEM integration, the subject matter knowledge needed in the development of PCK is not siloed but draws from all three disciplines. This model is the foundation of the research I will be conducting in this study and will hopefully be a tool for future research to frame PCK for data literacy. The model will be discussed further in Chapter 3 of this paper.

While most teachers have extensive personal PCK for subjects they received training in and have experience teaching, they will need to develop additional PCK for STEM integration and data literacy in order to enact these successfully in the classroom. While PCK can develop through experience alone, studies have shown that PCK development can be guided and enhanced through high-quality professional development that aligns with teachers’ learning needs and goals (Aydin-Gunbatar et al., 2020; Gess-Newsome et al., 2019; Heller et al., 2012; Wilson et al., 2018). The existing bioinformatics study seeks to develop teachers’ subject matter knowledge and PCK for data literacy in STEM integration through engaging them in professional development that is grounded in literature on characteristics of successful PD. This study seeks to add an extra PD treatment to that existing program which focuses more specifically on how to develop PCK explicitly. However, since all the participants in the extra treatment will have gone through the existing bioinformatics PD, it is important to understand how that PD was designed in order to explore what PD components support teachers’ development of PCK, which is one of the research questions I will be asking in this study.
2.3 Characteristics of High Quality PD for Data Literacy in STEM Integration

Despite some large-scale studies that have shown mixed results for teacher professional development (Garet et al., 2011; TNTP, 2015), there has been extensive research at smaller scales showing that high-quality PD can improve teachers’ practice and have a positive impact on student outcomes (e.g., Desimone & Garet, 2015; Fischer et al., 2018; Lynch et al., 2019; Yang et al., 2020). Research on PD for teachers has begun to unearth some universal characteristics that lead to high quality professional development (Darling-Hammond et al., 2017; Desimone, 2009). In this section I will briefly outline those universal characteristics which formed the backbone of the design of the existing PD. Then I will discuss two areas where the characteristics that support successful PD are less well known and agreed upon: PD for STEM integration and data literacy and PD for PCK. Then in the following chapters I will outline the ways in which the existing bioinformatics PD and my planned extra treatment PD build on them.

2.3.1 Characteristics of High-Quality PD

In the last decade or so there has been increasing agreement around how to design for high-quality PD (Darling-Hammond et al., 2017; Desimone, 2009; Desimone & Garet, 2015; Fischer et al., 2018). In her seminal paper on characteristics which support high-quality PD, Desimone (2009) laid out five critical priorities for developing PD that will lead to improvement in teaching and learning: a) PD needs to be content focused; b) teachers should be engaged as active rather than passive learners; c) goals of the PD should be consistent with teacher and student needs; d) teachers should be engaged with PD over a sustained period of time, at least 20 hours of contact time over an entire school
year; and e) teachers should participate collectively with others with similar teaching parameters. Subsequent research (Desimone & Garet, 2015; Fischer et al., 2018; Yang et al., 2020) has shown the effectiveness of Desimone’s model on teacher practice and student outcomes. In a more recent report, Darling-Hammond and colleagues (2017) examined 35 research studies on effective PD for changing teaching practice and student outcomes and, with feedback from Desimone, updated the original list of characteristics. The new list includes a) a focus on disciplinary content, both the concepts and pedagogies; b) addressing how teachers learn through active learning and sense-making; c) enabling collaboration among teachers; d) using models of effective instruction; e) offering coaching and expert support; f) dedicated time for feedback and reflection on practice; and g) sustained duration of PD participation. Though some have criticized Desimone’s model for focusing too much on the teacher and not enough on external systems, (e.g., Boylan et al., 2018; Opfer & Pedder, 2011) the seven characteristics highlighted by Darling-Hammond and colleagues serve as a strong foundation for designing impactful high-quality PD including our team’s existing bioinformatics PD.

Though shown to be necessary for designing successful PD, these characteristics are generalized and additional characteristics that are specific to the subject matter and context, in our case STEM integration and data literacy, need to be considered.

2.3.2 PD for STEM Integration

As teachers, especially secondary school teachers, have often not been provided with training or experience in disciplines beyond their own it is important to provide teachers with PD to support them in STEM integration (Aslam et al., 2018; Brown &
Bogiages, 2019; Nadelson et al., 2013; Vossen et al., 2020). In their review of literature on teachers’ perceptions of STEM integration, Margot & Kettler (2019) found that additional PD was one of the main areas that teachers believed would help them in engaging in STEM-integrated teaching. Because STEM integration is complex and involves expanding teachers’ content knowledge, PCK, and confidence beyond their current discipline, research into PD for STEM integration has found some characteristics that have supported teacher growth that are specific to STEM integration that are modification or additions to the accepted characteristics of high-quality PD generally (Darling-Hammond et al., 2017; Desimone, 2009).

In a meta-review of PD for STEM integration, Lynch and colleagues (2019) found a number of characteristics of PDs that led to not only improved instruction, but also improved student outcomes. These were, a) included a balanced focus on both content knowledge and PCK; b) provided teachers with quality curriculum materials for STEM-integrated teaching; and c) extended meetings into the school year that provided space to troubleshoot and discuss implementation. These findings align with other research that has shown that teachers need both content knowledge and PCK for STEM-integrated teaching (El-Deghaidy et al., 2017; Heller et al., 2012; van Driel & Berry, 2017; Yang et al., 2020;) and that lack of strong resources for teaching STEM is a primary concern for teachers (Asghar et al., 2012; NRC, 2014; Margot & Kettler, 2019). Though dedicated time for feedback and reflecting on practice is one of the characteristics of high-quality PD highlighted by Darling-Hammond and colleagues (2017), that is often viewed as an end of implementation activity, while Lynch and colleagues (2019) highlight the importance of space and time for troubleshooting during
teaching, especially for STEM-integrated units that tend to extend for a few days or even weeks. Additionally, although *enabling collaboration among teachers* is another one of the key characteristics of high-quality PD already emphasized by Darling-Hammond and colleagues (2017), it is particularly important for PD focused on STEM integration. Vossen and colleagues (2020) found in a study on PD aimed at developing PCK for STEM integration that through a PLC, teachers were able to build a collective knowledge base of connections between content and design. Other researchers have also highlighted the importance of collaboration for developing knowledge, skills, and confidence in STEM integration (Aslam et al., 2018; Johnson & Sondergeld, 2016).

Other researchers have emphasized the importance of explicitly discussing student-focused teaching strategies during PD to help teachers develop skill and confidence in using them (Asghar et al., 2012; Brand, 2020; Nadelson et al., 2013). Asghar and colleagues (2012) specifically explored PD that focused on the use of PBL as a tool for STEM integration. They found that teachers entered the PD with some understanding of how PBL could be used in the classroom, but with a discipline centered approach. By the end of the PD, teachers were able to describe how PBL could be used for STEM integration and cited that they felt more comfortable engaging in open-ended student-centered learning. However, Asghar and colleagues (2012) found that teachers still had trouble making connections between the content from different disciplines and between the content from the problem in the PBL to their larger disciplines. This aligns with other research that emphasizes the need to make connections between topics across disciplines explicit for teachers during PD (Aydin-Gunbatar, 2020; Cavlazoglu, B. & Stuessy, 2017; English, 2016). Additionally, some studies have shown that teachers do
not always come into PD with a prior understanding of STEM integration that aligns with the PD and/or intervention (Burrows & Slater, 2015; El-Deghaidy et al., 2017) which means that explicitly defining STEM integration to include those connections between disciplines can help increase teachers’ knowledge development.

Another unique area of focus for PD for STEM integration is the inclusion of technology. Technology use, as well as technological knowledge around its role in society is an important part of STEM integration yet technology integration often contrasts with the beliefs, knowledge, and skills of teachers in other disciplines (Ertmer et al., 2012; Tondeur et al., 2017). In a review of PD for integrating technology in science classrooms using inquiry-based learning, Gerard and colleagues (2011) found that teachers encountered many challenges with implementing technology, including alignment with curriculum and time to modify interventions and determine how new technology fits into their curriculum and classroom. They found that the most successful features of PD for technology integration were to a) provide teachers with extended time to not only learn how to use technology, but also how to fit it into their curriculum; b) access to mentoring and collaboration; and c) the need to ground learning in the context of the school and classroom technology capabilities such as speed of internet connections, access to computers, budget and funding, and even access to outlets. While the first two findings align with Darling-Hammond and colleagues’ characteristics of high-quality PD, the third is unique to technology integration and has been found by other researchers to be a vital component of PD to encourage teachers to bring technology into their classrooms (Lee et al., 2017; Hubers et al., 2020). Lee & Wilkerson
(2018) also addressed the need to support teachers in working with technology in their report on data use in middle school and high school as it relates to data literacy.

2.3.2.1 PD for Data Literacy. There has been little to no empirical research published on PD specifically for developing data literacy. The theoretical research on potential strategies for teaching data literacy (e.g., Lee & Wilkerson) have been developed without engaging teachers with those strategies during PD. Though there has been research on PD for statistics education (e.g., Meletiou-Mavrotheris et al., 2009; Schoen et al., 2019; Souza et al., 2015) those publications have little to add beyond what has been discussed and studied in the literature on PD for STEM and technology integration. As data literacy is an example of STEM integration, much of the same findings on PD for STEM integration apply to PD for data literacy as well. Lee and Wilkerson (2018) in their review of research on data science link PD for data literacy and technology integration in their report, offering four theoretical suggestions for supporting teachers in working with emerging technologies for data use: a) guide teachers through an entire cycle of inquiry with the new technology, allowing them to experience many different phases of working with data, including processing the complexity of data sets; b) when working with models or simulations, connections between the data produced by the simulation and content topics need to be made explicit for teachers, and they should be provided with examples of explanations and arguments that connect the data from the simulations to the underlying concepts; c) encourage teachers through PD to go beyond noticing student engagement, as technology for data collection and analysis can be exciting for students, to consider their conceptual thinking as well; and d) developing teachers’ comfort levels with large complex datasets such as public databases. Lee and
Wilkerson (2018) also emphasize the importance of helping teachers “to “step back” and understand these resources as sources of information, rather than communications of objective truth” (p. 30) which is a vital piece of developing data literacy and the ability to be critical of data.

Along with work being done in STEM and technology integration to foster development of stronger data literacy, there are a number of groups that have been focused on creating resources for teachers in developing data literacy (e.g., Concord Consortium, Oceans of Data, Data Nuggets). The team at the Concord Consortium, through an NSF grant, developed the Common Online Data Analysis Platform (CODAP) which provides free tools for teachers and students to engage with complex data in interactive ways (codap.concord.org). In the past few years, they have begun research into how students learn and develop data literacy through using CODAP with complex data sets (Deitrick et al., 2017; Erikson et al., 2019; Finzer et al., 2018; Wilkerson et al., 2018) and through their research on how students learn have developed a set of suggested “data moves” for teachers to focus on when teaching students to work with data (Erikson et al., 2019). However, these data moves have not been empirically studied. The Oceans of Data Institute (oceansofdata.org) has also developed resources for students and teachers looking to engage with complex data and has conducted research on how students work with big data and develop data literacy (Busey et al., 2015; Kochevar et al., 2015; Sickler et al., 2021). Though both the Concord Consortium and the Oceans of Data Institute offer small scale PD for teachers, neither has made it a primary focus of resource development nor of research on teacher knowledge or learning. As such there is a need for more research into understanding how to support teachers in addressing data literacy.
in their classrooms. Specifically, there is a need to engage with teachers to develop an understanding PCK for data literacy and the ways in which teachers can engage with and develop it. This project hopes to do this, by drawing on existing literature for developing PCK through PD.

2.3.3 PD for PCK

Research into PD for developing pedagogical content knowledge has produced mixed results over the years (Chan & Hume, 2019; Gess-Newsome et al., 2019; Wilson et al., 2018). However, multiple studies have shown that when characteristics of high-quality PD are applied to the development of PCK, teacher knowledge (e.g., Aydin et al., 2013), instructional practices (e.g., Nilsson, 2014) and even student outcomes can be affected (e.g., Bayram-Jacobs et al., 2019). Among the characteristics of high-quality PD, enabling collaboration and dedicating specific time for reflection are two of the most important for the development of PCK (Donnelly & Hume, 2015; Wilson et al., 2018). Additionally, the focus on disciplinary content, must explicitly include teaching strategies along with the concepts (Heller et al., 2012; van Driel & Berry, 2017). Two common methods for engaging teachers in building and refining their PCK are: Content Representations (e.g., Loughran et al., 2012), and reflections on their own and colleagues’ teaching by observing video of teaching (e.g., Wilson et al., 2018), and extended reflections through both written assignments and interviews (e.g., Williams & Lockley, 2012).

2.3.3.1 Content Representations (CoRes). CoRes were originally developed by Loughran and colleagues (2006) as a way to guide teachers to think about their practice
in creating collective PCK around a particular big idea (Carpendale & Hume, 2019; Cooper et al., 2015; Loughran et al., 2012). A CoRe is a template (See Appendix A) which prompts a teacher or group of teachers to think about teaching a particular subject 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? (See Figure 4 for an example completed CoRe from Loughran and colleagues’ work).

Figure 4
Example of a Completed CoRe for Genetics

<table>
<thead>
<tr>
<th>IMPORTANT SCIENCE</th>
<th>IDEAS/CONCEPTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Genetic information are passed from parent to offspring.</td>
<td>C: The total amount of variation within a species is relatively stable.</td>
</tr>
<tr>
<td>B: There is variation between individuals within a species.</td>
<td>D: An individual’s characteristics are determined by interaction between genetic information and the individual’s environment.</td>
</tr>
<tr>
<td>Evolution in a process that results in heritable changes in a population spread over many generations.</td>
<td></td>
</tr>
</tbody>
</table>

**Note:** From Loughran, Berry & Mulhall (2012).

Loughran and colleagues (Cooper et al., 2015; Loughran et al., 2012) found that not only were CoRes helpful tools for assessing teachers’ PCK, they could also be used to
develop PCK by enabling teachers to make their practice, and their thinking on their practice visible in a way that allowed them to then reflect on that knowledge. Loughran and colleagues (2012) describe CoRes in the following way:

A CoRe (Content Representation) provides an overview of how a given group of teachers conceptualize the content of particular subject matter or topic. A CoRe is developed by asking teachers to think about what they consider to be the “big ideas” associated with teaching a given topic for a particular grade level(s) based on their experience of teaching that topic. These big ideas are discussed and refined and then, when generally agreed upon, become the horizontal axis of a CoRe [see Appendix B for the template]. The big ideas are then probed and quizzed in different ways through the prompts that are listed on the left hand side vertical axis of the CoRe, so that specific information about the big ideas that impact on the manner in which the content is taught can be made explicit. Through this process, the CoRe becomes a generalizable form of the participant teachers’ pedagogical content knowledge as it links the how, why, and what of the content to be taught with what they agree to be important in shaping students’ learning and teachers’ teaching. (p. 17)

CoRes have been shown to guide the flow of knowledge between collective, personal, and enacted PCK within the communities in which they are used. While teachers often struggle to construct big ideas for a subject and answer the questions in a CoRe individually, when building knowledge collaboratively they have more success (Aydin-Gunbatar et al., 2020; Carpendale & Hume, 2019; Cooper et al., 2015). 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 (Loughran et al., 2012). This collective PCK then influences their personal and enacted PCK which in turn can be brought back to the group to contribute to additional collective PCK (Loughran et al., 2012; Vossen et al., 2020). Next, I discuss three studies that used CoRes as part of a
teacher learning experience in order to understand design details for using CoRes as part of PD.

In recent research conducted by Carpendale and Hume (2019), CoRes were used as the central activity in a workshop aimed at developing teachers’ PCK. The study participants were colleagues at a secondary school in which science was taught as an integrated discipline who were interested in improving PCK for physics, specifically electricity and magnetism. The CoRe design workshop consisted of two three-hour sessions for a total of six hours of intervention and placed teachers into heterogeneous groups based on their comfort level with topic. The first workshop session introduced teachers to the construct of PCK and the purpose of a CoRe in capturing and developing PCK. Then teachers spent most of the remaining two hours developing a CoRe in small groups on the nature of science, which was deemed to be a topic that teachers had strong content knowledge on, while the facilitator moved from group to group asking probing questions to encourage deep thought. At the end of the workshop, the three groups shared their CoRes and combined them into one larger CoRe. In the second workshop which occurred a week later, teachers again worked in small groups to develop a second CoRe, this time on electricity and magnetism. At the end of the workshop session, the CoRes were again combined into one CoRe with multiple big ideas. Teachers then taught a lesson based on the collaborative CoRe and were asked to reflect after the lesson on how they thought there personal and enacted PCK might have been affected by the workshops. Through classroom observations before and after the workshop and interviews, Carpendale and Hume (2019) found that the workshop enhanced teachers’ personal PCK
and enacted PCK, especially within the component of knowledge of student understanding and knowledge of teaching strategies.

This work was focused on developing teachers’ PCK for a specific subject within physics, and most research with CoRes has been conducted in PD for developing PCK in specific disciplines (e.g., Cooper et al., 2015 (physics); Aydin et al., 2013 (chemistry); Walen, Nilsson & Ewen, 2017 (primary school teachers). However, two very recent studies used CoRes in development of PCK for STEM integration (Aydin-Gunbatar et al., 2020; Vossen et al., 2020). In using CoRes to support PCK development for STEM integration, Aydin-Gunbatar and colleagues (2020) worked with a group of pre-service chemistry teachers (PSTs) who were enrolled in a semester-long elective STEM course. One of the goals of the course was to develop PCK for STEM integration. They had the PSTs complete a CoRe individually at the beginning of the course and again at the end. Throughout the semester, the PSTs engaged collectively with examples of “strong” and “weak” completed CoRes, discussing their strengths and weaknesses relative to the components of PCK. Additionally, during one of the course meetings, the PST created a CoRe collaboratively in small groups. While the subjects and big ideas used in all the CoRes were topics from chemistry, Aydin-Gunbatar and colleagues found that the combination of critiquing and creating CoRes while also engaging in STEM-integrated activities as learners, led to a strengthening of the PSTs PCK for STEM integration.

Vossen and colleagues (2020) worked with in-service teachers who were already teaching a STEM-integrated course called Research and Design for secondary school students. A primary focus of their research was to build a professional learning community (PLC) so all of the work with CoRes that the teachers engaged in was
collaborative. Vossen and colleagues (2020) were also specifically interested in developing teachers’ PCK for the connections between research (primarily in science) and design (in engineering) by focusing on the five components of PCK (Magnusson et al., 1999; Park & Chen, 2012). Their intervention consisted of four three-hour workshops over the course of five months. The first workshop meeting focused on clarifying the role of the Research and Design course and enhancing teachers’ content knowledge for teaching it. In the second session the teachers were introduced to the concept of PCK and its five components, then collaboratively built a CoRe on connections between research and design through “a group discussion structured by the facilitator” (p. 304). In between sessions 2 and 4, participating teachers taught a lesson based on the CoRe and in sessions 3 and 4 they shared their experiences with implementation and made adjustments to the CoRe and to their lesson plans based on those discussions. At the end of the intervention, through recordings of the sessions and interviews that asked participants about their experience creating CoRes and teaching from them, Vossen et al (2020) found that teachers had grown in their personal PCK for connections between research and design.

Common themes across these interventions are a) the focus on collaborative development of CoRes; b) framing the work with CoRes by defining PCK and its components; c) having teachers use the CoRe as a foundation for teaching; d) reflecting on changes in their personal and enacted PCK as a result of using the CoRe; and e) facilitating knowledge sharing between collective, personal, and enacted PCK through iterative design, enactment, and reflection. These themes will provide a design framework on which the extra treatment for this study will be built.
In a brief, non-systematic review of research utilizing the CoRe framework and template, Lehane and Bertram (2016) found many examples of CoRes used in pre-service teacher training or in-service PD to successfully develop participants PCK. While the CoRe template is designed to surface and develop PCK and assumes a relative complete and high level of subject matter knowledge, there were some studies highlighted by Lehane and Bertram (2016) which showed that the process of developing CoRes collectively also supported growth in teachers’ subject matter knowledge (e.g., Eames et al., 2011 cited in Lehane & Bertram, 2016). Loughran and colleagues (2006) also discussed this potential use of CoRes in their original book on the practice. The process of completing a CoRe requires teachers to problematize the content, which often leads to them examining their content knowledge and further developing their understanding of it (Loughran et al., 2006, 2012). Additionally, when working collaboratively to build a CoRe, teachers must engage in conversation of the big ideas of a topic. This conversation can lead to teachers pooling their content knowledge and therefore learning from each other (e.g., Eames et al., 2011, Vossen et al., 2020). So, while CoRes are primarily a tool for developing PCK, as PCK is intertwined with subject matter knowledge, I hypothesized that the process of completing CoRes would support development of teachers’ subject matter knowledge as well as their PCK for data literacy.

2.3.3.2 Reflecting on Practice. The other primary technique used to develop PCK during PD usually aims to engage with teachers’ enacted PCK through the use of reflections to prompt teachers to examine and make visible their thinking while enacting a lesson. In a review of literature on PCK interventions Evens and colleagues (2015) found that opportunities for reflection was one of the primary effective characteristics for
PCK interventions, and it is also one of the characteristics of high-quality PD generally (Darling-Hammond et al., 2017). However, some have noted (Monet & Etkina, 2008; van Driel & Berry, 2017) that not all reflection is created equal and that teachers need modeling on how to reflect effectively in a way that engages their higher-order thinking. The decisions that teachers make in the moment while teaching are fast-paced and often complicated, in order to capture their reasoning behind those decisions, some PD facilitators and researchers have relied on video of lessons to trigger teachers’ memory of the enacted lesson (Nilsson, 2014; Roth et al., 2011; Wilson et al., 2018). Others have used notes from classroom observations to stimulate recall of the lesson (Bravo & Cofre, 2016; Vossen et al., 2020; Williams & Lockley, 2012) through interviews.

One PD framework for working with videos to guide reflection and analysis is Science Teachers Learning Through Lesson Analysis (STeLLA) developed by Roth and colleagues (2011). The STeLLA approach to teacher learning is built on three essential features: a) integrating PCK development with content knowledge development; b) scaffolding teacher learning to lead teachers away from the need for strong supports towards an increased independence; and c) the use of classroom videos that are context specific for participating teachers (Wilson et al., 2018). It seeks to develop teachers’ PCK for two components: knowledge of students’ understanding of science and science content storylines, which is a combination of knowledge of teaching strategies and knowledge of curriculum (Roth et al., 2011). The STeLLA framework is usually implemented as a year-long PD in which teachers engage in a cycle of planning, enacting, and reflecting using classroom videos to highlight certain aspects of the two PCK components. At the beginning of the PD, teachers are introduced to PCK as a concept, the
two components, and indicators for those components. Then teachers are shown a clip of
an expert teacher exhibiting at least one of indicators, for example “using representations
that are matched to the learning goal,” and are guided in a discussion reflecting on the
video using questions such as “what evidence is there that the representation was helpful
or confusing to student?” (Roth et al., 2011). After a few cycles of watching videos and
reflecting, teachers are then provided with a lesson plan which they teach and video
record. They then go through a few more cycles of watching and reflecting on these
videos before designing their own lessons and going through the cycle again. In this way
teachers are scaffolded into greater and greater independence (Wilson et al., 2018).
Reporting on two separate studies which used the STeLLA method, one a randomized
cluster trial of in-service elementary school teachers, and the other a multi-year quasi
experimental study of pre-service science teachers, Wilson and colleagues (2018) showed
that not only did the treatment teachers exhibit significantly more growth in their PCK
than those in the control group, but student content knowledge was significantly impacted
in the treatment conditions. The use of videos as a trigger for reflection will be a central
component of the extra treatment for teachers in this study and will be combined with the
development of CoRes and the use of additional written reflections to capture teacher
thinking in between meetings of the workshop.

Written journal reflections have also shown to be effective at developing PCK,
especially when combined with CoRes (Aydin et al., 2013; Aydin-Gunbatar et al., 2020;
Evens et al., 2015; Williams & Lockley, 2012). Aydin-Gunbatar and colleagues (2020)
used written journal reflections in combination with both individual and collective CoRe
critique and development with prompts such as “During this lesson what were the
moments of STEM integration? Were there other moments where integration could have been made explicit?” They found that these reflections helped teachers solidify at a personal level the growth and learning that was happening collectively. Reflection questions work best when they challenge teachers to think more deeply about the lesson than surface level questions such as “what went well?” (van Driel & Berry, 2017). Henze and van Driel (2015) tried to guide teachers into higher-order reflecting by prompting teachers to provide evidence for their reflections, using questions such as, “What was successful for your students? How do you know? Explain your answer.” They found that this encouraged teachers to reflect in a deeper way and prompted them to become more aware of their own PCK. Though reflection questions on their own do not always elicit PCK development, as we found in analysis of the existing bioinformatics PD which relied primarily on this method for engaging teachers’ PCK (Miller at al., 2021), when combined with more robust strategies such as CoRe development and video analysis, they can enhance the learning that occurs (Aydin-Gunbatar et al., 2020; van Driel & Berry, 2017).

Using these strategies, and the frameworks discussed previously in this chapter, this research seeks to study the as yet un-studied construct of PCK for data literacy. In the next chapter, I lay out the framework for my study and explain in more detail how each piece will be conceptualized and measured.
CHAPTER 3. CONCEPTUAL FRAMEWORK

Professional development can affect teachers’ collective, personal, and enacted PCK which depending on their subject matter knowledge, both of which can in turn have an impact on student outcomes (Figure 5). This research project seeks to understand how the design of a PD can affect teachers’ subject matter knowledge PCK for data literacy which is inherently reliant on their subject matter knowledge of data literacy. The construct of student outcomes, though a vital part of the larger project, will not be explored in this study. Each of the other constructs that is addressed in this research is defined in greater detail below.

Figure 5
Conceptual Model for PD for PCK for Data Literacy in STEM Integration

Note: The blue text denotes areas of the existing bioinformatics PD that were modified for the 2021-2022 iteration in alignment with this research.
3.1 The Bioinformatics PD

The PD being studied in this research drew on design principles from three areas of literature: high-quality PD, PD for STEM integration and data literacy, and PD for PCK. The existing bioinformatics PD was designed by our team based on literature for general characteristics of high-quality PD and characteristics for PD for STEM integration and data literacy. The general characteristics of high-quality PD are well establish, and were outlined in section 2.3.1, therefore they will not be discussed again here. However, since the characteristics of successful PD for STEM integration and data literacy are less well established by previous literacy and have not been succinctly outlined by other publications on the existing bioinformatics project, I will outline them here.

3.1.1 PD for STEM and Data Literacy

Based on existing literature on PD for STEM and data literacy described in section 2.3.2, I have distilled five characteristics that support successful PD for STEM and data literacy that were consistence across the literature. They are a) using a full cycle of inquiry; b) providing time to modify provided technology and curriculum; c) focusing on the context of teacher’s school and classroom; d) making connections between subjects and disciplines explicit; and e) attending to the complexity of data. In order to successfully engage teachers in active learning for STEM integration, PD should provide teachers time and support to go through an entire cycle of inquiry with the data and technology, from the initial introduction of the problem, through the development of questions and determining how best to answer them, through the collection, organization
and analysis of data, to the development of solutions and visualizations to communicate them. As adopting technology in the classroom is a barrier for some teachers, *time needs to be set aside during PD for teachers to modify curricula and tools* to fit their classroom and student needs. Additionally, the *context of the classroom* needs to be attended to during PD, including the structure of classes such as their length and frequency of meeting, and their technological capabilities such as access to outlets, computers, or reliable internet access. In order to successfully integrate topics and skills across the STEM disciplines, *connections between the disciplines need to be explicitly established and explored* through discussion and reflection. Finally, PD for STEM integration should *address the complexity and messiness* of data associated with real-world problems.

Some of these characteristics overlap with the seven of high-quality PD or build on them to make them more specific to STEM. Additionally, the existing PD already included some of these design characteristics. Specifically, using a full cycle of inquiry, providing time to modify, and focusing on context. However, despite efforts by the existing PD to make connections explicit and attend to the complexity of data, these were two areas in which teachers in the first two cohorts consistently struggled (Miller et al., 2021; Shim et al., 2021). As such, the summer PD was modified for the summer of 2021 to more explicitly focus on these two design characteristics and the extra treatment PD that is the focus of this study put extra focus on designing for them in order to better support teachers in these areas.
3.1.2 PD for PCK

The third and final area from the literature that the PD drew on is the development of PCK through PD. Specifically it focused on the use of content representations and higher order reflection. The CoRes used by the PD were modeled closely after those designed and popularized by Loughran and team (Loughran et al., 2012; Cooper et al., 2015) and discussed in section 2.3.3.1 and were added to both the existing summer PD and used heavily in the Extension Workshop Series. For the purpose of this research, higher-order reflections are defined as reflection activities that promote higher order thinking and require teachers to reason from evidence. Though the existing PD did include reflection as aligned with the characteristics of high-quality PD, it was not designed to elicit higher-order thinking or require teachers to reason from evidence in their classrooms or experience. The discussion forum prompts for the existing summer PD asked teachers to reflect on their teaching, but as it occurred during the summer, divorced from enactment in the classroom, teachers struggled to connect their reflections to evidence from their students. Additionally, analysis of interviews from the first cohort showed that teachers did not naturally supply evidence for reflections on their lessons and the existing post-observation protocol did not included a prompt for this. As such, the post-classroom observation reflection questions were extended and modified for the teachers in this study to elicit high-order reflecting. Additionally, video reflections, as discussed in section 2.3.3.2 were a central design feature of the extra treatment Extension Workshop in order to further higher-order reflecting through the use of an artifact that was used to trigger deeper analysis of teaching.
3.2 Subject Matter Knowledge of Data Literacy

While this study is focused on PCK for data literacy, as has been argued throughout the literature and in this paper, PCK and subject matter knowledge are closely inter-related and for teachers, dependent on one another. This paper hypothesized that through the process of developing CoRes to surface PCK for data literacy, teachers would also interrogate their own subject matter knowledge for data literacy as well as the big ideas of subject matter knowledge for their students. While little research currently exists about what teachers view as important for students to know and learn about data and data literacy, and there is still a lack of consensus in the field on what the goals of data literacy should be in K-12 education (Kjelvik & Schultheis, 2019) for the sake of this study, I frame the subject matter knowledge of data literacy based on the framework discussed in section 2.1.1 the existing bioinformatics PD explicitly addressed five components of subject matter knowledge of data literacy: a) context, b) variability, c) aggregate, d) visualization, e) inference, and was modified for last summer’s iteration to include a sixth: f) complexity of data. In this section I outline the indicators for mastery of these components highlighted by the literature. The summer PD introduced these components to the teachers and supported them in reflecting on their own subject matter knowledge for data literacy. The Extension Workshop Series used these components as the topics for which teachers discussed big ideas and developed CoRes and extended both their own subject matter knowledge, but also their PCK for the goals and strategies for teaching this particular subject matter knowledge.
3.2.1 Data in Context

*Context* is the who, when, where, what, why and how of data collection and use. It is understanding where data sets came from and the bias that might be ingrained, as well as why it’s important to be aware of the context and the bias it can carry. Indicators of mastery of data in context will include a) asking probing questions about the origins of external data; b) providing context for personally collected data; and c) considering how bias stemming from intention, existing theory, and the tools of data collection can affect data.

When confronted with data you did not collect yourself, it is important to ask probing questions about the origin (Rubin, 2020; Wise, 2020). Some examples of these types of questions that probe for context of the data and are outlined in the existing PD are:

- Who collected the data? Are they reliable? Are they experienced at data collection? Are they well-funded? Peer-reviewed?
- Who is the data about? Is the sample of participants used representative of a larger sample?
- What was the reason for the data collection? Was it to answer a specific question?
- When was the data collected? Not just year, but time of year, and time of day.
- What were the methods used to collect the data? If asking questions of people, what questions were asked? Were certain definitions used such as "popular" or "high-risk" that may be subjective?

Evidence of use of these types of questions when confronted with data of unknown origin is one indicator of mastery of *data in context* but another is addressing these questions when collecting your own data. In the current bioinformatic curriculum, students collect
data on air quality which is highly sensitive to context. Paying attention to, documenting, and communicating this context and how it might affect the data is another indicator of mastery of data in context.

Finally, as Hardy and colleagues (2020) discuss, data is shaped by the intention with which it was collected and analyzed, existing theories about the world, and the tools that are used to collect the data, and each of these are a potential source of bias in the data. An indicator of mastery of this component of data in context is any mention of bias as it relates to intention, theories, or tools, including confirmation bias. Mastery of Data in context can be measured on these three indicators to explain teachers’ understanding of this component of data literacy.

3.2.2 Variability in Data

All data sets contain variability. Understanding variability requires knowledge of potential error in collection instruments but also the factors involved in real-world scenarios that can cause fluctuations and variability and how to account for these. Indicators for mastery of variability include a) the ability to measure and describe variability through range, standard deviation, and other statistical tools such as quartiles; b) the ability to identify different sources of variability and c) the ability to discuss uncertainty by referring to those measures and sources of variability. As Rubin (2020) points out, probability is an exceptionally hard concept for most people to understand which makes talking about uncertainty difficult as well. She also highlights that data collected over time and across space brings with it the complication of spatial and temporal correlation which needs to be addressed when interpreting variability.
Additionally, Hardy and colleagues (2020) discuss the difference between the variability of data that stems from the phenomenon being studied, and the variability that comes from the tools and method of measurement. All of these play into the indicators of mastery for variability in data.

3.2.3 Data in Aggregate

In order to find patterns across a dataset, it needs to be viewed in the aggregate, allowing for measures of centrality and large-scale patterns to be determined. This component of data literacy can be difficult to master because there is a tendency to focus on individual data points, especially when the data is in some way personal (Wise, 2020). Konold and colleagues (2015) outline a number of different lens for viewing data that can help clarify data in aggregate. They describe the different views as depending on the fundamental unit into which the data is organized. Some lenses for looking at data use each individual data point as the fundamental unit, while others involve grouping data points by similar characteristics and having these groups serve as the fundamental unit.

Applying an aggregate lens to data treats the entire data set as the fundamental unit. This understanding of an aggregate lens is also supported by Lee and Wilkerson (2018), in identifying ways in which students often struggle with data in aggregate. Given this, the indicators for mastery of data in aggregate include a) calculating measures of centrality for the entire data set; and b) applying those measures of centrality to describe patterns and themes evidenced across the entire data set.
3.2.4 Data Visualization

Mastery of data visualization involves both interpretation of existing visualizations and the creation of new ones, often with the use of technologies built for doing so. For both interpreting and creating visualizations, paying attention to labels is important. As Hardy and colleagues (2020) note, something as simple as the scale of an axis can drastically alter the way a data visualization is perceived. In fact, manipulation of axes is one of the primary ways that data visualizations can be used in biased ways to affectively “lie” while still using “real” data. Additionally, mislabeling or even leaving off labels all together from axes and trend lines is another source of confusion in data visualizations (see badvisualizations.tumbler.com for an excellent compilation of some of these misused data visualizations). As such, being able to identify issues with labels when interpreting visualizations created by others, and properly including them when creating visualization is one indicator of mastery with this concept.

Another important indicator for mastery of the creation of data visualizations is the ability to choose the appropriate type of visualization for the data and the information you are trying to communicate. People are biased towards types of visualizations that they are more familiar with such as bar graphs and pie charts (Gebre & Polman, 2016). However, these are not always the best format for communicating your information, and sometimes certain types of visualizations can lead to wildly misleading representations of the data, which leads to another indicator, which is the ability to communicate effectively through data visualizations. Choosing a useful type of visualization and labelling it correctly and helpfully permits for communication to occur, but the text and context
surrounding a visualization also adds to its ability to communicate, through descriptions, trend lines, colorization choices, and highlights of certain aspects of the visualization.

Finally, the last indicator of mastery of data visualizations is the ability to draw an inference and build a scientific argument using a data visualization that you may not have created yourself. As Finzer and Reichman (2018) point out, extracting information from a data visualization can be complicated and puzzling and is connected to mastery of the ability to ask the right questions about context in order to identify deep relationships as well as potential bias.

3.2.5 Inference with Data

A primary indicator for mastery of inference with data is the ability to build a sound scientific argument using the claim, evidence, reasoning (CER) framework (McNeill & Krajcik, 2011). The CER framework starts with a claim about the question or problem, then supplies evidence from the data that supports that claim and ties them together with reasoning that makes clear how the evidence supports the claim. Simply being able to build a CER that makes logical and scientific sense is an indicator of some mastery of inference with data, however Rubin (2020) points out that it is also important that the scientific argument sits within the context and purpose of the problem or question being addressed, that it addresses the uncertainty associated with the data, and that it makes generalizations that allows for prediction about future data collection. Inclusion of those three pieces of an inference indicates full mastery of this component of data literacy.
3.2.6 Complexity of Data Sets

The primary indicator for mastery of complexity of data sets is the ability to work with a data set that falls on the complex end of the scale on all five of the characteristics outlined by Kjelvik and Schultheis (2019) and discussed in detail in section 2.1.1. Here, “the ability to work with” means applying all of the other five components of data literacy successfully to a data set with high levels of complexity. Now that I have outlined indicators for the six components of data literacy, next I will do the same for the components and indicators of PCK for data literacy. Since PCK for data literacy is a yet un-explored concept, there is less previous literature to draw from so indicators are identified by modifying research from other STEM fields.

3.3 PCK for Data Literacy

The work of this study uses the modified RCM as a framework (Figure 3) for defining and understanding PCK for data literacy. Within the RCM are three types of PCK: collective, personal and enacted. Though they have the same components, they were measured using different tools and so are represented as separate but connected concepts (Figure 5). As discussed briefly earlier in this paper, there are five components that have been defined within PCK (Magnusson et al., 1999; Park & Oliver, 2008). Three of these will be the primary focus of this research and will be discussed here in further detail as to how they apply to data literacy. The three components of interest to this study are: a) orientations to teaching with and about data, b) knowledge of students’ understanding of data, and c) knowledge of instructional strategies for teaching with and about data. The other two, knowledge of curriculum and knowledge of assessment were
excluded because there are not yet established learning progressions or assessments for data literacy specifically and attempting to develop some is outside the scope of this research. Since there is no existing research on teacher knowledge for data literacy, this study draws on research on PCK for science (e.g., Freidrichsen et al., 2011) as well as research on students’ conceptions of data literacy (e.g., Lee & Wilkerson, 2018), and the goals for teaching data literacy to pre-college students (e.g., Erwin Jr., 2015) as a baseline for analyzing the components of data literacy surfaced in this study.

3.3.1 Orientations to Teaching.

A teacher’s orientations to teaching with and about data include their beliefs about the nature of data and its role in STEM-integrated instruction as well as their beliefs about the purpose of developing data literacy in the classroom. Orientations to teaching do not exist on a scale from novice to expert the way that other components of PCK and subject matter knowledge do, they are more nuanced and can shift from one orientation to another. Magnusson and colleagues (1999) proposed nine different potential orientations to teaching science: process, academic rigor, didactic, conceptual change, activity-driven, discovery, project-based science, inquiry, and guided inquiry. However, others including Freidrichsen and colleagues (2011) have criticized these for being too rigid and, as they were built from elementary school teachers, not aligned with the belief structures of high school teachers. Despite this, framing orientation to teaching data literacy as focusing on student-centered learning is a useful measure and aligned with the teaching and learning goals of the bioinformatics project.
In a review of how the component of orientations to teaching had been conceptualized across research on PCK for science, Freidrichsen and colleagues (2011) found what they identified as the three overarching orientations to teaching science: a) conceptions of teaching and learning science; b) conceptions about the nature of science; and c) conceptions about the goals or function of science education. If these are modified to apply to data literacy, they become a) conceptions of teaching and learning with and about data; b) conceptions about the nature of data; and c) conceptions about the goals and function of learning with and about data. In continuing with the analogy drawing from Freidrichsen and colleagues more detailed descriptions of these orientations, the conception of teaching and learning data literacy has the indicators of beliefs about the role of the teacher in learning data literacy and beliefs about making data literacy attractive and comprehensible to students. The conception about the nature of data would include as an indicator an epistemological belief about what counts as data. Finally, the conceptions about the goals and function of learning data literacy would break down into beliefs about learning to do data literacy and learning about data literacy, which mirrors the way I have conceptualized PCK for data literacy as knowledge for teaching with and about data.

As has been highlighted in previous research (Frank et al., 2016; Kjelvik & Schultheis, 2019) there is still much disagreement in the field about what the goals and function of learning data literacy should be. Much of the research conducted thus far on data literacy at the middle and high school level has focused on procedural goals (e.g., Erwin Jr., 2015, Lee & Wilkerson, 2018) such as calculating measures of aggregate, creating graphs, or using a spreadsheet software to analyze large data sets. However,
some research has proposed that the goal of data literacy is to shift students’ mindset about data and how it applies to their life (e.g., Doresy & Finzer, 2017; Hardy et al., 2020) including developing the habit of looking for data in any given situation and supporting students’ agency over their own data in the role of data producers. These two views on the goals of data literacy are not mutually exclusive and I look for indicators along a spectrum that includes both in this study.

3.3.2 Knowledge of Students’ Understanding

Having a knowledge of students’ understanding of data includes how students learn with and about data, their interest and motivations to do so, and their preconceptions about data including about its purpose, how to use it, and where it comes from. It also includes knowing where students will encounter learning difficulties when developing data literacy.

Indicators for knowledge of preconceptions and learning difficulties are most easily observed in the planning and discussing phase of PCK at all levels: collective, personal, and enacted and are similar no matter what the content is. These include planning activities and discussion prompts that predict and purposefully challenge students’ preconceptions and developing scaffolding to help students through areas they will most likely encounter difficulties. Additional indicators at the teaching phase of PCK include being able to recognize when a preconception or difficulty is creating barriers for a student and being able to pivot to address unforeseen preconceptions or difficulties.

While surfacing what teachers know about the content of those preconceptions and difficulties for working with and learning about data is one of the goals of the first
research question of this study, some previous reach has been done on students’
conceptions about the nature and purpose of data. Lee and Wilkerson (2018) compiled a
number of these conceptions in their literature review including; a) students can view
data as true or factual rather than as having inherent uncertainty, b) students often bias
“fair” treatment in choosing samples rather than using true randomness, c) students
struggle with viewing data sets in the aggregate rather than as a collection of single
points, d) students may view data visualizations as pretty pictures or illustrations rather
than tools that can be used to make inferences about data, and finally e) students are
likely to allow their predeveloped assumptions to guide their data interpretations.
Additional conceptions that students hold about data include that students have a
constrained conception of data and that they don’t see the connections to their everyday
life (Gebre, 2018). Finally, Harris and colleagues (2020) found that the process of data
collection and analysis was not inherently interesting to students, but that some became
more interested if they saw how it was relevant to them and their lives or if they truly
believed it could actually be useful to people with power to make change. This study
seeks to uncover the extent to which teachers have knowledge of these student
conceptions.

3.3.3 Knowledge of Instructional Strategies

Knowledge of instructional strategies for teaching with and about data is
knowing how to address students’ preconceptions and difficulties and create an
environment in which learning can occur. It includes knowledge of activities for
developing data literacy and which examples and representations to use to scaffold
learning and to break down the complexity of real-world data sets. Similarly, to knowledge of student understanding, this study seeks to surface teachers’ knowledge of these strategies but there is some previous literature on what those strategies might be. Wolff and colleagues’ (2019) design principles for supporting the development of data literacy discussed in section 2.1.2 of this paper can serve as indicators for knowledge of activities. They include, a) scaffolding data within a full cycle of inquiry; b) scaffolding the data sets being used from simple towards more complex; c) creating activities that work with data that is familiar or relevant to students’ lives; d) connecting activities to larger ideas and transferable skills; and e) providing activities for students to engage in data collection themselves so they understand the process.

Additionally, Lee and Wilkerson (2018) provide some suggestions for teaching approaches that they theorize will support students in developing data literacy. Their suggestions include a) data literacy should be taught within the context of authentic and meaningful but also familiar contexts, b) students should be supported in developing the ability to consider datasets as aggregates, c) interpretive work with data visualizations should be engaged, and d) data literacy should be taught frequently and connected to students’ learning in math. While the second suggestion is too unspecific to serve as an indicator for this knowledge component, the other three, echo some of the same strategies outlined by Wolff and colleagues and together they provide a foundation on which to measure teachers’ knowledge of strategies for teaching data literacy.

Given these conceptualizations of and indicators for the constructs of PD, data literacy, and PCK, the next section will discuss the context and methods for the PD program, the extra treatment PCK Extension Workshop and the research study in which I
sought to develop a greater understanding of PCK for data literacy and how it can be surfaced and developed with teachers.
CHAPTER 4. METHODS

This research engaged in an early-stage or exploratory study (IES/NSF, 2013) in which the goal was to examine what knowledge teachers held about teaching for data literacy, the construct of PCK for data literacy, and how that knowledge could be surfaced and developed, in order to guide improvements to the existing bioinformatics intervention as well as developments of future interventions in STEM integration and data literacy. As Berliner (2004) and many others since have acknowledged and studied (e.g., van Duzor, 2011), teachers have a wealth of expertise that is content and context specific and which can be utilized by researchers to expand knowledge and understanding about the teaching field. The intervention was designed to engage teachers’ existing knowledge and expertise through discussion and collaborative planning in a way that gave them space to surface their existing PCK for data literacy. I, as the researcher, then used the knowledge and expertise that was surfaced by the teachers to identify themes and potential components of PCK for data literacy.

In order to identify the themes, I used a qualitative multiple case study approach (Yin, 2018) which drew on multiple data sources to examine four teachers and their interactions in a synchronous workshop series. A constant comparative analysis (Glaser, 2008) was conducted on transcripts of interviews and meetings and triangulated with participants’ written reflections. In the next section I describe the context of the larger existing bioinformatics intervention project from which I cut a slice for the extra treatment study.
4.1 Context

My study took place within the larger context of an intervention designed to support teachers in implementing a STEM-integrated unit on bioinformatics in their existing secondary school science classrooms (Yoon et al., in press). Bioinformatics, and specifically health and air-quality, were chosen for this study due to their transdisciplinary nature and their perceived immediate relevance to students’ lives. Culturally relevant pedagogy has been shown to be important for engaging underrepresented students in STEM (Barton et al., 2008). As a result, we chose to focus the bioinformatics unit on the issue of unequal asthma rates across different communities within urban centers in the U.S. We hypothesized that this issue would be engaging for students and in previous years implementing the PBL curriculum, teachers supported this hypothesis.

In order to bring bioinformatics into secondary school classrooms, the team designed a PBL unit on addressing asthma and air quality issues in local communities. Both health and climate are highlighted by Nadelson and Seifert (2017) as good context for STEM integration because they are discipline general. The unit uses mobile learning through small air quality sensors connected to a phone app designed by a member of the team to engage students with active collection of real-world data to address the provided problem. Working with sensors has been shown to be an effective way for students to position themselves as data producers and experience the messiness of real data (Hardy et al., 2020; Philip et al., 2013; Wolff et al., 2016). Additionally, mobile learning can provide tools for students to engage with data from their local communities (Headrick Taylor, 2017; Philip et al., 2013).
The team then designed a PD experience for teachers to develop the knowledge and skills needed to implement the unit in their classrooms. The pilot year of the program took place in August 2019 with a three-week in-person workshop with six teachers from the School District of Philadelphia, five of whom went on to implement the PBL unit in their classrooms. Due to the Covid pandemic, the PD was moved to a virtual environment for summer 2020. Through analysis of data from year one and extensive feedback and input from the teachers in the pilot year, both the PD and the PBL unit were modified for the second year. The PD course was conducted mostly asynchronously over four weeks in July 2020 with weekly synchronous virtual meetings to support community building and progress through the course. Two of the teachers from year one served as facilitators in the online course during summer 2020. Both year one and year two included monthly virtual meet ups throughout the school year to support teachers through their implementation, and in year two, each teacher from both cohorts was assigned a facilitator from the research team who served as a point of contact and support during the 2020-2021 school year.

Year three, which was the primary focus of this study, followed the same pattern as year two, with additional modifications made based on feedback from the previous year. Additionally, four participants from the summer PD were invited to participate in a PCK Extension Workshop Series which extended through the fall into the winter. The year three format and content of the student-facing PBL unit and the summer PD, as well as the PCK Extension Workshop Series are described in more detail in the next sections.
4.1.1 The Bioinformatics PBL Unit

Bioinformatics PBL Unit: Asthma & Air Quality comprises twenty 50 minute lessons designed to create an authentic STEM-integrated experience for students in order to support them in becoming scientists and agents of change. The entire unit is grounded in a real-world problem (Figure 6) which engages students with their own community and motivates them to come up with a potential solution through collecting relevant data, analyzing it, and communicating their findings.

Figure 6
The Bioinformatics PBL Unit: Asthma and Air-quality Problem

There is a rising case of asthma in urban areas in the United States. Researchers have hypothesized that air quality contributes to this phenomenon. However, there may be a number of other causes, including a rise in smoking rates, industrial pollutants, and social stress from a lack of access to life resources. The Town Council of Philadelphia has announced a new community program that will fund projects that are likely to support risk reduction of local asthma cases. These projects may include planting more trees to purify the air, an anti-smoking campaign, and building a park to promote activities that can alleviate stress. The Town Council asks you to research the issue and submit a proposal describing a project that your team would like to fund with evidence from public health and environmental data that supports your proposal.

The six components of subject matter knowledge of data literacy are woven throughout the unit in both explicit and implicit ways. Table 1 outlines the lesson objectives and activities that address the data literacy components. All of the data literacy activities are grounded in content relevant to the problem of asthma and air quality that the students are exploring. Teachers were encouraged to have students conduct the entire unit working in groups of three to four. At the end of the unit, each group produced a report on their findings and presented it to the class. Then, at the end of the school year, Penn GSE will host a student virtual summit in which students from all participating
classes get to meet some researchers working in the field of bioinformatics and one or two groups from each class will present their final projects. The lessons indicated in Table 1 are the lessons that were the focus of both classroom observations and video reflections during the extra treatment. As can be seen in Table 1, the component of data complexity was an integral part of the existing PBL unit, however, complexity of data sets was not explicitly addressed in the summer PD during the first two iterations. Previous research on this project found that the complexity of the data sets was one area that some teachers struggled with implementing (Miller et al., 2021; Shim et al., 2021). A greater focus on this component was added to the summer PD during the most recent summer program in order to better support teachers in implementing this component of the PBL unit.

4.1.2 The Existing Summer PD

Just like the second year summers’ PD, the third summer PD took place over four weeks during July 2021 and participating teachers were encouraged to put four to five hours a day into taking the course. However, the timing was flexible as most of the course was conducted asynchronously on the edX testing platform called Edge, a platform designed for large-scale online learning. Synchronous meetups were held on the first day of the course and each Friday morning over Zoom to provide space for participants to engage in community building activities and real time reflection. The course is divided into eight modules: 1) Introduction, 2) Science Education, 3) Bioinformatics, 4) Data Literacy, 5) Computational Literacy, 6) Mobile Learning, 7) Review of the PBL Unit, and 8) Next Steps. Participants were asked to complete two modules per week, though not all participants stayed on track.
<table>
<thead>
<tr>
<th>Data Literacy Competency</th>
<th>Lesson Objectives SWBAT...</th>
<th>Activities</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context</td>
<td>Lesson 7: Describe the context of data and why context matters</td>
<td>Students engage in a small group discussion about a dataset that has been taken out of context and come up with questions they have about it</td>
</tr>
<tr>
<td></td>
<td>Lesson 9: Interpret an external data set</td>
<td>Students are given access to a small partially curated dataset from the EPA and interpret the variables in it</td>
</tr>
<tr>
<td></td>
<td>Lesson 12/16: Record relevant context while collecting data</td>
<td>Students collect data around the school/neighborhood with the sensors and app and keep notes about any relevant factors that might affect their data</td>
</tr>
<tr>
<td>Variability</td>
<td>Lesson 7: Understand that data has variation and uncertainty</td>
<td>Class discussion about the definitions of variability and uncertainty</td>
</tr>
<tr>
<td></td>
<td>Lesson 8: Calculate and discuss the range of a dataset</td>
<td>Students work with a data set in Google Sheets to calculate the range and discuss its meaning and usefulness</td>
</tr>
<tr>
<td></td>
<td>Lesson 17: Analyze the variability of their data set</td>
<td>Students analyze their data set, paying attention to the variability and creating hypotheses for any deemed outliers</td>
</tr>
<tr>
<td>Aggregate</td>
<td>Lesson 7: Explain why it is often important to look at whole data sets rather than single points</td>
<td>Class discussion about weather vs climate using a dataset of temperature from April 2020</td>
</tr>
<tr>
<td></td>
<td>Lesson 8: Articulate the difference between the mean, median, and mode and what each can be used for</td>
<td>Students work with a data set in Google Sheets to calculate measures of centrality and compare them and articulate the differences</td>
</tr>
<tr>
<td></td>
<td>Lesson 13/17: Explore patterns in air quality data at sites across the school/community</td>
<td>Students examine their own data set and compare it with data from other groups to observe larger patterns</td>
</tr>
<tr>
<td></td>
<td>Lesson 18: Compare their data to data collected by students at other schools</td>
<td>Students use the shiny app tool built by our team to view data collected by students at other schools and look for patterns across the whole city</td>
</tr>
<tr>
<td>Visualization</td>
<td>Lesson 7: Describe best practices for creating graphs</td>
<td>Students engage in a small group discussion about two biased data visualizations</td>
</tr>
<tr>
<td></td>
<td>Lesson 9: create a bar graph in Google Sheets</td>
<td>Students create a bar graph that visualizes a particular question they are trying to answer</td>
</tr>
<tr>
<td>Lesson 10: read, interpret and compare a number of different visualizations of the same data set</td>
<td>Students examine a number of different visualizations of the same data set: map, bar graph, box plot and discuss what the pros and cons are of each</td>
<td></td>
</tr>
<tr>
<td>Lesson 17/18: Create visualizations of their data set that help in analysis and communication and choose which ones are most useful</td>
<td>Students create visualizations of the data set(s) they collected which will help communicate their findings and justify why they chose the format they did</td>
<td></td>
</tr>
</tbody>
</table>

**Inference**

| Lesson 7: Build a scientific argument from data from PARGASITE | Students use air quality data from the PARGASITE website to develop a scientific argument using the CER framework |
| Lesson 10: Build a scientific argument about the safety of different US cities | Students build a CER using a number of different visualizations of the same data set as potential evidence |
| Lesson 13: Hypothesize potential causes for patterns using data | Students look at data from around the school and hypothesize causes for the patterns observed |
| Lesson 17: Develop a scientific argument in response to their research question | Students use the data they collected in the neighborhood to develop a claim with evidence and reasoning that addresses their research question |

**Complexity**

| Lesson 8: Organize data in Google Sheets | Students work in Google Sheets to clean up and organize a data set so that it is in a usable format |
| Lesson 9: Download a dataset and extract the relevant data | Students are given access to a small partially curated dataset from the EPA and select and extract the variables they need |
| Lesson 13/17: Prepare a data set for analysis | Students extract their group’s subset of data from a larger class-wide data set, clean it, and set it up for analysis, all in Google Sheets |
| Lesson 18: Retrieve and analyze data from public data sets | Students access the EPA and NIH websites in order to retrieve data that is relevant to their question, and extract and analyze the data in order to compare it to their own |
Modules three through seven each end in a capstone assignment which synthesizes lessons learned from the module. The core data literacy features are woven into the rest of the course, though are mostly in modules four, six, and seven. A description of the core data literacy features in the previous PD as well as the core feature that was added during summer three to support my dissertation, and how they all align with the PD design characteristics is outlined in Table 2.

Table 2
*Core Data Literacy Features of the PD Course*

<table>
<thead>
<tr>
<th>Core Data Literacy Features of Course</th>
<th>Description</th>
<th>Alignment with PD Design Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Features Included in the Existing PD Course</strong></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| Explicit Definition and Discussion of Data Literacy Competencies | Participants read sections of select articles on data literacy, watch a video on data literacy, and review summaries and examples of the six data literacy competencies, *including how to work with complex data sets* | • Content focused  
• Makes connections explicit  
• Attend to complexity of data |
| Engaging with Data as a Learner | Participants engage with a full cycle of data collection, analysis, and inference using the sensors, Google Sheets, and the shiny app tool, (Lessons 9-12 of the PBL unit). | • Active learning  
• Models effective practice  
• Full cycle of inquiry  
• Attend to complexity of data |
| Discussion Forums | Participants reflect with their peers on the learning process and make connections to their own classrooms and students | • Collaboration  
• Offers reflection time  
• Context focused |
| Review PBL Unit | Teachers take time to go through the entire PBL unit, make modifications for their classroom, and discuss implementation strategies with their peers | • Time to modify  
• Context focused |
| **New Feature Added in Year 3 to support this study** |                                                                             |                                                           |
| *Content Representation* | **Participants develop a CoRe for one of the six data literacy competencies and post it to the discussion forum where other participants and facilitators can give feedback** | • Collaboration  
• Time for feedback  
• Makes connections explicit  
• Context focused |

*Note: Activities that were added to the PD for summer 2021 are indicated by bolded, italicized text.*
Through the core features, teachers engaged as learners with all six of the data literacy competencies and practiced applying them while working with real-world datasets that they collected themselves, and then in comparison to large public datasets from the U.S. Environmental Protection Agency (EPA) and U.S. National Institute of Health (NIH). A content representation was added as the capstone assignment for module 4 as a new feature of the PD in summer 2021. Additionally, one subsection that explicitly outlines how to work with complex data sets was added to the previous module 4 where the other components of data literacy were presented in the previously existing PD. These two changes to the existing PD were added explicitly to affect the research being conducted in this study as well as other studies being conducted on the bioinformatics project.

4.1.2.1 The Existing School Year PD. Despite these components designed to develop teachers’ content knowledge and PCK for data literacy, extended time working with the material is important for prolonged effects on teacher practices. The entire community of teachers, including some teachers from years one and two met for an hour twice during the 2021-2022 school year as of this writing to discuss implementation strategies and provide feedback on the unit materials. An additional meeting is planned for the spring. Though these sessions were run primarily by other researchers on the larger project and were not designed to directly support this research, due to the timing of the sessions, the participants in this study ended up playing an outsized role in the meetings as they were the bulk of the teachers implementing in the fall. Only one other teacher, from cohort 2, implemented during the fall and winter months of the school year.
The first of these meetings was held in August 2021 and was primarily a logistics meeting to determine distribution of sensors, assignment of facilitators and guidelines for implementation research expectations. The second meeting was held in February 2021 after the last session of the PCK Extension Workshop Series, discussed in detail in the next section. During this session, two teachers, both participants in this study, shared short videos of their classroom implementation and a short discussion followed each. Since this session happened after the last workshop session, but before the final interviews were conducted, some of the reflecting on practice that occurred in that session may have contributed to the development of subject matter knowledge and PCK for data literacy for the teachers in my extra treatment. Finally, for the purpose of this research, a small group of four teachers met for extra treatment sessions during the fall semester to focus further on developing PCK for data literacy. As this Workshop Series is the primary intervention in this study, it is described in some detail in the next section.

4.1.3 Bioinformatics PCK Extension Workshop Series

Four participants from the summer PD were invited to participate in an extended school year PD which focused on PCK development. This small group met synchronously nine additional times, starting in August, through February, for between one and a half to two hours each session, for a total of 15.5 synchronous hours. Additionally, the participants were asked to complete some asynchronous work to prepare for the sessions which was intended to take about four and a half hours for a total of 20 hours of intervention time. Details for each session can be found in Table 3. The sessions were held on the video conferencing platform Zoom and were facilitated by me.
Table 3
Workshop Session Outline for Extra Treatment on PCK Development

<table>
<thead>
<tr>
<th>Session</th>
<th>Date</th>
<th>Time</th>
<th>Description of Session Activities</th>
</tr>
</thead>
</table>
| 1       | 9-17-21 | 1.5 hours| • Introductions and overview of the project (20min)  
|         |         |          | • Introduction of CoRe Template (Appendix A) and review of an example CoRe (10min)  
|         |         |          | • As a full group, participants brainstormed answers to the question “What is a Big Idea?” (10min)  
|         |         |          | • Participants collectively chose one of the six components of data literacy (*Inference with Data*) and in breakout rooms of two groups of two, brainstormed big ideas for that component. (10min)  
|         |         |          | • Through a full group discussion, two big ideas were chosen as a primary focus (5min)  
|         |         |          | • Participants collectively developed a CoRe in a Google Doc template. (15min)  
|         |         |          | • The CoRe was discussed with probing questions to encourage participants toward higher-order reflecting. (15min)  
| Asynchronous Assignment #1 |         | 1 hour   | • Participants read background information on PCK  
|         |         |          | • Participants wrote an individual reflection on their conceptions of their own PCK both for science in general and for data literacy specifically.  
| 2       | 9-26-21 | 2 hours  | • Welcome/Check in. (10min)  
|         |         |          | • Review of PCK information and reflections from asynchronous assignment. (30min)  
|         |         |          | • Explanation of the purpose of video reflections and introduction to the Video Reflection Framework (Appendix B). (5min)  
|         |         |          | • Participants watched a video clip of a bioinformatics teachers from cohort 2 teaching data literacy activities from Lesson 13. (10min)  
|         |         |          | • Participants engaged in a reflection discussion on the lessons that highlights how the clips display components of PCK using the Video Reflection Framework. (30min)  
|         |         |          | • Participants shared implementation plans and discussed challenges and perceived barriers. (30min)  
| 3       | 10-10-21| 1.5 hours| • Welcome/Check in (15min)  
|         |         |          | • Review of Data Literacy Components (15min)  
|         |         |          | ◦ Participants think individually and then share out on the following questions: 1) In thinking
about the six focal components of data literacy, what is one area where you think your students will shine? 2) What is one area where you anticipate students will struggle? What specific misconceptions and/or challenges do you anticipate them having? 3) How are you currently thinking about addressing those challenges?

- Participants collectively chose one of the six components of data literacy (*Data Visualization*) and then individually brainstormed big ideas for that component. (10min)
- Participants shared out Big Ideas and then each chose one to add to the CoRe. (10min)
- Participants collectively developed a CoRe in a Google Doc template with each participant filling in one column. (15min)
- The CoRe was discussed with probing questions to encourage participants toward higher-order reflecting. (15min)

| 4 | 10-21-21 | 2 hours | • Welcome/Check in (20min)
• Briefly reviewed Video Reflection Framework (5min)
• Watched a video of one of the participants teaching a portion of Lesson 9 (20min)
• Discussed the video using the Video Reflection Framework as a guiding tool (45min)
• Implementation Check in using the following prompts: 1) Where are you in the process? 2) What has gone well so far? 3) What are you and/or your students finding challenging? (20min) |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Asynchronous Assignment #2</td>
<td>1 hour</td>
<td>• Individually watch a video of one of the participants implementing a portion of Lesson 7 which focuses on <em>Data in Context</em> and <em>Variability in Data</em> and fill out the Video Reflection Framework</td>
<td></td>
</tr>
</tbody>
</table>
| 5 | 11-8-21 | 1.5 hours | • Welcome/Check in (10min)
• Discussion of the Asynchronous Assignment, the video using the Video Reflection Framework as a guiding tool (20min)
• Review of the Objectives for Lesson 7 (5min)
• Participants individually brainstormed big ideas for *Data in Context*. (10min) |
<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asynchronous Assignment #4</td>
<td>1 hour</td>
<td>• Individually watch a video of one of the participants implementing a portion of Lesson 18 which focuses on <em>Inference with Data</em> and <em>Data in Aggregate</em> and fill out the Video Reflection Framework</td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>--------</td>
<td>-----------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1-10-22</td>
<td>2 hours • Welcome/Check in (20min) • Discussion of the Asynchronous Assignment, the video using the Video Reflection Framework as a guiding tool (20min) • Participants collectively brainstorm big ideas for <em>Data in Aggregate</em> (20min) • Participants collectively developed a CoRe in a Google Doc template with each participant filling in one column. (20min) • The CoRe was discussed with probing questions to encourage participants toward higher-order reflecting. (20min) • Implementation check in (15min)</td>
<td></td>
</tr>
<tr>
<td>Asynchronous Assignment #5</td>
<td>0.5 hours</td>
<td>• Participants individually write responses to the following prompts: o What do you know about how students understand data? What knowledge and motivations do they come in with about data? How does this prior knowledge create barriers but also opportunities for hooks? Feel free to use bullet points or free write. o What strategies have you found or thought work well for teaching with and about data? Think about all the different strategies, activities, visualizations and tools we’ve talked about for the last few months and others you might have thought of while teaching or planning. Feel free to use bullet points or free write.</td>
<td></td>
</tr>
</tbody>
</table>
| 9                         | 1-24-22 | 2 hours • Welcome/Check in (15min) • Participants responses to Asynchronous Assignment #5 have been added to a Padlet (virtual concept mapping tool) with each separate idea as an individual virtual “post-it.” Participants log in to the Padlet and review the responses (15min) • Participants interact with the concept map, creating connections between ideas, adding new ideas, and grouping ideas together (20min)
• Discussion of the concept map starting with the questions: what patterns do we see? Are there any clear groups that we could give an overarching title to? (40min)
• Reflections on workshop and future work with the questions: What is one takeaway from this workshop that you would like to continue to think about/focus on going forward? What is a hope that you have for your students around the concept of data literacy? (20min)

The schedule of the sessions was not regular in length or in frequency due to the complicated nature of participants’ schedules.

The sessions included a number of different strategies, including explicit discussion of PCK, implementation discussions, and review of data literacy content components. However, the two primary strategies were collaborative development of CoRes (See Appendix A for template, discussed further in section 4.3.4), and Video Reflections on videos from their classrooms (See Appendix B for template, discussed further in section 4.3.5). During the workshop, four CoRes (and framing for a fifth was discussed) and five video reflections were completed. The CoRe template served as a guide for teachers’ discussions during the Workshop Series sessions. They were provided with the template as a Google Doc which they could all access and edit. Once big ideas had been agreed upon and added to the template, the teachers worked as a group to fill in the components of the template. This process took on a slightly different format in different sessions, but primarily consisted of a writing period in which teachers quietly added their written thoughts to different boxes of the CoRe template followed by a
discussion period in which the facilitator asked probing questions about certain boxes and guided the teachers to discuss what had been written and add to it or modify it.

Teachers were allowed a lot of autonomy in choosing which of the six components of data literacy to focus on for the first two CoRes and which big ideas to map within those components for each of the CoRes. However, as a facilitator, I guided them in forming those big ideas and provided feedback on the CoRes as they were being completed and during discussions after. The research conducted by Vossen and colleagues (2020) serves as an example for this work, as they also engaged teachers in creating a collaborative CoRe around a topic and big idea that was not from disciplinary content. In their workshop, they had teachers of a STEM-integrated curriculum develop a CoRe for the topic of building connections between research and design. In this Extension Workshop Series, teachers developed CoRes for the components and big ideas within data literacy that fall outside their disciplinary content. The big ideas they chose are presented in section 4.3.4 and discussed in the findings Chapter 7.

The clips of videos that were watched for reflections were chosen by me, the facilitator. The reason for this is that I wanted to have teachers watch and reflect on clips that highlighted specific components of data literacy and I worried that a) teachers did not yet have a strong enough grasp of the PBL curriculum to be able to choose the strongest examples; b) that they would have different priorities for choosing a clip even if asked to highlight examples of data literacy PCK; and mostly that c) the teachers would not have the time or capacity to watch hours of video to choose an appropriate 10-15 minute clip. The clips were chosen to highlight exemplary strategies for teaching with and about data in order to prompt reflection on how those strategies could be described, generalized, and
applied to other contexts. For example, one clip displayed one of the teacher participants guiding students through a discussion about context by encouraging them to question their assumptions about a provided data representation. The five clips used ranged in length from 9 m 20 s to 22 m 34 s and had a total length of 74 minutes and 36 seconds. The two shortest clips were of the same teacher and were discussed during the same session (Session 7). For four of the five videos, teachers watched on their own time and completed the Video Reflection Framework prior to attending the synchronous sessions where they were discussed. This served as both a time-saving strategy, but also took some of the nervousness out of the process for the teachers as they did not have to watch themselves teach in front of others. The Video Reflection Framework prompted teachers to identify instances of enacted PCK in the form of what student learning was being elicited and what specific strategies were being used to elicit it.

4.2 Participants

The participants for this study were chosen from the larger population of participants for the Bioinformatics project. Ten teachers participated in the Bioinformatics PD course in summer 2021, all from public schools within the School District of Philadelphia. From this cohort of 10 teachers, four teachers were invited to participate in the extension PD sessions focused on building PCK. The criteria for choosing the four participating teachers was primarily a convenience sample, as they were the only four teachers from the summer 2021 cohort who were planning to implement the PBL unit in the fall semester. All participating teachers needed to be implementing the full Bioinformatics PBL Unit in the fall semester so that their
implementation aligned enough for collaboration to be productive for all parties.

However, it happened that the teachers also had a range of scores on their pre-survey for teaching practices around data literacy with one of the four holding the lowest score, two scoring near the middle, and one having the second highest score of the original ten teachers. The pre-survey for teaching practices measured teachers’ predisposition to teaching with and about different components of the PBL including use of mobile technologies, embedded socio-scientific issues, data literacy, STEM integration, and bioinformatics.

The four teachers also had a range of teaching experiences and worked at schools with a range of demographics that represented variability within the district (Table 4). Additional important context includes the fact that the School District of Philadelphia was entirely virtual for high school students and teachers from March 2020 until September 2021 and that in Pennsylvania, Biology is the only science tested at the high school level, so there is a state-mandated exam called the Keystone that all students must take in May. The need to prepare students for this exam is an external factor for all Biology classes in Philadelphia. More detailed information about each teacher is presented in the following paragraphs.
Table 4
Demographics on Teachers and Schools

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Years Exp</th>
<th>Subject Certification</th>
<th>Gender</th>
<th>Ethnicity</th>
<th>School</th>
<th>Percent Minority</th>
<th>Low Income</th>
<th>St Pop 21-22</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hallie</td>
<td>22</td>
<td>Biology &amp; Env Science</td>
<td>F</td>
<td>White, non-Hispanic</td>
<td>A</td>
<td>96%</td>
<td>100%</td>
<td>241</td>
<td>Magnet</td>
</tr>
<tr>
<td>Mary</td>
<td>1</td>
<td>General Science</td>
<td>F</td>
<td>White, non-Hispanic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manisha</td>
<td>10</td>
<td>Bio, Chem, Phy, Special Ed.</td>
<td>F</td>
<td>South Asian</td>
<td>B</td>
<td>92%</td>
<td>100%</td>
<td>913</td>
<td>Public</td>
</tr>
<tr>
<td>Will</td>
<td>18</td>
<td>Biology, Env Sci, History</td>
<td>M</td>
<td>White, non-Hispanic</td>
<td>C</td>
<td>85%</td>
<td>78%</td>
<td>3483</td>
<td>Public</td>
</tr>
<tr>
<td>Avg</td>
<td>12.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>91%</td>
<td>93%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: Minority is defined as all races/ethnicities except white, non-Hispanic. Low income is defined by U.S. federal standards.

4.2.1 Hallie

Hallie was the most experienced teacher of the group. The 2021-2022 school year was her 23rd year in the classroom, which, as she pointed out a few times during discussions, was over half of her life. She had experience teaching science at the middle school level, as well as Biology, Environmental Science, and Health at the high school level. In her implementation pre-survey, she indicated that the thing that excited her most about teaching biology was “the pragmatic application of biology and environmental science for my students to improve their own lives. I find much of what I am required to teach to be esoteric.” This comment highlights one of her epistemological stances which surfaced often throughout the Workshop Series on the lack of relevance of most common science topics to her students. Additionally, she indicated in her pre-survey that she had
extensive experience with Google Sheets and felt extremely comfortable working and
teaching with it. However, her score on the implementation pre-survey indicated that this
comfort only partially transferred to her classroom, as her average pre-scores for data
literacy implementation and STEM integration were 3.83 and 3.75 out of 5 respectively
which was right at and just below the median scores of 3.83 and 4.0 respectively.

During the 2021-2022 school year Hallie chose to implement the Bioinformatics
PBL unit in three sections of General Biology with 11th graders. School A where Hallie
taught was a small non-selective magnet school where 100% of the student body
qualified as “low income.” The school was over 70% Black and over 21% Hispanic.
Historically, less than 5% of students scored proficient on the Algebra 1 Keystone, and
about 10% of students scored proficient in ELA. School A was rated a 1 out of 10 (the
lowest possible score) in Math, Science, and Attendance and a 3 in Reading
(greatphillyschools.org). In Hallie’s own words describing her students, “The typical
student has a reading level of about 5th to 6th grade and a math level of about 3rd grade.
About 20% of the students overall have IEPs [Individual Education Plans].” School A
operated on what they called a “waterfall block schedule” which was on a 10 day cycle,
every day of which had a different schedule. Over each two week period, Hallie saw her
students for three 80 minute blocks, two 45 minute blocks, and two 40 minute blocks, or
an average of about 3.5 hours a week.

4.2.2 Mary

The 2021-2022 school year was only Mary’s second year teaching in a formal
classroom and her first year teaching in person. Mary was a career changer who came to
formal classroom teaching from a background in informal education. She did not think of herself as a science teacher, stating multiple times throughout the Workshop Series variations of “I’m not really a science teacher.” This identity as something other than a science teacher made Mary predisposed to STEM integration, stating in her application that the thing that excited her most about teaching was “Working on project and problem based lessons and activities. Creating spaces to make things - food, a mess, change, gardens etc.” This predisposition aligned with her scores on the pre-survey for implementation. While Mary scored near the median for data-literacy implementation on her pre-survey (3.58 out of 5), her STEM integration score was much higher at a 4.5. She indicated that she agreed or strongly agreed with all the statements about STEM integration in her classroom (e.g., *I consider connections to other lessons when designing and teaching lessons* and *I make connections between math, science, and technology explicit for my students*).

Though Mary was also at school A, and in fact taught in the classroom next door to Hallie, her schedule with her students was very different. Mary only taught one class, an elective career and technical education course on Urban Agriculture for seniors, which accounted for half of the students’ credits for the year. Due to the rotating waterfall schedule, no day was exactly the same, but she saw her students for approximately 17.5 hours a week, or 3.5 hours a day on average with the shortest day being just over 2 hours and the longest days being nearly 5 hours of class time. That meant that though her students had the same demographics (low reading and math levels and low attendance rates) as Hallie’s, their teaching contexts differed significantly. Mary was not teaching a tested subject, so had infinite freedom in her curriculum design, she had approximately 5
times as many classroom hours with her students as Hallie did, and she only had one
class’s worth (24) of students to pay attention to, whereas across her entire teaching load,
Hallie had five sections of students.

4.2.3 Manisha

Manisha was in her 11th year of teaching, with previous experience teaching
physics, chemistry, biology, environmental science, and earth science, all at the high
school level. Manisha was the participant who entered the PD program with the lowest
predisposition to data literacy. She scored a 3.33 and 3.25 on the pre-survey in data
literacy implementation and STEM integration, the lowest scores in both sections of any
of the ten teachers in the summer PD. She neither agreed nor disagreed with most of the
statements in these sections and clarified in one of the open-ended questions on the pre-
survey which prompted participants to describe how they currently teach data literacy in
their classrooms. Manisha wrote, “I do teach data literacy - however not at the level of
depth that I set as a standard - in my mind, as to what I define as data literacy. Data
Literacy in my classroom usually consists of evaluating graphs, visual representations
and infographics. Discussions are around what they notice and comparisons.”
Additionally, during the summer PD she showed her inexperience with teaching Google
Sheets, writing in response to one of the prompts, “I haven’t used Google Sheets before
with my students. I will need to do it step-by-step with my students as we learn together.”

Manisha’s hesitation around teaching data literacy in the classroom stemmed
partially from the context in which she taught. Manisha was a certified special education
teacher, and her school, School B, used a sheltered instruction model for learners with
IEPs, which meant that every section that Manisha taught was composed entirely of students with IEPs. She chose to implement the Bioinformatics PBL unit with four of her class sections, two of which were 10th grade General Biology and two of which were 11th and 12th grade Environmental Science. School B was a medium-sized neighborhood public high school where again 100% of students came from low-income households.

There was about a 50% graduation rate at School B. Approximately 54% of the student population identified as Black and 33% as Hispanic. Similar to School A, students at Manisha’s school did not perform on grade level in literacy and math. In her words, “literacy level and maths level varies from second to fifth grade on general.” Attendance was also a barrier at School B with fewer than 50% of enrolled students present on any given day and only about 20% of students regular attenders. School B utilized an alternating day block schedule, so Manisha saw her students for 90 minutes every other school day.

4.2.4 Will

Finally, Will was in his 19th year of teaching, though only his fourth at the high school level. For the first 15 years of his career, he taught Environmental Science at the middle school level, moving to his current high school position and teaching Environmental Science and Horticulture for the previous three years. Will scored the highest of the four participants on the data literacy component of the implementation pre-survey and the second highest of the ten summer teachers overall with a 4.08 out of 5. However, his STEM integration score was just at the median of 4.0. These scores are supported by some of his answers to the open-response pre-survey questions. He
embellished on his data literacy practice, writing, “I ask students to collect data from experiments. Then I help them input their collected data into spreadsheets in Excel or Sheets and then translate them into graphs.” However, Will’s background was in History and despite his 18 years teaching environmental science he still had a modesty about his abilities as a science teacher, writing,

As a trained History Teacher, I understand the connections between the subjects of Social Studies, but in the realm of Science, I do not connect the dots (so to speak) between the various subjects of science as easily. I do not know where the content has cross-over and what content is course specific.

This supports his lower score on STEM integration as he felt less confident finding connections in science to other topics and ideas.

Will chose to implement the Bioinformatics PBL unit with his general level Environmental Science class. Will taught at one of the largest schools in the district. School C, a neighborhood public school which had over 3,400 students, was the most diverse of the three schools with a student body that was about 28% Black, 27% Hispanic, 23% Asian, and 15% White. It had a higher attendance record (5 out of 10 on greatphillyschools.org) and graduation rate (76%) than the other two schools. In order to break up its large student body into more manageable cohorts, School C used a small learning community (SLC) model to attempt to create spaces within the student body where students can get to know a portion of their class more robustly. The environmental science class that Will implemented with was part of the Natural Resource Management SLC. The students were all 4th year students who had been taking the majority of their classes together for the last two years. In Will’s own words,
This group of students, specifically, is primarily non college bound students. There's probably a handful out of this class that will go to college, but most of them are looking to go into trades. This is a mixed group of special needs students with IEPs as well as high performing students. There’s a very big mix of abilities in the class.

Will had the added advantage of having been recruited to participate in the Bioinformatics program by a colleague who was part of cohort 2 in summer 2020 and who had already implemented the PBL unit in the previous school year and could serve as a mentor and thought partner.

One thing that all four participants had in common was their motivation for pursuing continuing education. All four were participating in at least one additional PD program outside this one during the 2021-2022 school year. Additionally, Mary was taking graduate level courses to earn her full professional certification, and Manisha ran an after school robotics club. Though they all taught in different school contexts, they were all part of the School District of Philadelphia and subject to the expectations, rules, and regulations of one of the top 20 largest school districts and the 5th poorest large school district in the US (NCES, 2020). As will be discussed later in the paper, the context of the School District of Philadelphia during the 2021-2022 school was a particularly challenging one.

In order to build case studies of the four participating teachers and the learning community they formed and answer the research questions, many different data sources were collected and analyzed. Those are discussed in detail in the next section.
4.3 Data Sources

This research drew on multiple data sources in order to build a case study of teachers’ content knowledge and PCK development for data literacy in STEM integration.

4.3.1 Surveys

As part of the larger bioinformatics research project, all participants in the summer PD took two surveys, a content knowledge survey and an instruction survey, before beginning the PD and again at the end of the four-week program. The content knowledge survey consisted of sixteen open-ended response questions in the pre-survey and a subset of nine of those questions in the post-survey. The questions in the content survey probed for teachers’ understanding of the main tenets of the bioinformatics project, including data literacy. Four of the questions on the pre-survey and three on the post-survey referred to data literacy (e.g., *In your view, what are the most important skills students need to have in order to work with data?*) The instruction survey consisted of twenty-seven Likert-scale questions using a five-point scale from strongly disagree to strongly agree. The questions asked teachers about their instructional practice in relation to the main tenets of the project. Of the questions on the survey, twelve were relevant to data literacy instruction (e.g., *I teach my students to recognize and understand patterns in data.*) Appendix C contains a list of all data literacy relevant questions on the surveys. All surveys for the course were completed virtually through the Qualtrics software program.
Additionally, all participants in the summer PD completed an application form when applying to participate in the summer program. This application asked questions about teachers’ expectations and interest for the course and their previous experience with certain components of the program (e.g., *What excites you the most about teaching biology/environmental science?* and *What are your expectations for this professional development program?). The application and the two pre-surveys were used to create initial profiles of the four participants in this project and serve as a baseline from which to measure change.

### 4.3.2 Discussion Forums

During the summer PD which took place on the Edge platform, teachers participated in discussions on Edge’s built-in discussion board environment. There were over sixty discussion forum prompts throughout the entire PD course. The prompts were developed based on previous literature on community and trust building in online courses (e.g., Hew & Cheung, 2014; Ng et al., 2012). Sixteen of these prompts addressed data literacy and are therefore relevant to this project (e.g., *How do the ideas presented so far on data literacy relate to concepts you already teach? Once you’ve responded, read a few other responses and comment. Are your experiences similar or different to theirs?*). Posts in the discussion forums were scraped from Edge after the course ended and organized in an Excel spreadsheet for analysis. As this study only examined a subset of four participants from the summer course, their posts were isolated from the rest of the participants for a total of 67 posts across the 16 relevant prompts. There was no serious discrepancy between the number of posts by each participant. The discussion forum posts
were used to triangulate with the pre and post tests to determine the effects of the summer PD and to create a more robust baseline for the teachers at the beginning of the Extension Workshop Series.

4.3.3 Synchronous Virtual Workshop Sessions

The nine sessions of the Extension Workshop Series were conducted over Zoom. Each session was recorded using Zoom’s built in recording feature. A total of 15 hours, 20 minutes, and 48 seconds of video was captured. Zoom also has a built-in auto transcription feature. This was used to create AI produced transcripts. I then watched each video to correct any mistakes in the transcripts. These transcripts serve as the primary resource for answering the first research question.

4.3.4 CoRes

One of the primary design components of the Workshop Series were the development of CoRes (Appendix A). These served as a framework for discussion, but the written artifacts also serve as a data source to support and expand on the themes developed during conversations. During the nine workshop sessions, the participants collectively developed four CoRes and the framing for a fifth. The topic and big idea(s) for each CoRe can be found in Table 5 and an example completed CoRe from session 3 on Data Visualization can be found in Table 6.
### Table 5

*The Big Ideas Explored in the CoRes*

<table>
<thead>
<tr>
<th>Session</th>
<th>Data Literacy Topic</th>
<th>Big Idea(s)</th>
</tr>
</thead>
</table>
| 1       | Inference with Data     | • Anyone can make meaning of data. You don’t need to rely on the conclusions of others.  
          |                         | • There is no perfect data. All data is biased.                               |
| 3       | Data Visualization      | • An effective visualization is something that makes the information and intention clear (the what AND the so what)  
          |                         | • Data visualizations tell a particular story about a set of data at a particular point in time.  
          |                         | • What you exclude in a data visualization matters as much as what you include |
| 5       | Data in Context (Framing only) | • Interrogating data is useful and necessary  
          |                         | • Data can be used to tell different stories with different intent  
          |                         | • Data is often manipulated for non-scientific reasons, to lead you to a specific conclusion.  
          |                         | • People make decisions about what and where and when to collect data  
          |                         | • A lot of things get measured because they are easy to measure not because they are important to measure |
| 7       | Complexity in Data      | • Data can take many different forms                                          |
| 8       | Data in Aggregate       | • Sample Size Matters  
          |                         | • Different Aggregate Measures support different visualizations  
          |                         | • Trials, Trials, Trials                                                  |

Additionally, all teachers participating in the summer PD completed an individual CoRe as the capstone assignment for module 4 of the course. They were prompted to choose one of the six competencies for data literacy that they had just explored explicitly in module 4 and highlight one big idea within the topic of their choosing. These initial CoRes were completed without any background information or explanation of PCK or
### Table 6
*Example Completed CoRe on Data Visualization*

<table>
<thead>
<tr>
<th>Big Ideas</th>
<th>A: An effective visualization is something that makes the information and intention clear</th>
<th>B: Data visualizations tell a particular story about a set of data at a particular point in time.</th>
<th>C: What you exclude in a data visualization matters as much as what you include</th>
</tr>
</thead>
<tbody>
<tr>
<td>What you intend the students to learn from this idea</td>
<td>The what AND the so what. Trying out different ways to represent the data, and deciding which makes the intention clear, by trial and error is part of the process. There is more than one RIGHT WAY.</td>
<td>That those creating data visualization make choices &amp; those evaluating data visualizations need to think about what choices were made as they view the visualization</td>
<td>Know the parts that are represented, and those that are not. Know the value /meaning of the representation, as well as the shortcomings.</td>
</tr>
<tr>
<td>Why it is important for students to know this.</td>
<td>Data representation includes identifiable design choices, and a challenge in the process of making models is making meaningful choices.</td>
<td>It is important for students to understand that intended and unintended bias enters into communications around data.</td>
<td>Clarity of context, one time, place, situation, does not necessarily correlate/connect/ help predict another.</td>
</tr>
<tr>
<td>What else you know about this idea (that you do not intend students to know yet).</td>
<td>Pretty isn’t worth points.</td>
<td></td>
<td>Words biases/ manipulations / agendas</td>
</tr>
<tr>
<td>Difficulties/limitations connected with teaching this idea.</td>
<td>It can feel wishy-washy for students, who are accustomed to “only one of the choices is correct”.</td>
<td>finding/curating a variety of data visualizations that could be evaluated and interpreted would take time and effort. I think with the right examples this would be fairly straightforward to teach</td>
<td>Getting past feelings overhanging beliefs</td>
</tr>
<tr>
<td>Knowledge about students’ thinking which influences your teaching of this idea.</td>
<td>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”.</td>
<td>Sharing dramatic examples of manipulated visualizations would be an effective way to influence them. Especially something related to manipulating youth</td>
<td>Anchoring into beliefs, that clouds views towards new openings/ perspectives that are potential “blind spots.”</td>
</tr>
</tbody>
</table>
Other factors that influence your teaching of this idea. | Data visualization is everywhere - so once students are made aware perhaps they can discover examples to share out with class | The necessity, in this age of information overload... and mental media onslaught! Before at least you could go home to a TV. now you carry everything - everywhere.

| Teaching procedures (and particular reasons for using these to engage with this idea). | Represent a set of data in various ways, and critique the choices along with students, practicing a group decision making process can be valuable for many purposes. | 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 - offer specific ideas for how they might do this - and then have students present their visualizations to the larger group | Allow the unraveling!

| Specific ways of ascertaining students' understanding or confusion around this idea (including likely range of responses) | Have students identify design elements, and state how they increase the clarity of the story, or if they don't. | Students could then be asked to explain and support how they created their visualization and if they still think it is a good way. Were ideas changed by seeing the various options? | Listen to student responses, ideas, thoughts.

Note: This CoRe was completed collaboratively by teachers during session 3 of the Workshop

the purpose of a CoRe and so served as a baseline for assessing both teacher change and the usefulness of the CoRe as a tool for reflection and growth.

4.3.5 Written Reflections

For some of their asynchronous assignments during the Extension Workshop Series, participants were asked to watch a video of themselves or one of their peers teaching. After watching, they were asked to complete the Video Reflection Framework.
(Appendix B) which prompted them to identify instances of enacted PCK in the form of what student learning was being elicited and what specific strategies were being used to elicit it. As not every participant completed the framework for each assignment, a total of 9 completed frameworks were collected. Though the videos and frameworks were discussed during the workshop sessions, the written artifacts serve as secondary data to triangulate PCK themes developed during the discussion. An example completed video reflection by Hallie on one of Will’s videos is displayed in Figure 7.

Additionally, participants were asked to complete two sets of written reflections about their own PCK. These written reflections were completed once at the beginning of the workshop sessions, in between sessions 1 and 2, and again at the end, after the final session. The questions prompted them to think about their own PCK for biology or environmental science depending on their teaching subject and then separately specifically for data literacy (e.g., Where do you feel your PCK for science teaching in general is strongest? and Where are you most motivated to grow your PCK for data literacy?) These written reflections were used in combination with the interviews and session transcripts to determine changes in knowledge perception for each participant.
Figure 7
Example Completed Video Reflection Framework

Video Reflection Framework for PCK for Data Literacy

Data Literacy Content Knowledge
- What big ideas of data literacy are being focused on? (add as many lines as needed)

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Big Idea</th>
<th>Thoughts/comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>0:28</td>
<td>Inference around data (CER)</td>
<td>“so how do you think air quality in your neighborhood compares to other neighborhoods around Philadelphia”</td>
</tr>
<tr>
<td>4:27</td>
<td></td>
<td>“if you live closer to the park your air quality should be …” (student reasoning)</td>
</tr>
<tr>
<td>8:41</td>
<td>Looking at Data in Aggregate</td>
<td>“You can easily interpret and synthesize that whole messy data set”</td>
</tr>
</tbody>
</table>

Engaging Students’ Understanding of Data
- What strategies are being used to engage students’ understanding of data?
- What aspects of teachers’ knowledge of student understanding are displayed? How is the teacher displaying this knowledge?
- What aspects of students’ understanding of data are surfaced?

(Add as many lines as needed)

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Strategy being used</th>
<th>Display of teacher knowledge</th>
<th>Aspect of student understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td>2:10</td>
<td>Affirming “That’s some good thinking”</td>
<td>Students will be more ready to take intellectual risks if their risks are acknowledged</td>
<td>Reasoning matters, and risks may be rewarded.</td>
</tr>
<tr>
<td>6:23</td>
<td>Probing “Why did we have, why did I have you break the data down into locations? Why did I from those locations make you come up with the mean, median mode and range for each location”</td>
<td>Students may be reluctant to volunteer thoughts, but for the most part, they have some answers inside them, we just have to look hard sometimes.</td>
<td></td>
</tr>
<tr>
<td>8:06</td>
<td>Affirming “Thank you Pedro.”</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Were there any missed opportunities to engage with student understanding?

Instructional strategies for teaching with and about data
- What strategy is being used? (e.g., visual representation, model, vocabulary)
- How is the strategy matched to the learning goal?
- What evidence is there that the strategy used was helpful or confusing to students?

(Add as many lines as needed)

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Strategy</th>
<th>Connection to Learning goal</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>5:07</td>
<td>Vocabulary</td>
<td>“the categorical variables contain finite distinct categories”</td>
<td></td>
</tr>
<tr>
<td>9:44</td>
<td>Data Visualization</td>
<td>“What chart is going to be the best one for that Ryan?”</td>
<td></td>
</tr>
</tbody>
</table>

Other Thoughts/Comments/Questions/Feedback:
Maybe remind everyone how bar graphs use these on the X, or pie charts, where appropriate, whereas the continuous ones work better with line graphs. (mostly a note to self). Continuous might be from negative infinity to positive infinity… This would be a good opportunity to connect the type of graph you are choosing to the vocabulary reviewed just moments ago.
4.3.6 Concept Map

On the last day of the Workshop Series, participants collectively created a concept map for PCK for data literacy. For the asynchronous assignment leading up to the final session, they were asked to brainstorm responses to the following questions: 1) What do you know about how students understand data?; and 2) What strategies have you found or thought work well for teaching with and about data? They were encouraged to submit their thoughts in bullet point format the day before the final session. Their bullet points were then transferred to a virtual concept mapping tool called Padlet with their names attached to the thoughts. During the session, teachers were prompted to read through the posted ideas and create connections between them. They were then encouraged to add additional ideas, and to move the notes around on the board to group them into themes. This was all done synchronously, so the teachers were able to discuss the process of connecting and grouping the ideas. The facilitator then guided the teachers to identify themes within the structure and groups. The results of this concept mapping activity are discussed in detail in Chapter 5.

4.3.7 Interviews

At the end of the Extension Workshop Series, a semi-structured interview was conducted with each participant. The interviews asked teachers about their orientation to teaching data literacy (e.g., *In your view, is data literacy important to teach in a science class, why or why not?*) as well as asking them to reflect on the components of the Workshop Series and how they affected their knowledge and learning (e.g., *What parts of the Workshop Series did you find most supportive of your growth as a teacher?*) The
entire interview protocol can be found in Appendix D. The interviews were conducted virtually over Zoom and Zoom’s internal recording and transcribing tools were used to record and create transcripts of the interviews. The interviews ranged in length from 40 minutes to 58 minutes with an average length of 49 minutes and 43 seconds and a total time of 3 hours and 19 minutes.

4.3.8 Classroom Observations

As part of the larger bioinformatics project, each teacher was assigned a facilitator from the research team who was responsible for organizing and conducting classroom observations during implementation. I served as facilitator for the four teacher participants in this study. During classroom observations, a protocol was used collect notes on observations and inferences about actions taken in the classroom (See Appendix E). Additionally, video cameras were set up around the room to capture the lesson. One was set up at the back of the room, and one was set up to focus on a specific pair of students who were working as a group. The same pair of students was videotaped for each observation session. Due to schedules, timing of classes, and the number of sections of students being taught by each teacher, the amount of video for the teachers was not consistent ranging from just over 5 hours for Mary and over 28 hours for Manisha. The videos of the classroom observations were used for the video reflection activities within the workshops. The videos and notes were also used to triangulate information and ideas discussed in the workshops.
4.4 Data Analysis

While the teachers in this study were provided a definition of PCK and a section of the second session of the Workshop was devoted to discussing it, they were not asked to identify the components of PCK. The Workshop was designed to elicit their existing PCK and was structured around the larger components of PCK which served as guides for the duration of the PD. As such, the role of identifying the components of PCK from the transcripts and artifacts, fell to me as the researcher. This study took a qualitative case-study approach (Yin, 2018) to investigating the research questions it seeks to answer. Qualitative research is often multifaceted and non-linear and can take many different forms (Ravitch & Carl, 2016). As such, I used a number of different analytic exercises to explore my data, pull out themes, and make meaning of the multiple sources and the stories they told.

4.4.1 Memos

In order to track my evolving thinking on themes and answers to the research questions, I wrote a number of memos throughout the data analysis process. As Ravitch and Carl write, memos “are a way to capture and process, over time, your ongoing ideas and discoveries” (Ravitch & Carl, 2016, p. 159). After each session of the Workshop Series, I wrote a memo on my impressions about the session and planning thoughts for the next session. I used these memos to help plan future sessions and eventually to help determine the interview questions for the post-workshop interviews. Additionally, once the Extension Workshop Series was complete, I read through the transcript for each session and wrote a memo at the end of each read through. These memos helped me to
identify some themes and also to highlight and save specific quotes and moments that I thought might be interesting to include in the write up. Finally, I used the same process for the transcripts of the four interviews, reading through each and writing a memo at the end on my thoughts on each participants’ strongest points and their orientation to the different research question components.

4.4.2 Coding

Before beginning coding, the transcripts from all nine Workshop sessions and all four interviews were compiled into the online qualitative analysis tool Delve (delvetool.com). Before initiating coding, I added a number of big overarching categories to the available codes based on the components of PCK and my research questions, these were: a) conceptions of teaching and learning with and about data; b) conceptions about the nature of data; c) goals for teaching data literacy; d) knowledge of students’ understanding of data; e) instructional strategies for teaching data literacy; f) teacher learning and development; and g) workshop reflections. Within the final code for “workshop reflections”, I added two subcodes: “CoRes” and “video reflections”. I began by deductively coding snippets into these larger categories of PCK and the aspects of the Workshop Series. However, as I read through the transcripts, I used a constant comparative method to generate additional themes that fit within those larger overarching categories (Glaser, 2008). For example, I started by adding any statement that demonstrated teachers’ ideas about the goals of data literacy to the code “goals of teaching data literacy.” However, some themes within that code quickly began to emerge, so the codes of “learning to question” and “avoiding manipulation” were added. The
initial big bucket code of “knowledge of student understanding” was sorted into subcodes that included “fear of numbers” and “need for right answers.” “Sharing ideas” was added to the “workshop reflections” code. As new codes were inductively added, I went back through transcripts I had previously coded to deductively code for the new themes. I went through multiple cycles of this process of deductively coding snippets into categories, then inductively adding additional codes, then going back through the transcripts to recode snippets into the new categories.

As my list of themes began to grow, I collapsed my codes into more generalized themes that represented both the breadth and depth of the thoughts and ideas shared by the participating teachers. For example, the codes “tools are a driver” and “data literacy is powerful” and “data is a useful and powerful tool” were all collapsed into the code “tools of data literacy are powerful” because nearly all of the instances in the “data literacy is powerful” and “data is a useful and powerful tool” code were in reference to Google Sheets. The statement “Being able to compile it like we discussed in the class. Being able to teach students the process and the understanding of the information and then how to make the analytical much easier is a powerful skill that students need.” was originally coded as “data is a useful and powerful tool” because of the use of the word “powerful” but was also referencing Google Sheets through the “make the analytical much easier” phrase and so fit into the new collapsed category of “tools of data literacy are powerful.”

As I collapsed my codes and continued to cycle through the process of reading and coding, Delve provided me with an easily accessible count for each code so that I could determine the significance of each of the themes and the weight they each carried within the larger set of data. This led to another example of codes coalescing into larger
themes among the four codes: 1) scaffolding; 2) modeling; 3) teaching the tools; and 4) demonstrating mistakes. Each of these codes were within the larger overarching code of “instructional strategies” but had a lower code count than most of the other themes. Upon closer reading, it was determined that these codes could be combined into a larger theme of “scaffolding the tools through modeling” with “demonstrating mistakes” and “repetition” emerging as sub-codes within that larger theme.

Additionally, one code was teased apart during this process as well. A large number of instances had been coded as “student interest” within “knowledge of student understanding” however upon closer reading of these snippets from the transcripts, this code actually comprised three separate levels of student interest: 1) teachers’ knowledge about what interests or doesn’t interest their students; 2) teachers’ predispositions and orientation to interest and relevance as a fundamental factor of the role of data literacy in classrooms; and 3) specific instructional strategies for engaging students’ interest to learn with and about data.

Through this process of emerging and then collapsing codes, a final coding manual was developed (See Appendix F) which included 32 different codes organized into the original six big overarching categories which aligned with my research questions. Once I had settled on this set of themes that felt well balanced, I added additional data sources that were relevant to each research question to find additional support for each theme.
4.4.3 Secondary Sources

Though the transcripts from the Workshop Series sessions and the post-workshop interviews served as the primary data sources for all three research questions, there were a number of secondary sources that were used to triangulate the themes developed through the coding process. For research question one on the components of PCK the themes found in the transcripts were compared to the themes highlighted by teachers in the concept map created by teachers on the last day of the Workshop Series to determine alignment. Additionally, for research question one and for research question three on the PD components, the themes found in the transcripts were applied to the written CoRes and the Video Reflection write-ups to determine to what extent the themes were present in these additional sources. For research question two which examined the learning of each teacher individually, their pre and post PD surveys, their discussion forum posts from the summer PD, and their pre and post workshop written reflections were also used to build a case for each individually. This process is describe in more detail in the next section.

4.4.4 Building Cases

To investigate the first and third research questions for this study, I treated the four teachers as a single case of a learning community. The surfacing of components of PCK that occurred during the Workshop Series truly was a collaborative effort with teachers brainstorming collectively and building off one another’s idea during the synchronous meetings. Likewise, the components of the PD were designed to be engaged
collaboratively. As such, it seemed appropriate when determining the findings for these two questions that the participants not be separated into different cases.

For the second research question on teacher learning I conducted a multiple case study approach (Yin, 2018). Each teacher’s pre and post summer PD and post-implementation survey open-response questions, their pre and post workshop written reflections, their summer PD discussion forum posts, and their post-workshop interviews were compiled within separate “projects” in Delve by teacher. This allowed for a new round of coding that was unique to each teacher. The coding manual developed during the iterative coding process (Appendix F) was applied to each teacher’s interview and written responses. While not all codes from the learning community coding manual applied to individual teachers, no new codes were added. The written responses and interviews were analyzed and compiled in chronological order to allow for a sense of change over time. Some classroom context presented or discussed during the Workshop sessions was included in building the cases as well as facilitator observation notes from classroom observations. Once each teacher had been analyzed individually and a case had been developed, cross-case themes were determined and presented.

4.4.5 Researcher Positionality

Though this study did not engage in true participatory research, as the intervention, methods and analysis were all conducted by me (Cornwall, & Jewkes, 1995) my role as facilitator of the Workshop Series session made me a participant in many of the discussions held as well as arbitrator to some degree of the focus of the sessions. As such, my positionality as a researcher and a former educator is important to consider. All
Qualitative research is bound by the subjectivity of the researchers involved; thus, it is important for the researcher to establish their research identity, including how their own experiences and goals may affect the data collection and meaning-making process (Ravitch & Carl, 2016).

There are a number of important factors to consider about my positionality within the learning community established during this study. The first is that though my background is not in biology or environmental science and I am not currently teaching high school, I was a public high school science teacher in an urban environment for four years followed by three years as an informal science educator to middle and high school students. So, while I do not have experience teaching in the specific context the participants in this study are working in now, I do have experience teaching science to a similar demographic of students, and I drew on that experience to add anecdotes to the Workshop Series conversations on multiple occasions. Secondly, I played a major role in the development of the Bioinformatics PBL Unit and its corresponding resources. Though I tried to approach the process of supporting the teachers with an open mind, and I am aware of many flaws in the unit and its supporting resources, my role in its development put me in a position of authority over the curriculum which was acknowledged by the teachers’ frequent questions to me about the curriculum documentation.

A third important component of my positionality within the group was that in my role as researcher, I was able to visit all four teachers’ classrooms and observe them as they implemented. While the teachers did watch short clips of each other implementing during the Workshop Series, I had a much more complete view of what was happening in
each teachers’ classroom. This gave me the ability to make connections for the teachers that they were unable to make because they were each siloed primarily within their own context. I often pointed out these connections as a way to initiate further conversation and thought-sharing between the participants. Having this birds-eye view also allowed me to design the workshop sessions to focus on common themes or challenges I was seeing across their different classrooms and implementations.

Finally, I tried throughout this project to be acutely aware of the stressors and hardships placed on teachers, especially urban public school teachers, and especially urban public school teachers during a global pandemic. This led me to be forgiving of participants needing to miss or duck out early of Workshop sessions occasionally. It also led me to redesign the plan for the Workshop Series to better accommodate teachers’ schedules and available time for synchronous and asynchronous work. I mostly tried to defer to the teachers about which topics and big ideas they wanted to discuss or that they felt would be more beneficial to them. Despite the fact that they were being monetarily compensated for participating in the workshop I was aware of how valuable teachers’ time is and I wanted to ensure that the teachers found the time spent in the Workshop sessions to be beneficial to them. This often meant that the Workshop session discussions went in directions I didn’t expect, however the discussions were rich and resulted in robust findings, presented in the next few chapters.
CHAPTER 5. FINDINGS ON RQ1: THE COMPONENTS OF PCK FOR DATA LITERACY SURFACED BY TEACHERS

As PCK for data literacy is an under-studied topic, one of the primary goals of the Extension Workshop Series was to work with expert science teachers to surface details about what PCK for data literacy might look like. The first research question for this project thus sought to answer this question: What components of PCK for data literacy were surfaced by teachers during the PD experience? Specifically, I focused on the components of a) orientation to teaching with and about data; b) knowledge of students’ understanding of data; and c) knowledge of instructional strategies for teaching with and about data. Within each of these components a number of themes emerged which were solidified through the analysis process described above into topic specific knowledge components that are unique to teaching data literacy. Figure 8 shows a summary of the seventeen components of PCK for data literacy that I identified from the knowledge surfaced by the teachers during the PD. The remainder of this chapter presented each component in turn with supporting data and examples.

The process of surfacing specific, descriptive components of PCK for data literacy within students’ understanding of data and strategies for teaching to that understanding was an engaging process for the four participants that stretched over multiple months. As Manisha described in session three of the Workshop Series, the conversations around data literacy and how students and teachers understand it were often complicated and multilayered, saying
This is fascinating because, just as much as we are trying to unpeel, to get to the data and the literacy and understanding, here we’ve just exposed another layer. So, we’ve got compounded levels of all the stuff and all we are trying to do is unpeel the simplicity of just data literacy. And it’s fascinating. It exposes all these layers that we’ve got piled on top of it as we’re trying to dig down to the facts or the clarity around it.

Figure 8
*Summary of the Components of PCK for Data Literacy*
In this statement, Manisha captures both the interest and enthusiasm the participants showed throughout the process for continued interrogation of the process of teaching and learning data literacy, but also the way that the richness of the conversations often led to even more questions as well as deeper layers of understanding.

To present an overview of the complexity of thought that went in to trying to answer the question of what PCK for data literacy looks like in a specific, tangible way, Figure 9 shows a concept map built by the participants during the last session of the Workshop Series. Participants were asked to brainstorm asynchronously before the session in response to prompts about students’ understanding of data and strategies for teaching data literacy (See Table 3 for full prompts). Their responses were added to the Padlet virtual platform as distinct ideas. Then, during the session, participants interacted with these ideas, creating connections between different thoughts, adding new ideas, and ultimately seeking out patterns among the posts to identify a few larger themes. Figure 9 serves to demonstrate the truly collaborative nature of the Workshop Series sessions.
While Chapter 7 will present some findings on personal and enacted PCK, the findings presented in this chapter are of collective PCK among the group of four participants and me as the facilitator. The themes from this concept map, as well as others that surfaced during my analysis of the extended data sources will be presented in this chapter as somewhat discreet ideas. However, I will refer back to this concept map as a way to convey the interrelated nature of these themes and the way that many of them built on and supported each other to create a concept of PCK for data literacy.
5.1 Orientation to Teaching With and About Data

A mistake that we in the educational community make is to think that the person best suited to teach a subject comes from a wealth of content knowledge without consideration for how hard won that knowledge might be. Math people rarely become effective math teachers, particularly if they cannot articulate their processes in a way that is accessible to their students. There is a degree of humility and insecurity with the subject matter that can bring out greatness as teacher and students discover their learning together.

– Hallie
Hallie gave the above reflection during session 2 of the Workshop Series during the review of the asynchronous learning on PCK. I start this section of the findings with her quote because not only does it summarize from her own perspective an understanding of PCK, but it also emphasizes the way that all four teachers approached teaching data literacy, with a degree of humility. Orientations to teaching with and about data includes beliefs about the nature of data and the role and purpose of developing data literacy in the classroom. This humility allowed for open and reflective conversations which created space for teachers to truly question what they believed about data and data literacy. This section is divided into the three primary subcategories of orientation to teaching with and about data literacy; a) conceptions of teaching and learning with and about data; b) conceptions about the nature of data; and c) conceptions about the goals and function of learning with and about data.

5.1.1 Conceptions of Teaching and Learning Data

Due to the self-selection nature of the participants in this study, it is unremarkable that they all believed data literacy to be a vital life-skill for students to have that needs to be explicitly taught. However, there were two subtleties to this belief that emerged as themes during this research: a) it is important to teach data literacy while grounded in relevance to student’s lives; and b) the tools of data literacy are powerful and need to be explicitly taught. These two themes added specificity to the conception that data literacy is important to teach explicitly. Each theme will be explored in the sections below.
5.1.1.1 Data Literacy Must be Grounded in Relevance. Relevance to students’ lives and experiences was a connecting theme through all components of PCK and content knowledge for data literacy. Mary was the most vocal and explicit about her epistemological stance towards relevance for learning data literacy, beginning in the second workshop session during a discussion of how the RCM (Figure 2) mapped onto the components of subject matter knowledge for data literacy (Section 3.2), Mary said,

I feel like, another important piece is, “why should they care about any of this” that is sort of, I mean, I have a sense of why they should care and with a group of students there’s so many of them and they all might have different reasons that they might or might not care or might or might not feel overwhelmed. But I feel like contextualizing it even in a meta, more meta way, like how they might ever use this information in their actual lives would really be helpful.

By “any of this” Mary was referring to the six components of data literacy subject matter. Throughout the workshop sessions, Mary repeatedly stated her orientation to the importance of teaching data literacy specifically as it related to students lives and as she said above, how it can be useful to students. She sustained this predisposition with many small comments such as, the following from session 5, “So to me, I'm really practical. Like, why would a student care about air quality data?” culminating in her post-workshop interview when she shared her anxiety around the need to present data literacy in a way that was relevant to students saying,

It's so easy to turn a kid off and then they're turned off of this whole subject. And that’s the thing I'm always afraid of is that I'm going to be so excited and I'm going to be so cheery about this thing that they're like, "This is the most boring thing. Why are you making us do this?"
While Mary’s focus on relevance to students’ lives as a major orientation to teaching and learning data literacy was the most prominent and vocal, she was not the only one for whom this orientation manifested and was an important factor. In session 4, during the reflective discussion on his first classroom video, Will spoke about he was also using relevance to students’ lives as a driving focus for how he was teaching the data literacy concepts, saying,

I just made it relevant to them about Covid. Like everybody's worried about masks and filtration and we went over the different N95 masks versus the cloth masks and filtration in the school system and in their homes…Yesterday we talked about analytics and how analytics is used in everything from you know sports, schools, and medicine. I'm trying my best to make it as relevant and pertinent to them, so that they draw those correlation between what we're doing in school to how it improves them in the future.

In this statement, Will echoed Mary’s orientation towards grounding the teaching of data literacy in students’ own lived experiences, and also their future goals. When he said, “how it improves them in the future” that harkened back to Mary’s comment from the previous session about how helpful it would be to have a concrete explanation for “how they might ever use this information in their actual lives.” Manisha also displayed an orientation toward the importance of grounding data literacy in relevance to students’ lives, saying in her interview,

There's a lot of data, but it's more at a higher level and our kids don't care for it. I think the one piece that I kept on harping about, and I still do is the data that is relevant to our kids in terms of the numbers in the population, because that's relevant to them and their families, a lot of the surrounding data was not.
Here, Manisha also made reference to the explicit context of the data used in the Bioinformatics PBL Unit. The reference to “numbers in the population” was in connection to asthma rates and how those are directly related to students lives, but “the surrounding data” which referenced the air quality data is not. This was again an echo of Mary’s comment from the sessions, “why would a student care about air quality data?” Manisha was also thinking about data literacy as it related to students’ lives and how the “higher level” data wasn’t as important to teach, because it wasn’t grounded in that relevance.

Though the concept map created during the last session of the Workshop Series was focused on knowledge of students and strategies, “Relevance” was a theme identified and added to the map by Will. By zooming in on the top left section of the concept map (Figure 10) the interconnectedness of the concept of relevance can be highlighted. While Will originally added the Theme Box (in Green) for relevance, he added it without a description, just as a theme title, during the ensuing discussion about the theme that Will had identified, comments made by Mary, “Right, there's like a, why does this matter to students? Because really if we can't make it matter, what is it, why would they care?” and Hallie “The ‘so what’? [in chat]” were both added to the theme box to the box along with a further idea about Students’ Orientation to Data through motivation. This demonstrated both the collective nature of this orientation and the way it was connected to other components of PCK.
Note: This is a zoom in on the top right section of Figure 9. The Theme boxes are colored green. The red stars represent knowledge of students’ understanding of data and the yellow stars mark instructional strategies related to relevance.

While all teachers in the study held an orientation to teaching and learning with and about data that data literacy needs to be explicitly grounded in students’ lives, Mary was the most affected by this orientation and struggled throughout the Workshop Series with not only how to manifest this orientation in her teaching, but even how to articulate
it sometimes, as seen in this comment from session 9 during the final reflections on the Workshop Series.

I've been struggling a lot with this, it's like what, to drill down to like why, I mean they don't, in my particular class I realized the... at my end and I think it's gotten me really stuck because there's so many great lessons and there's so many things I know that they would benefit from knowing but it's not really relevant to their lives necessarily. And it’s never wrong to know things, knowledge for knowledge’s is sake is a good thing, but given that they’re not going to invest their time and energy unless they find a connection to something… what's the connection to what matters to them?

In this quote, Mary was articulating how the idea of relevance to students’ lives had been a preoccupation and also a barrier for her in the classroom. This quote also demonstrated how this orientation to teaching and learning data created a through line to knowledge about students’ understanding of data and strategies for teaching with and about data and how the components of PCK are all interconnected. Here, Mary acknowledged her knowledge of students’ motivations around data literacy and how that had affected her orientation to teaching with and about data and led to questions about the best strategies for teaching “the things they would benefit from knowing.” These connections are also demonstrated in the concept map in Figure 10 where examples of knowledge of student understanding (designated by red stars) are interconnected with knowledge of instructional strategies (designated by yellow stars).

5.1.1.2 The Tools of Data Literacy are Powerful. The second theme that emerged for conceptions of teaching and learning about data, is the idea that not only is data literacy important and should be explicitly taught, but within that, specifically the
tools of data literacy need to be explicitly taught. While the teachers had varying levels of experience and comfort with Google Sheets coming into the PD experience, they all left with a deep appreciation for the power of it as a tool and a newly fortified commitment to explicitly teaching students how to use it. In his final interview, Will spoke about how surprised he was that the students had never used Google Sheets before even though they were seniors, and the need to address that.

So being able to teach students the process and the understanding of the information and then how to make the analytical much easier is a huge skill that students need. You heard today; they'd never used Sheets before this unit. And that blew me out of the water, saying, "What do you mean? You're seniors! How have you never used Excel or Sheets or anything?" And it was just incredible, so clearly we need to address that more.

Here Will was articulating his conception that the tools of data literacy, here Google Sheets, is a skill that needs to be explicitly address for students. While Mary’s students had the opposite proficiency with Google Sheets going into the unit as Will’s did, she still highlighted this conception of the tools as being important, saying in her final interview,

And honestly, they had done the Google Sheets, most of them. This was not new to them, but it was much different. They were learning and, they were layering on skills that they had. So, I do feel like there's a lot of power in helping kids gain proficiency in the tools that help them make a complicated world, less complicated.

Again, this quote articulated a conception held by all four teachers by the end of the Workshop Series about the power of Google Sheets as a tool and the belief that it should be explicitly taught to students. In their final interviews, both Hallie and Manisha spoke about their intention to bring Google Sheets into their classroom in a more robust way.
going forward. In response to the question: *What were some of your biggest takeaways from the workshops?* Hallie said, “Well, I think one of the big takeaways is earlier and more often, and any excuse to collect data and any excuse to manipulate it and create visualization of any kind, like starting right off the bat using the tools.” Manisha had a very similar response to the question about plans for teaching data literacy in the future, saying, “Well, clearly I'm going to use the Google Sheets tool a lot more. From having the data, transferring the data to graphing the data is clearly something that is not a maybe, it's a must have.” These two responses showed how collective the thinking among the group became by the end of the Workshop Series. The way they expressed this orientation to teaching and learning about data was nearly identical, down to the components of the investigation process: collection, analysis, and visualization and how the tools of data literacy can support that process.

The strength of this orientation was evident in the fact that all four teachers individually brought it up as a vital component of teaching data literacy in their interviews without the immediate influence of the collective PCK around the conception. However, this display of individual PCK built on a collective belief in the power of Google Sheets and the need to explicitly teach it. Similar to the previous theme, this theme also manifested in the concept map created by teachers on the final day. Figure 11 shows a magnified image of the bottom left corner of the concept map with some of the surrounding ideas hidden in order to emphasize the posts related to the concept of teaching the tools explicitly. Ideas which explicitly mention Google Sheets are marked with an orange star. The Theme boxes are colored green. While Manisha added the Theme box about Modelling the Process, I added the Theme box titled “Explicitly
Teaching the Tools (Google Sheets) as a theme based on the comment from Hallie during the discussion phase of the session in which, when asked if there were any other themes that were jumping out at people, Hallie replied, “Using Google Spreadsheet as a tool and it's underrated.” Building on that comment, others jumped in with Mary following up to say, “I mean it's a really powerful tool, so I feel like these ideas of how versatile some tools are and that of course you need to be shown how… the visualizing of data that Google Sheets can do is the twist in this particular unit for me.” Then Manisha responded directly to Mary saying,

I think what I’m really just alluding to is the fact that it's really clear that we need to explicitly teach this. Whether we use direct instruction, explicit instruction, but it needs to be shown and it's been stated by a lot of people here, step by step, and the value of it and [Mary] you just reinforce that, but that explicit need is definitely there and as [Mary] says, like the tool is the essence of it.

The concept map also highlighted that this conception of teaching with and about data literacy, that the tools of data literacy need to be explicitly taught, became another through line to additional components of PCK. The purple stars mark examples of an explicit instructional strategy “Demonstrating mistakes” that will be discussed further in section 5.3.5. Both Will’s comment from his interview about how surprised he was that students’ did not have prior experience in Google Sheets and Hallie’s post in the concept map about her difficulty guiding students to acquire the Google Sheets process skills, showed how this orientation was connected to knowledge about students’ prior experiences and skills when it comes to data. The connections in the concept map between the idea that explicitly teaching the tools is important and one of the strategies
for doing so, modeling the process, was also evident. As such, the belief that the tools of data literacy are powerful and need to be explicitly taught was a driving factor for other components of PCK for data literacy.

**Figure 11**

*Concept Map on PCK for Data Literacy: Google Sheets Section*

*Note:* The green boxes depict themes added during the conversation about the ideas. The orange stars mark ideas that explicitly mention Google Sheets and the purple stars mark explicit examples of an instructional strategy addressed later in the chapter.
5.1.2 Conceptions of the Nature of Data

While the nature of science and beliefs held by teachers about the nature of science is an entire field of study, the nature of beliefs held by teachers specifically about data is less well documented. During this research project, there was a lot of discussion around two main components that fall under the category of the nature of data: a) what actually counts as data and how it can be bounded, and b) the power that data has to mislead and manipulate.

5.1.2.1 Defining Data. In the seventh session of the Workshop Series, the group engaged in conversation about the definition of data. Though this theme didn’t fully manifest until the last third of the Workshop Series intervention, the conversation was robust, and spawned some deep reflection which then carried into the remaining sessions. The conversation initiated in response to brainstorming big ideas for the data literacy content knowledge component of complexity of data sets. The group had settled in the previous session on “Data can take many different forms” as the big idea they wanted to build a CoRe on. While discussing the first box of the CoRe template: what you intend the students to learn from this idea, Mary launched the group into a discussion about the definition of data. The beginning of the conversation between Mary, Hallie and me as the facilitator is presented in full here to frame the conversation that ensued.

Mary: Can you can you explain… I'm not 100% sure I know what you mean by data can take many forms. I know this is… I'm sorry.
Facilitator: No, that's fine! What are, how are you understanding it right now [Mary]?
Mary: Well So data can be a bunch of numbers that are like… I don't know, I mean that's why, I guess that's why I'm asking. It feels like um… So, data can be graphs I guess data can be charts, data can be tables...
Hallie: Well, data doesn't have to be numbers.
Mary: That’s true.
Hallie: it can be qualitative things.
Mary: So, Okay So what you, so you're basically saying the big idea is that
data is information and information is expressed in many ways…
[crosstalk]
Hallie: [crosstalk] quantitative and or qualitative
Mary: Right okay.
Facilitator: So, is any information data?
Mary: No. I'll say no.
Facilitator: Okay, so how do we… I mean so [Mary] says no.
Mary: Well so data is… some is information that was collected that… okay,
  So I think we should define data.

In this initial exchange, Mary was the one who expressed uncertainty about the definition of data as a concept and Hallie responded with more conviction about what data can be. However, as the conversation went on, others reflected on Mary’s question and expressed their own uncertainty. After a pause, Hallie came back to the question about whether all information is data saying, “Well, I'm thinking about it right now trying to find anti example you know. So, what would be my counter example to information that is not data and I don't have anything yet.” She followed this thought up with, “I wouldn't call superstitions information and I wouldn't call faith information” and Mary agreed with this example of what data isn’t and added a characteristic that might help define data saying, “I mean, there's so much that people believe that isn't actually data. So, data is possibly reproducible or… Data is measurable. I mean I'm not saying it's not messy, but it is reproducible.”

While Mary and Hallie were having this conversation about what data and information is and isn’t, both Manisha and Will had been quietly using the internet to
help them process the question of how to define data because neither felt confident in
providing a definition on their own. Manisha shared what she found first saying,

This is interesting because I just looked it up and according to philosophy
[data is] things known or assumed as facts making the basis of reasoning or
calculation. And then there's another one that’s a simple definition and data
is defined as facts or figures or information that's stored or used by a
computer. I don't know about that, in my interpretation.

In the above quote Manisha turned to the internet to help her reflect on a definition for
data but then rejected part of what she found, suggesting that though she didn’t
necessarily have the words to describe it, she clearly carried a belief about the definition
of data. Through Mary’s questioning the definition of data, the participants all grappled
with their own understanding of the definition of data. Though there was no cohesive
conclusion reached among the group about a definitive definition, the understanding that
each of the participants had a belief about what counted or didn’t count as data
emphasized this conception as an important piece of teachers’ orientation to teaching
with and about data through a conception about the nature of data.

Later in the session, as participants continued to work on filling out the CoRe for
the big idea that data can take many forms, Hallie connected the conversation they had
had earlier about defining data for themselves, to how that would affect their students,
saying, “If you don't know what it is and that would be data in general, then that's already
very alienating and it's going to be tough to develop a relationship to something that's
undefinable for you.” The idea of defining data became another through line that
connected teachers’ orientations to teaching with and about data to other components of
PCK for data literacy as the teachers continued to consider in this session, and going
forward, how their students’ understanding of data also contained a component of defining and creating boundaries around what counts as data and what doesn’t.

5.1.2.2 The Power of Data to Manipulate. The second theme for teachers’ conception of the nature of data was ubiquitous throughout all the conversations that teachers had, which was a predisposition toward a somewhat pessimistic view of the power of data for manipulation and the role it plays in spreading misunderstanding. This orientation to the nature of data was apparent from the very first session of the Workshop series during which one of the two big ideas highlighted by the participants to build a CoRe on was the simple statement that “Data is biased.” Mary and Manisha in particular held strong views about the unreliability of data and the way that it is used to manipulate people and were driving forces in settling the group on that particular choice of big idea. The conversation that led to an agreement on data is biased as a big idea is included here.

Manisha: So, one critical idea is, what organizational tools help make meaning of data, because with artificial intelligence and all this machine learning, it's already creating a sense of… inference in itself before we can even get there.
Mary: Right, it's private inference. It's like the algorithms are proprietary so you don't even know why you can't get parole you can't get parole because somebody… you don't even know. I feel like that's another like the sort of the data, big data, how big data is being used to influence us.
Will: Computer programs are only as good as the information you put in and what bias has been programmed into those computer programs, to give you the response you're looking for.
Mary: Why it's so important is that data seems like it's not biased. Which is exactly what the point is, it's like data is as biased… data is biased, maybe that's the big idea.
Manisha: Oh, I like that, data is biased.
Will: The bias depends on, it depends on the reader of the information.
Hallie: There's no perfect data set.
In this exchange, it was Manisha who first set the group on the path toward considering the uncertainties of data and then Mary who directed the group towards thinking about manipulation by highlighting “how big data can be used to influence us.” While Will and Hallie were active participants in this exchange, their orientation towards this particular conception of the nature of data was less strong. Manisha however restated later in the same session the depth of her predisposition towards seeing data as inherently biased and used sometimes with less savory intentions, saying,

Sometimes I wonder if science is as objective as it hopes to be, you know I’ve been looking at the Covid data right now…I’ve lost my respect for the scientific community actually, I really have. And the political agenda, the data is not objective at all. It’s a hodgepodge, it’s a mess, and whatever the political agenda is.

Here, Manisha demonstrated that she doesn’t view data that she encounters in the world as objective. She linked this epistemological stance to the Covid pandemic and the way she saw data as having been manipulated for political gain. While in session 1, she seemed to be presenting this as a new orientation for her, by session 7, she was still firmly grounded in this belief, bringing up her same idea about Covid data and its unreliability during the conversation about the definition of data.

Data can be manipulated, so how different is that from belief? It’s just to reinforce the belief so it’s really tricky. Even what we’re now talking about as data, I’m questioning, especially after this Covid experience. We’re living in in these bubbles where these algorithms are feeding us things that are reinforcing our beliefs so it’s very… we talk about data being messy, but I think our world is pretty messy.
Manisha’s orientation to the conception of the nature of data as being inherently untrustworthy to the point of being indistinguishable from a belief was grounded in both her experience of the Covid pandemic and her mistrust of social media algorithms.

Mary’s predisposition towards skepticism about the objectivity of data was even more pessimistic than Manisha’s. While Manisha viewed data as untrustworthy because of the biases inherent to its existence, Mary repeatedly described her view of data as a tool for active manipulation, saying in session 5,

I think people really do try to manipulate data to tell a marketing story to sell you stuff to take advantage of you to make a point. Some of that's malicious some of its ignorance, but there is a lot of like really questionable data that looks really fancy so we believe it.

Here Mary explicitly pronounced her orientation to the nature of data as a tool for manipulation. She did acknowledge that it isn’t always with malicious intent, but that sometimes it is. She brought manipulation up multiple times throughout the sessions, saying in session 6,

So, there's using data and there's misusing data and there's understanding data, so I feel like there's lots of real life examples of ways that data is used to make decisions about you, for you, without you.” and in session 8, “I feel like that’s the big idea is that you are being manipulated all the freaking time because you're not really paying attention.

While Mary was fixated on the idea that data is used to manipulate, and Manisha was predisposed to mistrusting data, it was Hallie that connected the idea of “data is biased” concretely to the content of data literacy and the complexity of data sets and the fact that all data is in context. During the conversation in session 7 about the definition of
data, Will suggested a definition he found on the internet which he shared as, “data is the raw information unprocessed and then once it’s processed and interpreted it becomes information.” Hallie responded by immediately connecting this definition to the idea of manipulation that Mary had repeatedly mentioned. Hallie reflected that,

I think that's the room for manipulation and the rest of it right, because one of the things that happens with data is that you're presented the numbers with missing pieces of the context. Sometimes I'll read an article and I'll look to try to figure out what some of the context of the thing is and they've maybe even intentionally left some details out of the paper. The slick things that folks do with data to create information is where one must be suspicious. We're looking at the curating of data. It's messy so you have to clean it up. Well which pieces do you clean out? The ones you don't like.

This connection by Hallie between the way data is analyzed and how that process can lead to manipulation and bias creates a bridge which shows how an orientation to thinking about the nature of data as inherently biased and a tool for manipulation can affect other components of PCK, specifically teachers’ conceptions of the goals and functions of teaching with and about data. As will be discussed in the next section, the orientation to data as a tool for manipulation guided teachers’ goals for teaching data literacy, and as will be discussed in section 5.3.4, one of the primary strategies for teaching data literacy surfaced by teachers was supporting students in interrogating data for context and potential bias which is certainly aligned with this orientation to the nature of data if not a result of it.
5.1.3 Conceptions of the Goals and Functions of Teaching With and About Data.

Through the development of big ideas for different data literacy content areas and the building of CoRes for some of those big ideas, the participants discussed many wide-ranging specific goals for what they wanted students to know and be able to do in relation to data literacy. However, as shown in Figure 2 and discussed in section 2.2.1.1, PCK can be applied at many grain-sizes. In order to build a more holistic representation of the primary conceptions of the goals of teaching data literacy, these concept specific big ideas and objectives were gathered into themes that apply at the topic level of data literacy as a whole. Three themes for participants’ orientation to the goals and function of teaching with and about data emerged: a) Learning to interrogate and question data; b) Learning to tell a compelling story using data; and c) making interdisciplinary connections.

5.1.3.1 Learning to Interrogate and Question Data. One of the primary goals of teaching data literacy expressed by all four teachers was to help students develop the skills and comfort for interrogating and questioning data. They primarily spoke and wrote about this goal in relation to data that the students did not collect themselves but encounter through their everyday lives. This orientation to the goal of teaching data literacy is aligned with and grows from the belief about the nature of data as inherently biased and prone to use for manipulation. As all four teachers held a belief to some degree that data is biased and can be confusingly or even maliciously utilized, it logically follows that they view supporting students in confronting those biases and misinformation as a primary goal.
During session 3, while working on a CoRe for data visualization, and attempting to determine big ideas for data literacy, Manisha and Mary had an exchange about wanting students to ask questions of data visualizations that they might encounter.

Manisha: I think it's something that we really want get kids to get into in terms of not just the design, but what's included what's not. So that they are clear with some sense of integrity, definition of integrity, some sort of standard by which you should really dig into this. It's like okay sure anybody can throw out some data anybody can throw out some pretty, as you said, visualizations and all that but what are the key things we want our kids to guard against or watch out for?

Mary: What questions should they ask when they look at data? What questions should always be… how cynical should they be most of the time, basically.

Manisha: Well, they should be searching for integrity, I think, and what are the factors that are pulling out integrity of data.

Manisha and Mary struggled with how to create a set of standards for students to interrogate the integrity of data and wondered how to support students in being skeptical of data and guarding against data manipulation. Mary’s comment “what questions should they ask when they look at data?” was directly related to the data literacy subject matter component of data in context.

Much of what the teachers emphasized when discussing their orientation that a goal of data literacy should be to interrogate data was related to guiding students to be more aware of the context in which the data was collected. Hallie put this explicitly in session 5 when she reflected, “Part of the probing for that interrogation is like okay what can we find from this? What can we discover about this context and then, what do you still wonder?” Hallie then described a teaching strategy that she thought might be used to support students in developing a skill for interrogation and Mary responded saying,
It seems like the goal of that is that the students should understand that interrogating data is always essential. If somebody is giving you a table giving you a chart or giving you a graph like don't take it at face value, look at it a little bit more carefully.

While the majority of the conversation on the goal of developing data interrogation skills in students was focused on determining the context of data that they encounter externally, in later sessions, participants also connected this goal to building inferences with data that students had collected themselves. In session 6, while reflecting on the complex data set his students had collected with the sensors and in conversation to build a CoRe for complexity in datasets, Will reflected,

I think the biggest thing with data for me is understanding, like everybody said the bias and what lens are you viewing it through and being able to identify how you're interpreting it and saying you know okay well, is this data we collected important, why is this data important to me? How could it be important to somebody else? So, seeing what data that you deem important, why do you deem that important and others unimportant?

In this quote, Will connected the orientation about biased data to biases his own students might have when interrogating the data they collected themselves. When he said “the biggest thing for me” he was referring to the biggest goal he had for his students. The orientation he was showcasing here was towards the goal of teaching data literacy not only to support students in interrogating and questioning data they encounter in their lives, but also to question and interrogate choices they are making about how they are interpreting their own data and creating the context through which that data will be viewed.
The orientation to one of the primary goals of data literacy being to develop students’ ability to interrogate data was also supported by the CoRes the teachers developed. Of the five CoRes built during Workshop sessions by the participants, three of them included big ideas related to this orientation. For Inference with Data, one of the two big ideas the teachers chose to focus on was “There is no perfect data. All data is biased.” In completing the CoRe for this big idea, in the section titled, “What you intend the students to learn from this idea,” teachers included the following points: 1) It is good and proper to interrogate data; 2) Identifying potential biases, and inaccuracy or issues with data (e.g., insufficient data); 3) Knowing who (and how?) data collected → evaluating sources of data for credibility; and 4) What are the agendas, biases, objectives for the data and collation methods used. All four of these student objectives supported the belief that interrogating data is an important goal for teaching data literacy. The CoRe on Data Visualization included the big idea “What you exclude from a visualization matters as much as what you include” with one of the student objectives reading “To find the value and meaning of the representation as well as its shortcomings.” And finally, during the framing activity for the CoRe on Data in Context, one of the big ideas the participants settled on was “Interrogating data is useful and necessary.”

The second goal for teaching with and about data that the teachers showed an orientation towards involved the next step after data has been interrogated which is how it is communicated.

5.1.3.2 Learning to Tell a Compelling Story Using Data. A second goal for teaching data literacy coalesced around the theme of storytelling. Using the language of
stories and storytelling first appeared in session 3 when teachers were brainstorming big ideas for the CoRe on data visualizations. Mary said,

I feel like another central concept is that you are making choices to tell a story, you know your visualization is emphasizing and editing parts of the, of this, you know your editorial. Decisions about the visualization are sort of key to the effectiveness of telling the story that you're trying to tell, that I mean visualization matters in in a pretty powerful way not necessarily to the science of it, but to the communication.

Here Mary was thinking about a big idea not for interrogating existing data but for creating a data visualization out of a raw data set and using it to communicate a story. After some finessing and Manisha adding the idea of a temporal component, one of the big ideas that was eventually used in the CoRe for Data Visualization was “Data visualizations tell a particular story about a set of data at a particular point in time” and in the section titled “Why it is important for students to know this” teachers wrote “Data representation includes identifiable design choices, and a challenge in the process of making models is making meaningful choices.” This supported the idea that teachers were thinking about supporting students in being able to communicate data through data visualizations and representation as being a goal of learning data literacy.

The language of telling stories was used again by Will while the group was brainstorming big ideas for data in context during session 5. He said,

I think my biggest goal and my objectives are to have the students take out the data that speaks to them, that makes the story for them. Yes you're 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. That's what I want them to understand is you can use data to, if you do it properly, to support any position that you want as long as you have the data set to support that. There
is no right answer there's no wrong it's, what can you use the data sources that you have to justify and support your position.

Will explicitly stated that he viewed being able to tell a story with data as one of his primary goals for teaching data literacy. He also implicitly built on his knowledge of students’ understanding of data with his phrasing of no right or wrong answer and the fact that students are often seeking a “right” answer, a component of PCK which will be discussed further in section 5.2.2.

Hallie followed up Will’s comment by connecting the idea of telling any story with data by choosing to ignore the “whole other world of data” to learning to ask questions about data by thinking about what else could be learned from that data you’ve chosen to ignore in telling your story and bringing it back to her idea about encouraging students to wonder, saying, “A lot of that falls within interrogating data sets. But it's not so specific that it's even like what do you still wonder about this dataset it's about what do you still wonder about the world.” This comment tied into her comment from earlier in the session about interrogating data to discover what you are still wondering. This highlighted the fact that though these orientations are distinct themes in many ways, they are also tightly connected.

The theme of supporting students in being able to tell stories with data continued to surface as a primary goal for teaching data literacy. In session 7, while discussing the CoRe for complexity of data sets Will said,

Students always want to wrap things up nice and tight and be as quick and concise as possible and be done with it, but I think it's important that they learn their data tells a story and what story are you trying to convey with the information that you have is important
Additionally, in the final session, in reflecting on their primary takeaways from the Workshop Series, Mary said, “I want them to talk about visualizing telling a story with data and that's the part that I want them to, for them to feel like they are they're going to become storytellers and they're going to be change makers.” In this final comment from Mary, she connected the orientation towards learning to tell a compelling story with data as a goal of teaching with and about data to inspiring student to become storytellers and changemakers. She stated her orientation strongly not just as a goal for learning but a goal for identity.

5.1.3.3 Making Interdisciplinary Connections. The final theme for conceptions of the goals and function of teaching with and about data is the role data literacy plays in STEM integration. While much of the STEM integration was implicit in the way teachers discussed and reflected on the role of technological tools and math skills in the teaching of data literacy, there were moments when it became explicit that served to illuminate the underlying belief about using data as a way to address STEM integration.

During session 9 of the Workshop Series, the teachers had an explicit discussion about the connections between math and science when working with data, specifically, around the use of vocabulary and symbols and how the same concepts are often not presented in a coherent way from one discipline to the next, or even from one subject to the next. Hallie summarized this conversation well in the following quote.

So, fractions are the same thing as division. Like, there are so many different ways to symbolize because this is the thing with symbolic language. Once you understand the meaning, then just translating into a different set of symbols is fine, but if you don't actually stop and say hey, this is just a
different set of symbols that means the same thing that you already know…
that’s what is not happening.

Hallie emphasized why the disconnect that occurs between math teachers and science teachers in school can result in confusion for students because they don’t always make the connections between a concept they learned in math and working with data in science. This highlighted the importance of making those connections. Mary wrapped up that train of conversation with an explicit statement of this orientation, saying, “I feel like the interdisciplinary opportunities for all of this data stuff is, they’re so powerful, they’re huge.”

Even when the teachers were not explicitly discussing the cross-disciplinary nature of data literacy, it was implicit in the way they spoke about their students’ math skills and how their students’ prior knowledge in both math and technology played a major role in their understanding of data literacy. Will reflected on this in his interview when asked if there were any specific skills that he felt students needed to have in order to work effectively with data. His response went straight to the students’ math skills. He said,

They have to be strong in math. Absolutely. A lot of my class time was spent on why... What is the mean? What is the median? And why are they important? And those are, in my mind, come first and foremost in math, before science, and science kind of ties in with it, but I always thought they were introduced first in basic algebra. And just the cross-curricular nature of math is that it's essentially everywhere, but that was huge, not knowing that. And seeing that highlighted that this is an area that needs to be addressed earlier.
This statement summarizes a theme throughout the sessions of teachers lamenting their students’ math skills and how their “fear of numbers” made working with data more difficult. However, there was also discussion of how data and the technology of data, specifically Google Sheets, could actually help make math more accessible for students. Will commented on this as well during the discussion of his first video in session 4. He said,

> When I first introduced [Google Sheets] to them, I explained, this is how I do my household budget at home. It's useful beyond schoolwork. So, they were really excited to hear that you know it's not just a school thing and it's not just science, we use the math tool in music, history, it’s cross curricular.

Will was discussing here how he introduced Google Sheets to the students as a tool that can be used for many different types of data in different disciplines and for non-school related usage and how making that connection made students more excited to learn how to use the tool.

The orientation to believing that data can be used to make interdisciplinary connections, though not as explicit as the other themes discussed here, had vast implicit power through many of the other components of PCK for data literacy, especially in knowledge of students’ interest and motivation for learning with and about data and the strategies for making data accessible and tangible for students. The components of PCK that form knowledge of students’ understanding of data are presented next.
5.2 Knowledge of Students’ Understanding of Data

Having a knowledge of students’ understanding of data includes how students learn with and about data and their interest and motivations to do so. It also includes knowledge of their preconceptions about data including about its purpose, how to use it, and where it comes from and where those preconceptions might lead to learning difficulties when developing data literacy. These concepts have been organized here into two primary components of knowledge of students’ understanding of data: 1) their interest in and motivation to learn data literacy, and 2) their orientation to data. This second category echoes in some ways the teachers’ orientation to data literacy presented in the previous section in that it includes students’ preconceptions about the definition and goals of data, but it also focuses on how those preconceptions and orientations might present barriers to developing their data literacy.

5.2.1 Students’ Interest in and Motivation to Learn Data Literacy

This was the number one theme among the components of PCK for data literacy in that it surfaced in almost every session of the Workshop Series and played a significant role in teachers’ implementation plans and execution. The teachers’ knowledge about students’ interest in and motivation to learn data literacy was a driving force which had a strong effect on both their own orientation to data literacy as discussed in the previous section, and their strategies for teaching data literacy as will be discussed in the next section.

The concept of students’ motivation to develop their data literacy first came up in session 2 when participants were reflecting on the RCM and its relationship to the subject
matter components of data literacy. Manisha asked, “How do extrinsic and intrinsic motivators fit into the paradigm?” Mary immediately jumped on this question, following it with,

So, I really appreciate that the motivation word got brought up here because I feel like a lot of what we are all struggling to do is find intrinsic motivators that connect our students to this work, knowing that there's multiple intrinsic motivators for each kid and some kids aren't, you know they're not there yet. They're just at the like how many points does this have?

The struggle to motivate students seemed to be a factor that the teachers lamented across all subject matter and topics. Hallie gave it a name in a later session, “minimaxing: doing the minimum amount of work to pass.” Hallie also commented during the same session that “sometimes it’s just hard to maintain attention.” This was a reflection that was also supported by classroom observations and the field notes from those visits. Anecdotally, I was constantly noticing and commenting on how disengaged students were in general, on their phones, playing video games, watching YouTube on their laptops, or straight up sleeping. This sort of behavior was observed less in Will’s class than the other three but was still present. This was especially troubling given the Bioinformatics PBL’s attempt to engage critical relevant pedagogy and will be addressed further in the Discussion Chapter.

While interest and motivation were a challenge across the board for students, the teachers were particularly concerned with students’ motivations for learning data literacy. In session 7, while working on the CoRe for complexity of datasets, Mary commented, “Well, I mean the difficulty is all of it seems kind of boring to students, it’s like why do I care about this, you know, like, next.” and Hallie quickly followed up with “I really feel
like it seems really impersonal.” With these two comments Mary and Hallie were expressing knowledge of students’ need to see personal connection to what they are learning, and that data literacy does not inherently have that connection for students. However, teachers also discussed how when the students did see the personal connection, then their interest was engaged. During session 3 Hallie commented on how her students became more interested when they were able to connect weather, and through that, air quality, to their own personal health saying,

I mentioned that you can actually get air quality forecasts like you can for weather and it will actually recommendations for like people who are sensitive to pollutants should like not go outside today. And that they all, they found interesting and many of my students started looking for those and found them and they were like ‘oh okay, today, the air quality is fair.’ So, I think that they will find it interesting like how does one come about the conclusion of the air quality is fair. They see the practical application, because every single one of them either is asthmatic themself or someone who's a first degree relative mother, father sister brother, child, is asthmatic so in among my 66 biology students 100% could identify immediately with asthma.

In this reflection, Hallie was discussing how her students became more interested when she was able to not only connect the idea of air quality to asthma, which she acknowledged is a health issue that effects all of her students in one way or another, but she also connected the concept to their cell phones, and apps that they could actively use and explore. These personal and tangible connections increased students’ interest. In order to make those connections, Hallie was displaying knowledge of what would interest and engage her students about air quality data.

Similarly, Will spoke about how when students saw personal connection to the data, they were more engaged and more interested in developing their data literacy.
Reflecting during session 5 on the process of actually using the sensors to collect data with his students Will said,

Now we’re getting into the fun, we’re out of like the background knowledge and getting to the actual meat and potatoes of it, now they’re making the connections and seeing. And there were actually some of them that are going backwards to look at what they submitted in the past to say okay now I see why I had to do that.

Will referenced how his students actually went back to artifacts from the previous lessons about using Google Sheets and building data literacy skills once they saw a personal connection as to why those skills could be useful to them. He referred to the data collection piece where the students got to collect their own data as the “meat and potatoes” and the “fun” part which is a reflection of him observing that his students were having more fun with that part of the PBL unit.

Finally, connecting back to Mary’s reflection from session 9 that was first presented in section 5.1.1.1 to show how this knowledge of students’ motivation and interest shaped teachers’ orientation to teaching with and about data and the need to ground it in relevance to students’ lives.

I think it's gotten me really stuck because there's so many great lessons and there's so many things I know that they would benefit from knowing but it's not really relevant to their lives necessarily. And it’s never wrong to know things, knowledge for knowledge’s is sake is a good thing, but given that they’re not going to invest their time and energy unless they find a connection to something… what's the connection to what matters to them?

The question of what the connection is between the subject matter of data literacy and what matters to the students is one that all of the teachers sought to answer throughout the
process of implementation. Their knowledge of their students and what interests and motivates them was sometimes deep enough to lead to engagement and sometime incomplete in a way that caused barriers for implementation. However, as teachers strove to engage students with developing their data literacy, it was an important component of PCK to recognize that data literacy is not inherently motivating to students and needs to be connected to their personal experiences to be of interest.

5.2.2 Students’ Orientation to Data

Another important component of PCK for data literacy is understanding students’ preconceptions about data, how they relate to it, what they think it is, what it can be used for, and what role it plays in their lives. As teachers were discussing and writing about their knowledge of students’ orientation to and preconceptions about data, there were five primary themes which arose that were uniquely relevant to teaching data literacy in a way that was different from students’ orientation to science, or at least the biological and environmental science that the teachers in this study were trained and experienced in teaching. The five themes were: a) students have a fear of numbers; b) students have a constrained conception of what data actually is; c) students struggle with the ambiguity of authentic, real-world data; d) students are not used to questioning data; and e) students have a predisposition to and a tendency towards building arguments from assumptions rather than data-based inferences. Each of these themes were supported with examples of teachers’ PCK in these areas in the following sections.

5.2.2.1 Students Have a Fear of Numbers. The knowledge that their students often have a fear of numbers was well formed for all four of the teachers. This seemed to
stem primarily from their knowledge of students’ prior experience with and knowledge of math in general. The topic of students’ “math grade levels” was brought up often by all four of the teachers, with all of them acknowledging that their students were all “significantly” below “grade level.” Manisha in particular was hyper-aware of her students’ math and literacy levels, which spoke to her role and experience as a special education teacher working with student with IEPs. She continued to reflect on this in her interview at the end of the Workshop Series, saying,

I deal with students with IEPs that are below grade level in terms of math and literacy. So, their numeracy comfort is relatively well, it's not on grade level. So therefore, it makes it a smaller pull to say, so you can handle this number, but you may not be able to handle a fraction or something like that.

Students’ comfort with numeracy was a concept that Manisha also brought up during the very first Workshop session while working on the CoRe for Inference with Data. She added “Comfort with numeracy” to the section of the CoRe labeled “Difficulties/limitations connected with teaching this idea.” Other teachers added to this, with Hallie writing “Math phobia can turn students off looking into data” and Mary noting “Students’ abilities to understand beyond the numbers.” These are all different ways of recognizing that students’ relationship to numbers is a barrier to developing their data literacy. The thread of this theme ran through all of the sessions, culminating in the final session in building the concept map for PCK for data literacy (Figure 9) a section of which is displayed in Figure 12.

Figure 12 shows the section of the concept map depicting the area that displays the different thoughts connected to this theme. The theme title (green box) was actually
suggested by me, the facilitator during the discussion about the concept map, and then the
participants built on it to fill in the description. The beginning excerpt of that
collection went,

Facilitator: I think there's all of these pieces that really, like there is this, students have this fear of numbers.
Mary: They feel stupid around them.
Will: Especially in the sciences, where you know you're going from the metric system to the you know standard system and they don't convert very easily so they get overwhelmed and give up.
Hallie: One of the things we don't do is actually teach students, not just like what some of those symbols mean but, like the actual meaning of the symbols, like, division and fractions are the same thing, it's the same math and we don't teach that explicitly. We just figured that they'll figure out that fractions are division.

The comments the teachers made outlined this theme and made it more robust by providing reasons why students have a fear of numbers. This is a strong component of PCK because it displays knowledge not only of students’ preconceptions of data and specifically numbers (highlighted by red stars in Figure 12), but it also identifies some of the reasons why they might have those conceptions which opens a path to addressing the conceptions and shifting them. Will again connected this component to relevance to students’ lives and offered one instructional strategy, (highlighted by yellow stars in Figure 12), of enlightening students to the data all around them. Later in the above conversation, Will acknowledged that addressing this component of students’ orientation toward data might involve stronger connections with math teachers saying, “We have to sit down with the math teachers and say okay, what are you teaching and how are you teaching this so that I can help my students understand the language that I'm speaking.”
While Fear of Numbers was its own unique theme, there was another theme that surfaced that was related to and emerged from it, which is the knowledge that part of the reason why students’ fear of numbers is a barrier to teaching data literacy is that they equate data with numbers and don’t see the wider definition and possibility of data.

5.2.2.2 Students Have Constrained Conception of Data. Teachers’ understanding of how students often define data grew later in the Workshop Series as a
direct result of their conversation about their own understanding of the definition of data. As they talked about how they conceived of data and what it meant to them and how they put constraints on it, they also built on their knowledge of students’ fear of numbers and hesitancy around data to articulate that how students define data can be a barrier to learning. This is summarized nicely by a quote from Hallie, presented previously in section 5.1.2.1 about teachers’ own definition of data, but extended here. Hallie said,

“So, if you don't know what it is and that would be data in general, then that's already very alienating and it's going to be tough to develop a relationship to something that's undefinable for you. So, like if you think it's just numbers and you're not that wild about numbers, then it's you know once again going to be alienating. It's when you see that so much of the things that we all take for granted are also data that you can start to develop a relationship to data and then you know, find some confidence in your own ability to interpret it.

In this quote, Hallie connected the conversation the teachers had about the definition of data to her knowledge that students struggle with data to conclude that one of the reasons students struggle with data might be that they don’t understand what it is, or that they have a constrained preconception of it as being “just numbers.” She then took the idea a step further and reflected on how if students’ conception of data could be widened to include “the things that we all take for granted” then they could overcome the barrier of fear of numbers because data would no longer be just about numbers, it would become personal.

Manisha did mention students’ conception of data briefly in session 2 while reflecting on the example video of a previous teacher implementing a data literacy lesson saying, “He is, there's an assumption he made that the students know the meaning and the
significance of the data.” However, this comment was not picked up by the other teachers and the thread of it was lost until the later sessions, after the session 7 conversation about the definition of data. However, once the teachers had engaged in that reflective conversation, the idea that students’ preconceptions about the very nature of data needed to be acknowledged and addressed began to coalesce into a component of PCK for data literacy. In session 8 Will reflected that “connecting [data] to their personal life is where you can make real gains.” And then went on to explain an activity that could make these connections was to ask students to discuss how they make decisions about getting to school in the morning like how early to get up and which route to take. He explained that “This is data collection that they don’t even realize they are doing on a personal level and the processing of it happens naturally. Getting the students to see that they are surrounded by data challenges them.” Will was expressing his knowledge that students don’t see data as being a ubiquitous part of the world and of their personal experiences and that guiding them to see that could challenge their preconceptions.

While this theme did not surface explicitly in the concept map created in the final session, there were four posts in particular (Shown in Figure 13 with their original connections to each other but disconnected from other posts) which represented teachers’ thinking around this component of PCK. Three of these posts also appeared in the top right section of Figure 12, highlighting the interconnection between this component of PCK and the previous component of Fear of Numbers. Will again restated his knowledge that students don’t connect to or even acknowledge the data all around them in their lives. He clarified this even more with a second post which specified that students’
preconception of data is often “scientific and math stuff,” relegated to science and math classrooms rather than present in the world around them.

**Figure 13**
*Concept Map on PCK for Data Literacy: Data Conceptions Ideas*

*Note:* Posts have been removed from the larger map but with their original connections to each other intact.

Hallie and Manisha both added to the knowledge of students’ preconceptions about data with additional specification about the way those conceptions can be constrained. Hallie noted that students conceptions of data are constrained in grain size in that they often think of data disconnected from the larger context in which it was collected. Manisha added that students are constrained as well by time, conceiving of data as static rather than fluid and part of an ongoing process of collection and analysis. This post was a compaction of a comment she made in the previous session about the same idea reflecting,

There's this idea of the end versus, you know, a static place that you stop for a minute and analyze, and then it continues, and it changes, and everything changes. This idea of seeing change and interacting within that change, that's something… that we need to anchor in our kids.
This longer statement by Manisha demonstrated that she was connecting her knowledge of students’ conception of data as static and constrained in time to a need to support students in developing a stronger comfort with seeing and interacting with change. This connection lead into the next theme among teachers’ knowledge of students’ conception of data, which is their struggle with uncertainty.

5.2.2.3 Students Struggle with the Ambiguity of Real-world, Authentic Data.

In a continuation of her statement from session 8 presented in the previous paragraph, Manisha connected student’s discomfort with change to their need for there to be a definitive end with a “correct” answer. Manisha continued,

I think that's a critical piece that kids need to get comfortable with, because we are always talking about results, and this idea of an end. There is no end; there's constant innovation and interaction and things like that. So, first of all it's like moving our kids through that “is this right?” Kids will come into our classroom saying, “is this right?” at the beginning of the year and they'll start again the following year, doing the same thing.

In this extension of her original comment about students being constrained by viewing data as static, Manisha also identified that students want there to be an end to the process of data analysis. They want there to be an answer, “is it right?” and Manisha acknowledged that this preconception of data is not aligned with the reality of data analysis which is that there is no end, it’s a process.

Even when not discussing the temporal and cyclic nature of data collection and analysis, students struggle with the ambiguity of data. Mary attempted to explain this in session 7 stating, “The same data can support a lot of different conclusions turns out, and
that's always been the case, but I don't think non-scientists necessarily realize that that often.” This tied back to the teachers’ and particularly Mary’s epistemological stance about data being a tool for manipulation in that since data can support a lot of different conclusions, those conclusions can be chosen to align with needs and goals in ways that can lead to manipulation. It also speaks to a knowledge that students are not trained scientists and are likely to have the preconception that data can only support one conclusion, that there is a right answer. This struggle with ambiguity of data is an important conception to consider and address as Will pointed out in session 7 explaining.

One thing my students always want is, is it right or wrong, and that’s not the answer that you're looking for. There is no right or wrong, and they have to know that. That's why I always try and teach that data, it’s not one size is correct one size is wrong it's this is what you got.

Here, Will was making a point about his knowledge of students’ tendency to want to know the right answer and connecting that to his own approach to teaching data literacy by explicitly addressing this preconception that students have. Hallie explicitly connected this struggle that students have with data to the way it limits their ability to grow their data literacy during the session 3 CoRe discussion on data visualization. Hallie reflected,

Students can be overly cautious with taking intellectual or design risks, to the point of paralysis. Many won’t start until they know they are doing it “right”. It can feel wishy-washy for students, who are accustomed to “only one of the choices is correct”.

The conception held by students that data leads to one particular answer, or in the case of Hallie’s comment, one particular “correct” visualization, is a barrier as Hallie explains because it can lead students to paralysis where they are unable to even attempt to grow
their knowledge of and comfort with data literacy. While students are predisposed to second guessing their own data and looking for ways to match it to a correct answer, they are not used to questioning data presented to them from external sources, which was the next theme to arise from the participants’ knowledge of students’ understanding of data.

**5.2.2.4 Students Are Not Used to Questioning Data.** 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. As demonstrated in section 5.1.3.1, teachers saw the ability to interrogate and question data as a primary objective for teaching data literacy. The focus on that particular goal, stemmed in part from their knowledge that this is a skill that students are not comfortable with. During session 1, 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.

Manisha also reflected on this understanding of students predisposition to data during her post interview, calling back to the conversation from months prior during
session 3 and presented in section 5.1.3. When asked what skills students would need to be successful at working with data she reflected,

So, the skills that my kids don’t have yet but are going to need is how to navigate, interpret a question, the credibility of information, qualitative and quantitative quickly, and what skills and what questions are they going to ask quickly to say, well, this is valid. And the sense of this is questionable versus this has got much more integrity. It's relative integrity, right?

In this reflection, Manisha harkened back to a conversation with Mary from months early when they discussed a desire for students to have the ability to question data and judge it for its integrity. She explicitly stated in this quote that this ability was one that her students did not yet have. The through line of this knowledge for Manisha and it’s connection to her especially strong orientation to considering the biases of data led her to put a lot of focus on this theme during the concept mapping activity. In Figure 14, which highlights the bottom right section of the full concept map, it can be seen that a lot of her original notes (all ideas colored blue) were collected together and then she added a larger theme card to the concept map which depicted again her focus on wanting to support students in asking questions of data.

While Hallie’s idea wasn’t represented in this section of the concept map, she did make a connection between encouraging students to asks questions and the required state competencies, saying,

For the science competencies, at the end of your lab write up, you have to say, like these are my recommendations for further study. Like what did you notice that you didn't actually draw any conclusions about? What you did gain from this to make a recommendation of like how could you look at this differently?
In this reflection, Hallie displayed her knowledge of curriculum for data literacy which was a component of PCK not explored by this particular study, however, as she brought it up in response to a conversation about supporting students to better ask questions of data, she is also connecting this curriculum knowledge to both knowledge of students’ understanding of data and knowledge of teaching strategies.

**Figure 14**

*Concept Map on PCK for Data Literacy: Interrogating Data Section*

*Note:* A zoomed-in image of the lower right of the full concept map. Notice that most of the ideas are from Manisha

### 5.2.2.5 Students Build Inferences from Assumptions Rather Than Data.

The final theme for knowledge that teachers have about students’ orientation to data is that
they have a tendency to take a self-centered approach to analyzing data which leads them to build conclusions from their own feelings or experiences rather than from the actual data available to them. This theme was mostly elicited through the video reflections that teachers conducted and then enhanced through the discussions that followed. In preparation for session 7, teachers watched two short clips of Manisha’s classroom where she was working with students in building inferences. In her video reflection framework responses on her own video, Manisha noted under students understanding of data, “Assumptions, misconceptions and conclusions - that it is “old information,” based on student unfamiliarity of the Slack app.” This referenced a moment when a student, in looking at a set of data about social media which included Slack as an option as well as other familiar social media sites, assumed that the data must be old because she had never heard of the Slack app. All three of the other teachers also noted this moment in their written reflections and it naturally came up during the discussion about Manisha’s videos.

Mary started the conversation by referencing that moment from the video and explaining that she had noted that she, “thought there was a lot of work done to try to point out that assumptions are get you in trouble. Your assumptions are what we're trying to avoid. We're trying to actually find what do we know and how do we know it.” Here Mary not only acknowledged that students make assumptions, but that those assumptions get them in trouble and that they should potentially be taught to avoid them. Hallie immediately jumped in and disagreed.

I disagree. We don't want to avoid assumptions; we want to acknowledge when we were making them. We call it inference in science right. We are putting the pieces together in our own minds, but that doesn't mean that's how it was in the world. You need to recognize and acknowledge what
you've assumed, and then attempting, maybe to test it if you can. That's kind of what science is.

While Hallie disagreed with what students should be taught to do with their assumptions this came from an agreement with Mary’s knowledge that students do indeed make a lot of assumptions. Mary referenced the same knowledge again near the end of the discussion about the definition of data saying,

People have assumptions and I feel like this is what you were talking about a lot in your lessons with your students [Manisha]. We all make assumptions based on our own personal observations and experiences. That's not data. And I feel like you were trying to make that really clear like you know your personal opinion is not data.

This time Manisha responded immediately to Mary’s reflection about assumptions and the role they play saying, “I kind of like that, because I think that we really need to get our kids away from this idea of opinions. If our lens is clearly on just opinions we're not going to dig underneath it to see the credibility of it.” In this response Manisha dug into why the tendency to make assumptions can be a barrier to student learning, because if they build conclusions from assumptions or opinions, they won’t be motivated to interrogate the data.

Later in session 7 when the participants were collaboratively working on the CoRe for Complexity in Datasets, Hallie added a thought to the section “Knowledge about students’ thinking which influences your teaching of this idea” which read, “Recognizing that students start off with “lame” claims that are sometimes more like assumptions.” Then she added CER (which stands for Claims, Evidence, Reasoning) to the next section titled “Teaching procedures” I, as the facilitator, saw her add these and
verbally called attention to it and asked the other participants what they thought. Mary replied,

She's using her knowledge about students thinking. It feels like you know she recognizes that the claims students come up with are sort of lame starting off, but by having a process that requires them to dig a little deeper they come up with some better ones. So, it's a teaching procedure which is completely because of her knowledge of students’ thinking.

Mary referenced the exact way that teachers’ PCK about students’ understanding and instructional strategies for teaching can interact to support themselves and their students in building robust lessons and classrooms that directly address students’ preconceptions. Indeed, the teachers’ knowledge about student thinking and motivation led them to modify existing strategies and brainstorm new ones for how best to engage their students in developing their data literacy knowledge and skills.

5.3 Knowledge of Strategies for Teaching With and About Data

Knowledge of instructional strategies for teaching with and about data is knowing how to address students’ preconceptions and difficulties and create an environment in which learning can occur. In the previous sections teachers’ own beliefs about and orientations to data were explored as well as what they know about their students beliefs and orientations to data. There was a lot of alignment between those two components of PCK and sometimes knowledge of students’ orientation would shift the teachers’ own orientation to data and sometime examining their own beliefs about data lead them to more concretely conceptualize their students’ beliefs. However, knowledge of their own and their students’ conceptions of and orientations to data are only really
practically useful in how they then guide teachers’ knowledge and choices of strategies to implement in the classroom. During the Workshop Series discussion and CoRe development, five core strategies emerged as useful and important for teaching with and about data. They were: 1) making connections to students’ lives; 2) making data tangible; 3) framing scientific argumentation as storytelling to build inferences; 4) interrogating data by questioning assumptions; and 5) scaffolding through modeling. Though most of these teaching strategies can be generalized to strategies used for teaching other topics in science, through the application of their knowledge of students understanding about data, teachers were able to discuss specific examples for how these strategies could be applied to teaching data literacy.

5.3.1 Making Connections to Students’ Lives

While making connections between content and personal experience is an important tool for basically all learning, it is especially important for data literacy which is not intrinsically motivating for students. Teachers came up with a number of ways to connect data literacy, and specifically air quality data, to students lives. One of the most obvious ones, which is exemplified here is creating connections to the Covid pandemic. In session 4 Will explained,

I just made it relevant to them about Covid. Like, everybody's worried about masks and filtration and we went over the different N95 masks versus the cloth masks and filtration in the school system and in their homes. One of the things one of my students said today, when we talked about using the sensors, he was like well why do we have to do the school monitoring, you know schools are already monitored all the time for air pollution and I was like no they're not. They were amazed to find out that even though we shut
In this example, Will described a strategy he had already used in his classroom to make air quality data collection relevant to his students’ lives by talking about the Covid pandemic and their current lived experiences. He also shared some evidence that this strategy had worked with students and raised their interest levels. In a later session, he again shared a strategy that was specific to this particular project with mobile sensors and air quality which was sending the air quality sensors home with students. He explained, “Each person got their group sensor for a night and they took readings on their way home and on their way back to school the next day, so that there is some more of a personalized approach to it, and it's more relevant to what they're experiencing.” Hallie also supported the benefits of the strategy of having students collect their own data, sharing in session 7, “When students collected their own data, they got excited about seeing it in the spreadsheet. The more personal you can make it the better.” By providing students with data collection tools that supported them in collecting their own data, that data was made personal for them and they saw the connection to their own lives.

The teachers also discussed a number of strategies for making the foundations of data literacy more generally relevant to students lives. In the CoRe from session 8 on Data in Aggregate, Will wrote, “Connecting it to their personal life is where you can make real gains.” When pressed by the facilitator for an example, he shared,

I start off the year by discussing how they determined their fastest route to school, how late they can sleep and make it to school on time, how many hours they can work and still keep good grades, etc. This is data collection that they don’t even realize they are doing on a personal level and the
processing of it happens naturally. Getting the students to see that they are surrounded by data is what I challenge them with early each year.

In this example Will shared a number of different pieces of information that are relevant to students lives that could be used as a hook to support student to begin thinking about data in a more expansive way beyond the constraints they may be predisposed to. By choosing to engage students with data from their own lives that they already use, Will made data more accessible to students which in turn allowed them to find motivation and interest in developing their relationship with it.

Along with sharing strategies they had already used in the classroom, teachers brainstormed ideas for new activities that they could try out to connect data to their students’ lives. Mary provided an example for how she was thinking about connecting data to students’ lives, this time specifically for the data literacy subject matter area of Data in Aggregate, she hypothesized

We could probably come up with ways in to help people understand [data in aggregate]. Like, if I knew what you had for breakfast this morning that’s not going to help me know what you eat. Right? So, if I want to know what you eat how many days, do I have to find? If I asked you for a week is that, well is it a week in the summer, is that different than the winter?

In this example, Mary explained a way to connect teaching about data in aggregate, which involves sample size, to what students eat, which is personal and most likely varies a lot from one student to the next, so this example provides as Mary said “a way in” to hook students into asking question about what is needed to collect data that will lead to a useful aggregate. Another suggested example for how to connect data literacy skills to students’ lives came from Manisha during session 9 when she said, “I think that that
should be taught with money. I think that in the earlier years they need to perhaps have relevance with that kids into as teenagers I think they tend to pick up the dollar thing a lot well.” Hallie agreed right away saying, “Certainly, when I say okay money, then, because that is something that they think about and work with that they do see.” Here Manisha and Hallie were once again using their knowledge of students’ prior experiences and what interests and motivates them to come up with a specific strategy for engaging students in learning about data literacy. Notable here is that none of these examples of strategies mentioned asthma which was hypothesized by our team to be a relevant connection to students’ lives.

5.3.2 Making Data Tangible

While the previous section discussed a number of strategies for creating relevance for students by making connects to their lives, most of the presented strategies were engaged primarily through conversation and still had an element of abstraction about them. Another strategy that teachers discussed throughout the Workshop Series was how to make learning data literacy into a more hands-on experience through the use of tangible examples which would allow students to interact sensorially in the data they collected and analyzed to create not only relevance but a kinetic connection as well. Hallie, who was often looking for practical examples amongst the conversations we were having hypothesized a number of different concrete examples of activities using tangible objects. One example was using bags of different kinds of beans that had been “randomly” mixed and having students determine the best way to display the data visually. A second example was having students measure and graph their wingspan
versus their height and look at the trend line as a way to make outliers tangible. While sharing this she recalled how memorable doing the activity as a high school student had been for her saying, “I did a wingspan versus height thing when I was in 10th grade And Aaron Anderson his wingspan is 10 centimeters longer than he is tall, and my wingspan is three centimeters shorter and everybody else was like within a centimeter.”

Another example that Hallie shared during the session 6 discussion about complexity in data involved pennies and she described the activity she was imagining in detail:

I was making notes about what kinds of data we could have students collect on their own and maybe try to make some meaning of it by visualizing them. One of the things I was thinking is, if you could get like 100 pennies and you could have them like sort them by date and try to figure out how to make a meaningful graph out of that, but there was some dimes and some Canadian pennies in there, and I'm sure that you can find some old pennies that the date is unreadable. They'll figure out that like some of these pennies are unreadable, some of these aren't even pennies, and I think that it would be interesting, then, to when you get messy data later like okay, can you find the dimes in this? What are we just going to toss out because it's meaningless, where the date rubbed off you know, because we can't make meaning of that data.

The pennies example showed how deeply Hallie was thinking about all the components of data complexity and how to translate them into a hands-on tangible activity which could then serve as a metaphor to be referred to later when students were working with less tangible data, like that from the sensors.

While Hallie’s examples were all ways to engage students tangibly and kinesthetically with data more generally, Mary provided a concrete example for how to create a tangible component to the study of air quality, suggesting that students use white
felt to measure smog in the air. The following excerpt exemplified how Mary was figuring out her thoughts around this idea as she was speaking.

Some other way we you know really have it, have a white card that you carried around until it got gray, how long would it take to turn gray? I remember actually one of our teachers, like I don't think they did a lot of, like they posted things on different places, like they glued them to walls, but really we could, staple it to, we can do felt, white felt. We staple it to poles in the neighborhood to see which ones turn gray faster. Like that would be a really easy but tangible air quality assessment that wasn't technical but would be very visually cool.

As she solidified her thoughts, Mary came to a strategy that she was excited about that could help make air quality data collection into something tangible for students so that they would have a hands-on experience that could support the less tangible collection of data into the spreadsheet with the sensors.

Finally, Will took it a step further to connect a tangible, hands-on activity not only to the air quality project, but also tying it back to connecting the data to students’ lives by giving them some choice in the data being collected. He suggested that the students actually engage in building their own sensors, saying,

You could always look into having the students design and build sensors that do different things through Arduinos. That's something that I am considering in the future. Having the students decide which things do we want to measure for air quality and building sensors and maybe having students build different sensors to measure different things that are important to them and then comparing the data and trading off the sensors that they built.

What Will suggested is a great way to bring engineering into the Bioinformatics PBL unit as the one STEM discipline that isn’t really represented in the current focus and content.
Engineering and Making are often excellent ways to engage students with tangible data collection and analysis.

The next section provides some examples for how to support students in figuring out what to do with data once they have it by encouraging them to think of data communication as telling a story and both being able to tell their own story and determine the stories that other people are trying to tell.

5.3.3 Framing Scientific Argumentation as Storytelling in Order to Build Inferences

As discussed in section 5.3.1, learning to tell a compelling story with data was one of the primary goals that the teachers identified for learning data literacy. While teachers used the concept of “storytelling” a lot during their discussions and writing, this was primarily used as a more colloquial and familiar way to talk about building a scientific argument. By framing scientific argumentation as storytelling, teachers made it feel more accessible to themselves and also to their students. There were a number of strategies discussed with varying degrees of specificity for how to do this. Again, Hallie was often the one who came up with the most concrete and well thought through examples. During session 6 she shared a potential activity that she had been thinking about as a way to introduce students to telling stories with data by collecting survey data about planning for prom and giving students different responsibilities for planning different parts of the event and having them figure out which data they needed to tell their story.

They would have to pick through which questions are even relevant to their problem, and so they could filter out like okay, where you live doesn’t matter
for what we're eating. Have a number of dietary restrictions sort of questions that really shouldn't matter for the music. Have “Do you dance or not?” be a question on there, and so like are the preferences of the kids who dance more important than the preferences of the kids who don't dance? Is there any of this data that weighs stronger than others, like everybody else enjoys eating pork, but this kids Muslim so should we have pork or is like is some data like heavier than others?

This was another example activity that Hallie thought might be a useful strategy for engaging students with complexity of data, but she also connected it to supporting them in telling a story, saying later “When they’re looking at [the survey data], here is storytelling around the data and is that what the data says, or is that what they say the data says kind of thing. When we're looking at this data, it can be used to tell different stories.” Hallie connected her strategy for teaching the complexity in data subject matter to the skill of storytelling and of recognizing that data can be used to tell different stories based on the choices the storytellers are making.

In the same session Will also talked about a strategy to support students in telling the story of their data and tied it to his knowledge about students’ desire for a right answer. He said,

It’s just how do you get the students to build upon what they know and what they don’t know and taking it a step further and putting it on them. I tell students I don’t have the answer, and you don’t have the answers either, but let’s use what information we have let’s use the data, the details that we do know and extrapolate from that information what we can use to tell the story that we wanted to tell and not dictating what story it’s going to tell until after you have broken it down into smaller pieces smaller chunks.

Though this strategy is less concrete than Hallie’s, it highlighted the need to give students’ agency over the story they are telling with their data by explicitly explaining to
students that there is no “correct answer” and then supporting them in breaking the data down into “smaller chunks” which then can be used to build a story. This is essentially the process of building a scientific argument but using the language of storytelling.

5.3.3.1 Telling a Story with Data Visualization. Making choices about data in order to tell a story was a primary theme during the building and discussion of the CoRe for Data Visualization during session 3. One of the three big ideas that teachers chose to focus on for the CoRe was “Data visualizations tell a particular story about a set of data at a particular point in time.” This was again a reframing of the idea of using a visualization to enhance a scientific argument into the more familiar language of storytelling. Mary added an activity idea to the CoRe writing, “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 - offer specific ideas for how they might do this - and then have students present their visualizations to the larger group.” Hallie added to this, writing, “Represent a set of data in various ways, and critique the choices along with students. Have students identify design elements, and state how they increase the clarity of the story, or if they don’t.”

These are both examples of the same basic activity: building and then comparing multiple data visualizations to support a scientific argument based on the same set of data, just with different levels of scaffolding and support. Hallie defended her choice to provide students with more support in doing the process collectively first saying,

So, in the past, when I worked with students on making models of things, when we get to the point where it's like okay let's look at this against our criteria, what might improve the clarity of this model and students often say like well color. Okay what's the color going to do? and at first they’re
dumbfounded because they're like well it's just everybody always says that thing needs more color, like color is a good thing. It's like well okay, but for why, if you don't know why you're introducing a design element don't. It’s not simply about pretty. You want it to be pretty when you're done, but pretty by itself is not meaningful if you don't know what the pretty’s accomplished.

Hallie drew on her past experiences teaching data visualizations to explain why students need more support when making choices about how to tell a story with data visualizations. Students often make choices about their data visualizations that are not grounded in the framework of trying to support a scientific argument. Mary added to this saying,

Right, it's really about choices you make to convey information, which is as much a design decision, as it is anything else. You're making choices to tell a story. Your visualization is emphasizing and editing parts of your editorial. Visualization matters in a pretty powerful way not necessarily to the science of it, but to the communication of it.

Through this response Mary made a clear connection between Hallie’s explanation of students’ preconceptions about creating data visualizations and the way that can support or hinder them in telling a compelling story.

5.3.3.2 Using the Claim, Evidence, Reasoning Framework. Another more specific strategy that surfaced for supporting students in developing scientific arguments appeared in the form of the Claims, Evidence, Reasoning (CER) framework. This strategy was discussed in some detail during session 7 when the participants were building a CoRe on complexity. The statement “Using the CER format to challenge students to figure out what their evidence and reasoning are to get more precise and come
“up with more robust claim” was added to the CoRe. During the discussion about the CoRe, Hallie noted,

Well, I've been kind of running with that CER format there. Even when their claim is really, let's say lame, once they're challenged to say what's the evidence and then what's the reasoning is when they identify the evidence and then explain how that evidence means the thing they said they do get much more granular and do a much better job of expressing what it is that they're trying to say.

While the nature of the way participants were collaboratively adding to the Google Doc means that it is unclear whether Hallie added the original statement to the CoRe or not, using the CER format to help students expresses the story they are trying to tell more clearly was a strategy she has already been attempting in her classroom with some success. Will followed up on this explanation by Hallie by connecting CER to telling a story saying,

I think it's important that they learned their data tells a story and what story you are trying to convey with the information that you have is important. You could tell a story it's not black and white, there are many different ways you can translate that and get that across. Have them look at you know, take a step outside, use the CER statements as [Hallie] said. You know challenging them to go back and make their claims more fact based with evidence but allow them to use their creativity.

This exchange between Hallie and Will manifested in the asynchronous video reflection assignment conducted after this session as Hallie recognized in her Video Reflection Framework on Will’s video that he was “using CER to get them to defend the story they were telling.” Connecting an established strategy for teaching scientific argumentation in
science to the goal of supporting students in telling a compelling story with data gave the teachers a concrete tool on which to begin building data literacy with their students.

5.3.4 Interrogating Data by Questioning Assumptions.

The ability to interrogate data was another primary goal of teaching data literacy highlighted by the teachers. While the teachers discussed strategies for reaching this goal with students often, the examples for doing so were less concrete or robust than some of the other strategies and focused primarily on using the questions of data in context to push students through class discussions to interrogate data. This amorphous way of thinking and talking about the strategies for supporting students in interrogating data is summarized in the following quote from Manisha during session 6.

It's like to ask those questions, just think about it. Because once they think about it it might steer them in a different direction to where they thought before. Everything's handed to them, but they aren't questioning is it manipulated or not? The manipulation piece is huge. Who's censoring it? Who chose for you is huge and, before that even got presented to you.

In this passage, Manisha referred to questions about the context, specifically a set of questions related to the who, what, when, where, why, and how of a data set when she mentioned “those questions.” She seemed to be suggesting that simply giving the students the questions and encouraging them to ask them would be enough to get them to think differently about the dataset. Manisha solidified this reflection in the following session while working on the CoRe for Complexity in Data by adding to the Teaching strategies section, “Articles that show different types of data that are used and have them
question it. → What is missing, what context is missing, what types of data could be added.”

Teachers identified “interrogating data” as a potential strategy on their Video Reflections as well. After watching Manisha’s video, both Will and Hallie noted that Manisha was using the strategy of data interrogation didactically with her students, encouraging them to dig deeper and ask questions. Will wrote, “she is using probing questions to lead the students to think about the context: Who was surveyed, where are the teachers from: urban, suburban, rural, different country, another planet?” Hallie similarly wrote, “Interrogating data: What do you notice about this data? What other conclusions… How do you know… What’s the point? What questions do you have about this data?” Both of these reflections showed teachers identifying “asking probing questions” as a strategy to support students in interrogating data and interrogating their original assumptions or conclusions about the data.

In session 3 Mary shared a thought on a more specific way to scaffold students toward those probing questions by first having students simply identify the components of a data visualization before they even begin asking questions. She reflected,

What occurs to me is that, even just as an introductory thing, having students look at a variety of charts and graphs and you know things that look full of facts, that they could just even identify the elements of them, not even you know judging them just identifying them but there's you know, there are axes, there are colors, there are titles.

This reflection by Mary acknowledged that students might not have the experience to immediately jump into interrogating data and might not have the tools to even identify which parts of a data visualization contain information that can be interrogated. She
suggested supporting them in gaining those tools by beginning with simple identification.

This idea caused Hallie to reflect on her own over the next month and off record after a classroom observation share with me an idea she was forming about a strategy to support students in interrogating data visualizations in particular. During our next Workshop series session, which was session 5, I asked her to share her thoughts with the whole group.

5.3.4.1 Thinking About a Visualization as a Text. Hallie began to form an idea for an instructional strategy of unpacking a data visualization 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,

What I was saying is um there's a way of interrogating a text, I think that is possibly unique to looking at data and what I would like, in this lesson seven, to do more of is looking at data from different contexts and have the students sort of interrogate the data and see what they can come up with. If we were reading a short story, we say what’s the background we can assume this character might have had because of the way that they speak or whatever. There is a story that [the data] is telling. So, using a kind of text analysis [the students] can ask questions like who were these people? Why were they asking [these questions]?” If you gave them like paper copies of some of these, of some data representations that they could actually like write on what their questions are, what they think they know, and what they're assuming is true, similar to what they might do with a text in history.

Hallie connected Mary’s idea from the previous session about starting students off with just identifying the different parts of a data visualization to a pedagogical tool from outside traditional science classes to articulate a specific strategy for supporting students in learning to interrogate data. By connecting data interrogation to text analysis, Hallie
found a metaphor that could connect students to the process of data interrogation through a skill they have potentially already developed in their history or English classes.

5.3.5 Scaffolding the Tools Through Modeling

The final strategy that teachers identified was specifically related to their belief that the tools of data literacy, and especially Google Sheets should be explicitly taught to students. For all of them the obvious way to do this was through modeling the tool for the students. The teachers first began to think about what this process looked like in practice when they watched and reflected on the example video of a teacher from a previous cohort during session 2. Mary noted,

I thought he was demonstrating pretty clearly how to think about raw messy data. I think that demonstration of you know you're asking questions. He was just sort of riffing on like oh I'm looking at this, oh I noticed this, oh I noticed this now. What do you think? He was, I think, trying to show that you know it's a process. You notice one thing and then you notice another thing and then you jump to another thing.

Mary recognized that the teacher in the video was not only modeling the process of analysis within the tool, he was modeling his thinking that led him to make decisions about what and how to analysis. The strategy of modeling the tool of Google Sheets for the students was one that the teachers often pointed out on the Video Reflections of each other’s videos as well. An example is Will noting on his Video Reflection Framework on one of Manisha’s videos “She showed students how to move around in the data set. Has students edit and remove the data after she demonstrated this on her display screen.” And the idea of modeling the process of data analysis surfaced as a theme in the concept map
in the last session, highlighted by Manisha (Figure 15). This section of the concept map pulled together a number of notes about showing students specific skills or examples of data analysis with Google Sheets. Two specific subthemes within modeling the tools that received a lot of attention from the participants were making mistakes visible (highlighted in Figure 15 by yellow stars), and repetition (highlighted in Figure 15 by a blue star).

**Figure 15**
*Concept Map on PCK for Data Literacy: Modeling the Process Section*

*Note:* The Green box depicts the theme. The yellow stars highlight the subtheme of making mistakes visible and the blue star highlights the subtheme of repetition.
5.3.5.1 **Demonstrating Mistakes.** The strategy of demonstrating mistakes was mentioned in ideas in the concept map by both Mary and Manisha (Figure 15, yellow stars) as a way for students to learn through errors by seeing the mistakes that their teachers made either purposefully or not and then seeing how to fix those mistakes. Manisha had pointed out this tactic first when watching the example video in session 2, observing,

He was sorting data, so he had the data and now he was going through it. The strategies used, he used the technique of troubleshooting and simplifying the data for usable format as an approach to analyze or initiate the sorting. So, he was looking at what was missing what went a little off and then he used that to figure it out and he let the student see that.

Here Manisha was referring to a moment in the video when the teacher deleted some data that he actually needed and then figured it out because he noticed the hole later and demonstrated to the students how to go back and fix his mistake. Will referenced the same strategy during the discussion reflection on his video during session 4. He shared with the other teachers,

You know, you can always go back if you make a mistake and I made several mistakes in it, especially the first couple days, because I was you know, having brain farts the entire class. And we talked about it, and it was great because I'm showing the kids that you can make mistakes. I even make mistakes still using it and I've been using it for years and it's okay to make mistakes.

The “it” Will was referencing is the tool Google Sheets. He discussed how not only was the process of making mistakes during modeling the process helpful because it demonstrated for students what to do when a mistake is made, but it also modeled that
making mistakes isn’t a big deal, that everyone does it. Mary expressed a similar sentiment during her post implementation interview reflecting,

I do think it was kind of fun for them to sort of see that I had no idea how to do a lot of these things and I kept getting confused. And my getting confused led me to figure out other ways of doing a thing which they were then able to see.

Here Mary shared that watching her make mistakes and muddle through using the Google Sheets tool was actually fun for her students and that they learned new ways of doing things with the tool as she was figuring them out.

5.3.5.2 Repetition. Another strategy for modeling the tools was the strategy of repetition. This is a strategy that is universal pedagogy used across almost all teaching subjects and environments, but the teachers found it particularly important for scaffolding the students in working with Google Sheets. Will observed,

I actually had to do this three times that day, three separate sheets walking them through that; and then the next day again two more sheets for those students that were absent. If I went to one student and I helped them and then another student had a similar issue, and then by the time I hit number three with the same issue, alright we’re going to do it again.

Here, Will described using his enacted knowledge of student understanding that he was developing in real-time while teaching to help him determine when and what needed to be repeated and demonstrated for a second or third time. Manisha also employed the strategy of repetition to an even larger scale, though we never had a chance to discuss it as a whole group. After winter break she found that a lot of learning loss had occurred with her students, so she went back and retaught the entirety of lessons 8, 9, and 10 which
focused on developing Google Sheets skills. She had the students step through all of the exact same worksheets and activities they had done in December for a second time and she found that even though the students professed to not know how to use Google Sheets, most of them picked it up faster the second time through.

Finally, Mary spoke about repetition as well in session 4 when she was talking about teaching the students Google Sheets. She said,

It does seem like a few students are paying attention at any given time and so repeat again. This is something that I have to practice myself. I think I've said something five times and maybe it's been heard by two students. Maybe I have to say it 15 times for it to be heard by 10 students and 20 times for all my students. I feel like I don't want to bore my students by repeating myself, but I think it's the opposite, it's not boring them, it's helping them. That repetition of you know that back arrow is really your friend like you can't mess up, you can go back multiple times, if you need to, to where you started. You can start again. That's really powerful. It's okay to make a lot, a lot of mistakes.

She acknowledged that sometimes it is hard for teachers, herself included, to repeat things as many times as they often need to be repeated for students to absorb them. She also connected repetition to demonstrating mistakes because the thing she was talking about needing to repeat 5, 10, 20 times was how to use the back arrow on Google Sheets to fix mistakes.

Through all of the beliefs about data literacy, goals for teaching it, preconceptions that students might have about data, and strategies for addressing those goals and preconceptions, teachers in this study surfaced a number of components of PCK for data literacy. In the next section a summary of all the components discussed and the
connection between them will be presented as a more wholistic idea about what PCK for data literacy might look like.

5.4 Summarizing the Components of PCK for Data Literacy

The components of PCK for data literacy surfaced by the teachers combine much of their existing PCK for science with some components of knowledge about math and technology as well as even a pedagogical tool from humanities classes. While the strategies for teaching focused on are not entirely unique to teaching with and about data, the ways that they are applied in relationship to the other components of PCK create a foundation on which teachers can develop knowledge for teaching data literacy. Within the 17 components surfaced there are a number of important connections and through lines that help define PCK for data literacy more holistically.

The first is the thread of data interrogation which builds from teachers’ belief about the nature of data that data has the power to manipulate, combines with the knowledge of student understanding that students don’t have experience interrogating data and that they tend to rely on assumptions rather than actual data, to solidify the goal of teaching students to interrogate and question data and then lands on the strategies of explicitly teaching data interrogation through questioning and using storytelling as a way to frame inferences from data. All of these different components of knowledge grow together and affect each other. As teachers reflect on what they know about students’ understanding of data, that can shift their goals for teaching data literacy which can in turn affect which strategies they choose to use and using the strategy of interrogating data
in the classroom can build their knowledge of students’ understanding and potentially shift their own orientations to the nature of data and data literacy.

Another thread is the relevance of data literacy to students’ lives which again connects all three major components of data literacy with teachers’ knowledge that data literacy is not inherently motivating for students feeding into their own belief that data literacy must be grounded in students experience and therefore must be taught by not only making connections to students’ lives but also by making data tangible, sensory and kinesthetic for students. However, these teaching strategies are also built on the knowledge that students have a fear of numbers and that their understanding about what data even is is constrained in a number of different ways. Making connections to data in their every-day lives and conducting activities that interact with data tangibly are also addressing these other preconceptions that students have. Additionally, the belief that the tools of data are powerful and should be explicitly taught serves to guide teachers in thinking about ways to implement that explicit teaching through modeling the tools in a repetitive way that highlights and normalizes potential mistakes. Also, the knowledge that students are often seeking a “correct answer” out of data is the driving factor for the theme of storytelling in both the goal of data literacy and the strategies.

While most of the findings in this chapter arise from knowledge that was already held by the teachers, it was the structure of the PD that allowed for that knowledge to be surfaced by the teachers and then identified into the components of PCK presented here. The next chapter presents the components of the PD and which of those components were most successful at supporting teachers’ in exploring PCK for data literacy.
CHAPTER 6. FINDINGS ON RQ2: HOW THE PD COMPONENTS SUPPORTED CHANGES IN TEACHERS’ KNOWLEDGE

The second research question addressed in this study was, What PD components appear to support teachers’ development of subject matter knowledge and PCK for data literacy? As will be discussed in more detail in the next chapter, the most valuable and tangible shift in teachers’ knowledge was simply becoming more aware of the knowledge they already held and being able to articulate it in a way that made it easier to apply to their implementation. As Mary stated during the end reflection of session 9, “I think that what teachers need is these, is good ideas about how to connect with students around complicated concepts and having these workshops I think have been really powerful.” To that end, the two primary design components of the Extension Workshop Series, video reflections and CoRes, are presented in some detail below with teachers’ thoughts on how they each contributed to their learning. The teacher post-workshop interviews served as a primary resource for presenting these findings from the teachers’ perspective and the artifacts produced, the written video reflection frameworks and the written CoRes served as secondary sources. Additionally, the facilitator memos written after each workshop session served as a framework for the findings presented in this chapter.

6.1 Video Reflections as Affirming Anchors

All four of the participating teachers spoke in their interviews about the value of doing the video reflections and getting the chance to see each other teach as well as receive feedback on their own teaching. One of the biggest takeaways that the teachers
had about the video reflections was the simple affirming nature of doing them. Both Manisha and Will spoke of this succinctly in their interviews. Will said,

The meetings with the cohort was great, because every day I would go, and I didn't have the best support at times in the class, so it would always be oh, that was the worst, that was a horrible lesson. And then to see my videos, and other teachers are like, "Oh my God, that was amazing," I was just... It didn't feel like that!

In this quote Will was talking about how while teaching, and often while reflecting after with his student teacher, they would be incredibly critical of the lesson, but then when he shared his lessons with the Workshop group, the team was effusive about how great the lessons were. Manisha had a similar reflection during her interview saying,

Yeah, I mean, it was very confidence-boosting. Because, like I said, I felt like I was failing every day I was doing the lesson. I'd say, "I'm really struggling, the students I don't think are getting it," and then... Because when you're yourself, you're very critical. And then hearing that other teachers are saying, "No, you really did this great," you don't have the ability to see that. And: this was done very well, hey, I never thought about doing it that way. That was really helpful.

Here Manisha was not only talking about how the video reflection process served as a confidence boost for her, but it also sparked conversations with her peers about teaching strategies because her peers noticed strategies she had used while implementing and discussed how that was a new idea for them. This confidence boost and recognition of the strategies she used was helpful in building Manisha’s confidence for teaching, but also in helping her recognize the strategies she was using and be able to identify them as discrete tools. Hallie also had almost the exact same sentiment as Manisha during her interview about how doing the video reflections increased her confidence for teaching the content
but also helped her parse some of the strategies she was using in the classroom. Hallie reflected,

So, one of my biggest takeaways really is that I'm not getting the whole picture when I'm in the middle of the thing, doing the thing. And theoretically, I think I knew that, but to just listen to other people, analyzing what I'm doing and to analyze what someone else is doing and then hearing their reflections like, Wow, I felt like a miserable failure at that moment. So, the part where you're seeing something insightful happening here is encouraging, right? So, just having the encouragement of your peers is something that we can't get enough of in education because we're so isolated.

The encouragement that the participants received from each other while discussing the video reflections helped to boost confidence in teaching the data literacy components across the board.

Another aspect was specifically being able to compare their classroom context and their students to other classrooms and students with similar contexts and demographics. Mary spoke about how this was confidence boosting for her when she said during session 8, “I mean like watching the videos is really helpful because in some ways it's like oh, you know his students aren't asking any more questions than my students per se makes me feel better.” Seeing that other teachers were having the same challenges with their students and encouraging engagement helped Mary feel more confident in her own abilities to implement. Manisha spoke about this feeling in more detail during her interview, reflecting about the value of being able to see into other classrooms like hers with similar populations of students.

Even seeing other teacher’s classrooms, then you could see yourself in it. The questions that were asked, the kind of responses I would get out of my
kids, would they think in that way? That’s my pool of kids and could I stretch them to think that way, because that’s their peer group. It allowed me to see where my kids are, and their peers are in other population groups in the same district and how they were thinking. What it gave me access to is their peers that was beyond my classroom. And the teachers and their interactions that are my peers in similar settings. Unless you’re taking loads of time away, will you ever get access to that?

Manisha was describing the value to her of being able to see students like her students in other classroom settings achieving success while engaging in data literacy learning and how it encouraged her to have higher expectations for her own students because she saw that other students in similar contexts with similar barriers were able to reach those levels.

Will, due to the timing and alignment of his implementation was able to share videos for group reflection twice during the Workshop Series and after the second round of video reflections on his classroom he commented on how the process helped him think about his teaching differently and make changes to his implementation saying,

Being able to see each other teach their videos and then talk about how to make it better for next time. How to take what you're doing and how to make it even better. Take ideas that somebody else has done and improved. That's what teaching supposed to be I think. That's why I like this. This is how it's supposed to be done. You're supposed to be able to take a period off and go observe another teacher in your school.

As discussed in the previous findings chapter, Will had the most opportunity to use the knowledge surfaced during the Workshop sessions to change and improve his implementation, and as such was the most successful at doing so. This comment supported the role that the video reflections played in not only increasing his confidence for implementing data literacy lessons but also shifting his practice.
The finding from the teachers’ reflections that not only did the video reflection process increase their confidence, but it also helped them articulate teaching strategies was also supported by the written video reflections and the shift in the language teachers were using while filling in the framework over the three rounds of individual write-ups they completed. One example of this came from Will, who identified two instructional strategies in his initial individual video reflection on Hallie’s video “Share/reflection: Ask students what their thoughts are.” and “Engaging students to think about the data.” A month later in his video reflection write up for Manisha’s videos, he identified similar strategies but with more specific and robust language. He wrote,

Probing questions: Teacher knows the answer but does not provide the students the answers because she wants them to answer. She keeps probing and asking in different ways to get them to the answer she is looking for. Asks students what data is important. Students are drawing inferences from data set.

In between session 4 and session 7 the teachers had engaged in not only a discussion about Hallie’s video but also in-depth conversations about Data in Context and Complexity in Data and those conversations led to Will more clearly identifying a specific strategy being used by Manisha to “engage students to think about data.” Another example of the video reflections serving as a tool to solidify ideas developed during workshop discussions was in Mary’s write up on Manisha’s video she wrote as a strategy, “Exploring what is missing: questions and follow-up to the students to elicit specific observations about the story being told.” This is a direct reference to a conversation had during session 6, the session immediately before this round of video reflections was done. While there isn’t evidence that writing the video reflections directly
shifted teachers’ knowledge independent of the Workshop conversations, they certainly served as a way for teachers to solidify the knowledge that was being surfaced and developed during the Workshop discussions.

Overall, the video reflections served primarily as an affirming anchor that boosted teachers’ confidence and allowed them the space and structure to identify strategies they were using already in the classroom and make those strategies explicit. Manisha summarized it well during her interview when speaking about the video reflections saying,

It starts bringing up other questions on what you're doing. And I was like, oh, that's what I was doing. Oh, that, oh, gee, I missed that. And along with all the other noise that's going on in our lives, because there's so much going on, it's been a crazy time and it's been noisy, but that anchored us into that moment and thinking about that, because if it wasn't for that, we would not have that moment of time to just coagulate some of the happenings.

The video reflections allowed teachers space to pause and reflect on what they were doing in their classrooms and receive feedback from peers that highlighted certain moments in the classroom that might otherwise have been lost in the noise. The CoRes also served as a framework that allowed space for reflection, though less connected to actual implementation and more around potential for future implementation.

6.2 Content Representations as Guidelines for Discussion and Reflection

Again, all four participants thought the CoRes were a useful component of the Workshop Series. They primarily reflected in their interviews about how the CoRe Template provided them a specific framework that allowed for more nuanced and
focused reflection. When asked about the role the CoRes played in her learning during the interview, Manisha said, “They were helpful to think about it. It was more broken down thinking about and more focused reflection. So, that was good about it.” Mary had a similar response to the usefulness of the CoRes saying, “Some of the strategies and the idea of you know PCK. These ideas, the big ideas, but also the really nitty gritty practical stuff which like PDs don't tend to be very practical they’re more conceptual they’re intellectual. The more practical, the better.” Both Manisha and Mary were talking about how the CoRes allowed them to have more focused reflection on practical components of teaching.

Additionally, Hallie and Will both reflected during their interviews about the usefulness of the conversations engaged around the CoRes. Hallie spoke to how being forced to articulate her own thinking to fit into the framework of the CoRe led to deeper thinking for her. She said in her interview, “So, I'll say just the brainstorming around what to do and how to do a thing because it's a lot easier for me to think deeply about a thing when I'm trying to express it to someone else than it is when I'm just trying to turn it around in my own head.” Similarly, Will focused in his interview on how listening to other people articulate their thoughts in response to the CoRe framework was useful for him, reflecting during his interview,

So that helped me focus on, I was more looking at what the other teachers were talking about, so that if there's things... Obviously, I know what I already know and I know what I don't know but seeing what other teachers were saying to... what they were focusing on and what they knew opened my eyes: okay, oh, I hadn't thought of it that way. Oh, I didn't even think of that concept. So that was very beneficial to change my thinking in how I was going to present this or how, if I were to do it again, I would change the focus and make sure the students got this understanding of the content,
a whole other perspective changes the way you view the classroom, so doing that definitely helped.

In this reflection, Will was talking about how the discussions around the CoRes and the ideas that the other teachers came up with and the knowledge they surfaced helped him to grow his own knowledge. They opened his eyes to new concepts and new strategies that he hoped to employ in his own classroom going forward.

During session 1 after working to complete the first CoRe of the Workshop Series, when reflecting on the process, Manisha commented,

It's like emptying your brain onto paper. Like, you made it visible and it's like okay, this is what I'm thinking, and this is why what you're doing intuitively but it's like okay, you know that stuff we keep on doing with my kids which I think I feel sometimes they interpreted as harassment and I'm asking 1000 questions, just to squeeze them to get it out of their brain. I guess now it's my turn, yeah.

This supported the goal of doing the CoRes during the session which was to allow teachers to make visible their knowledge of teaching in a way that would allow them to discuss it and refine it. Manisha was able to articulate this well after the first collective CoRe process, and that sentiment continued through the remaining sessions as the teachers continued to use the CoRe framework to make their thinking visible and more open to discussion.

As discussed in Chapter 5, there were a number of moments where something that someone had written on the CoRe collaborative Google Sheet prompted a conversation that built on that idea and often extended or refined it. One example of this was during session 1 when Manisha added “Comfort with numeracy” to the section of the CoRe
labeled “Difficulties/ limitations connected with teaching this idea.” Which prompted a number of follow up thoughts from both Hallie and Mary building on and expanding this idea. During session 3 while building the CoRe on Data Visualization, Hallie and Mary added similar but slightly differently scaffolded activities to the CoRe and then had a fruitful debate about the merits of each approach. A final example is from session 7 when participants were working on the CoRe for complexity in data and “CER” was added to the CoRe which prompted a discussion about the value of CER and lead to both Hallie and Will adding additional thoughts to the CoRe about strategies for using CER and how it can benefit students.

6.2.1 Developing Big Ideas

While the CoRe template was useful for guiding discussions and reflections on teaching with and about data to be more focused and nuanced, perhaps the most useful part of completing the CoRes was identifying the big ideas for data literacy. As was evidenced in their pre-surveys and their discussion forum posts during the summer PD it was clear that teachers were not used to thinking about data literacy as subject matter that needed to be explicitly taught rather than as a tool to be used to teach other concepts in science. Having sensed this predisposition among the teachers from reading their discussion forum posts and pre-surveys, I, as the facilitator built extended time into the Workshop sessions for discussing and identifying big ideas before building each CoRe, yet it was often still not enough, as evidenced by the fact that during session 5 participants became so immersed in the discussion of big ideas for data in context that we ran out of time to actually create the CoRe which led to the CoRe for Complexity in Data
deliberately being designed to take two sessions so that there was more time for
discussion of big ideas.

Simply coming up with big ideas that were appropriately specific but also
appropriately generalized was a challenge for the teachers but a challenge that led to a lot
of growth and recognition and articulation and refinement of knowledge that the teachers
already held. However, they often needed support and guidance to make the big ideas
concrete. As a facilitator there were many times when I had to encourage or remind
teachers to take a long winded idea they had shared with the group and try to simplify it
into a concrete big idea. An example of this process was shared in 5.1.2.2 when the
teachers built on each other’s thought about the role that data can play in manipulation to
land on the big idea of “data is biased.” Will spoke to this process in his interview,
reflecting on how the teachers built off each other ideas to build bigger and more robust
ideas and how he appreciated the open-endedness of the CoRe prompts because it
allowed for that discussion and meaning making to happen.

Not being tasked with: answer this question, but: here's a prompt, let's
discuss this out loud. And it felt like all of us took something from another
previous speaker and added to it, and then it just kind of became a bigger
end goal or end thing than if we worked individually on it.

Another example is from session 6 when teachers were brainstorming big ideas
for Complexity in Data and had been sharing a number of ideas for specific activities
such as Hallie’s penny activity, I had to step in as the facilitator to try and guide them
towards some actual big ideas, saying, “I want to bring us back to big ideas. What are sort
of the big ideas that we want students to take away in terms of when we're talking about
these activities for data complexity?” This prompt by the facilitator led to another lengthy conversation that was robust but not focused, so as the facilitator, I jumped in once again to attempt to synthesize what they had been discussing. I said, “So, I think what I’m hearing is that many different types of information can be thought of as data and that that is sort of one of these big ideas that will help students see how they are more experts than they think they are at data.” This synthesis of the discussion they had been having finally gave them something more concrete to work with and allowed them to focus on a single big idea, which after some wordsmithing turned into the big idea “Data can take many forms.”

The teachers came up with many more big ideas than they actually ended up focusing on in the CoRes they built, however having the CoRe as a framework allowed them to decide which of the big ideas were the most important or salient, sometimes to them as a whole group or sometime to them individually depending on the structure of the CoRe building. So, while the brainstorming process for the big ideas themselves was a crucial piece of surfacing knowledge about teaching data literacy, the CoRes allowed the participants to focus in on specific big ideas and connect those to knowledge about students’ understanding and specific teaching strategies.

6.3 Building a Community of Support

Outside of the structured components of the video reflections and the CoRes, the participants also commented on the value of the learning community that was built during the Extension Workshop Series. Though developing a learning community was not an explicit goal of this PD intervention, one was developed, and the teachers were grateful to
have it. In her post-workshop interview Mary reflected, “Well, one takeaway is that we all do things differently. We all have different kinds of actual classroom environments, but there is a real community. I mean, there's a real sense of mutual support, mutual aid.” In this comment, Mary explicitly referred to the community she felt was created among the four teachers in the study. She continued to explain that having that community made the ideas and feedback shared during the sessions more valuable to her, “And so that's the reason why these workshops are cool. It's just the ideas that come, it's like, "Oh, oh, oh, oh." And it's like the reality is when it's like [Manisha]'s idea as opposed to some random teacher online that I'm Googling then it feels like the collegiality of having a connection.”

In her post-workshop interview, Manisha similarly reflected on the collaboration and support and the power of feeling that support saying,

It's hard to say most supportive, it's really hard to say that because there was so much support and there is so much support. You can't underestimate the power of that, going through those, the anchoring of the sessions is helpful. The collaboration is helpful. My greatest takeaways, I mean, I love the collaboration.

Manisha was explaining that the support she felt through the collaborative efforts of the group was the most powerful component of the workshop for her, and that she continued to feel that support even after the Workshop Series had ended. In my continued conversations with teachers passed the end of the Workshop Series as some of them continued to implement and as I continued to check in with them, I was amazed that they were still exchanging occasional emails with each other and sharing resources. As evidence that the community that the teachers built continued, over a month after the final workshop session, I received a text message from one of the teachers containing a
photo of all four of them hanging out together in person at a district sponsored weekend PD workshop that they all ended up attending because they communicated with each other about whether or not to go.

Will spoke to the sense of community and encouragement after the discussion of his video during session 8 saying,

Like [Mary] said, this is perfect, because we get to see what other people are doing, we get to improve our craft and hopefully the enthusiasm we gain from each other translates so if we're feeling like a 50% like [Hallie] said, it brings us to that hundred. It's like watching almost like a motivational speech. I can do this, I know I have Kate and [Mary] and [Manisha] and [Hallie] watching me, so I gotta do a good job.

Will expressed the support he felt from the other teachers in the group, but also the way the community made him feel inspired to “do a good job” and that sessions brought enthusiasm that gave him energy to keep going and trying and stay excited about the project. As Mary concisely said in her interview, “People and time is so valuable. Having that time to connect with other teachers, really making the time and the space where we are talking to each other, that was the gift that you gave us.” The teachers viewed the Workshop Series sessions as a gift of time and space to support each other and learn from each other. As a researcher, a teacher educator, and a former classroom teacher, I am so glad that this PD intervention was able to provide them with that time and help them grow a supportive community that they can lean on going into the future. Whatever other findings this study produced, putting my practitioner hat on, that outcome makes this study a success.
The primary goal of the Workshop Series for the participants was to support them in surfacing different components of their own personal PCK for data literacy and by sharing them and reflecting on them collaboratively through the structures built into the PD, build a beginning framework for collective PCK for data literacy. While the participating teachers were focused on digging in to what PCK for data literacy is and how it can help them implement in their classrooms, the facilitator of the Workshop Series had additional goals, one of which was to track and try to understand how the Workshop Series combined with implementation of the Bioinformatics PBL Unit served to change teachers PCK. The next chapter presents case studies on three of the four teachers to explore change in each of their personal PCK throughout the entire PD and implementation process.
CHAPTER 7. FINDINGS ON RQ3: THE SUBJECT MATTER KNOWLEDGE AND PCK FOR DATA LITERACY DEVELOPED BY TEACHERS

The third research question for this study asked *How do teachers increase their PCK for data literacy during PD experiences that have been designed to explicitly focus on data?* While this chapter will focus primarily on teachers’ personal PCK development, first I will present findings on teachers’ collective subject matter knowledge development, as subject matter knowledge is inter-related with PCK and the process of exploring PCK led to shifts in perceptions about the goals of data literacy.

7.1 Growth in Teacher Collective Subject Matter Knowledge of Data Literacy

While teachers’ individual growth in subject matter knowledge and PCK could be determined to some extent by analyzing their individual written responses and interviews, and will be presented in the next sections, a lot of the learning and growth that may have occurred happened during the Workshop discussions which were a collaborative process. To understand better what learning might have occurred during those sessions, I again examined the cohort as a single case learning community to explore the way the knowledge shifted over time.

As was discussed in Chapter 6, the development of Big Ideas for data literacy was a process that supported the teachers shift in awareness of data literacy and its goals and purpose in the classroom. Table 7 displays the big ideas developed in relation to each of the components of data literacy. Only a few of the big ideas were then used to develop
CoRes but examining the entire set of big ideas allows for a wider lens on how teachers were thinking about the subject matter knowledge of data literacy.

One big pattern that arises from examining Table 7 is that the majority of the big ideas surfaced by the teachers were not procedural. While a few of the big ideas referred to actions (e.g., Finding patterns and trends can allow for connections between variables, and messy data needs to be cleaned up and summarized in order to be analyzed) the majority of them were conceptual and focused on the nature of data literacy. As presented earlier in this chapter, both Will and Manisha showed signs of moving from more procedural to more robust and conceptual definitions of data literacy in their own knowledge development.

The focus on the nature of data and data literacy more than the skills of data literacy also suggests that there may be a disconnect between what the field is focused on as the big ideas of data literacy and what teachers believe to be important. To borrow a phrase that Hallie was fond of using, “the what, versus the so what.” While researchers, curriculum writers, and college professors have been more focused on “the what” of data literacy teachers seem to be much more interested in engaging their students with “the so what.” Rather than teaching students the skills of how to analyze data, teachers are thinking more broadly about data literacy as a mindset and way of interacting with the world. This potential disconnect will be probed further in Chapter 8.
Table 7
The Big Ideas Identified During the Workshop Sessions

<table>
<thead>
<tr>
<th><strong>Inference with Data</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• There are many tools available to help make meaning of data</td>
</tr>
<tr>
<td>• The tools (and their capabilities and biases) that are used can impact our inferences</td>
</tr>
<tr>
<td>• Big data can be and is used to influence our lives</td>
</tr>
<tr>
<td>• Data is biased, there is no perfect data</td>
</tr>
<tr>
<td>• Anyone can make meaning of data</td>
</tr>
<tr>
<td>• Not limited to one conclusion, there could be numerous interpretations</td>
</tr>
<tr>
<td>• Not all data is good data, some data is “trash”</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Data Visualization</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• An effective visualization is something that makes the information and intention clear (the what AND the so what)</td>
</tr>
<tr>
<td>• Numbers can be visualized in many ways (different types of visualizations)</td>
</tr>
<tr>
<td>• It’s a snapshot of information: a data visualization is bound by time</td>
</tr>
<tr>
<td>• Decisions about the visualization affect the story that is being told</td>
</tr>
<tr>
<td>• The categories/axes/titles with which you design the visualization matter for how it’s interpreted</td>
</tr>
<tr>
<td>• What you exclude in a data visualization matters as much as what you include</td>
</tr>
<tr>
<td>• A visualization is emphasizing and editing a set of data.</td>
</tr>
<tr>
<td>• Data visualizations can be misleading and can/should be examined for integrity and purpose</td>
</tr>
<tr>
<td>• Data visualizations tell a particular story about a set of data at a particular point in time.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Data in Context</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Interrogating data is useful and necessary</td>
</tr>
<tr>
<td>• Data can be used to tell different stories with different intent</td>
</tr>
<tr>
<td>• Data is often manipulated for non-scientific reasons to sell you something or lead you to a specific conclusion.</td>
</tr>
<tr>
<td>• People make decisions about what and where and when to collect data</td>
</tr>
<tr>
<td>• A lot of things get measured because they are easy to measure not because they are important to measure</td>
</tr>
<tr>
<td>• The context helps to understand where the differences in arguments and stories are</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Complexity of Data</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>• Actual randomness often has high variability</td>
</tr>
<tr>
<td>• Types of information is data - Data can take many different forms</td>
</tr>
<tr>
<td>• Data is not “other” or outside their experience - it’s not totally abstract</td>
</tr>
<tr>
<td>• Complex data includes a human factor as well as other pieces of the context</td>
</tr>
<tr>
<td>• The questions we really want answers to are complex and complicated</td>
</tr>
<tr>
<td>• Messy data needs to be cleaned up and summarized in order to be analyzed</td>
</tr>
</tbody>
</table>
### Data in Aggregate

- Sample size matters when looking for patterns across a data set, because your sample needs to be representative of the population.
- Multiple trials “smooth” out variability from human error and other variables (e.g., time) and uncertainty.
- Finding patterns and trends can allow for connections between different variables.
- There can be interesting data in the outliers and finding the aggregate measures help us identify outliers.
- Averages can be useful for comparing between data sets in a way individual data points are not.
- Aggregate measures are not always useful or appropriate and may take up more space than they have earned in the media/culture.
- Different aggregate measures support different data visualizations.

---

### 7.2 Growth in Teachers’ Individual and Personal PCK and Subject Matter Knowledge

The four participants, though teaching in the same school district and having similar demographics of students, had vastly different experiences implementing the Bioinformatics PBL unit. They also all entered the program with different orientations to data literacy and different goals for what they wanted to get out of the program. Despite spending over 15 hours working collaboratively to build meaning around PCK for data literacy, these different experiences led them to transfer the collective PCK into their own personal PCK in different ways. This chapter explores three of the participants’ learning around data literacy separately and then pulls the cases back together to highlight themes across their respective personal PCK shifts. The three teachers who are presented as cases here are Will, Hallie, and Manisha. Will had the most successful implementation of the four teachers and was able to use that success to build his confidence in teaching data literacy and shift his teaching strategies and goals for future units and classes to align...
with this increased confidence. Hallie came into the program with the strongest subject matter knowledge for data literacy of the cohort as well as being the only one of the four who already had a sense of the definition and purpose of PCK, however the external stressed of the school year caused her to shift her orientation to teaching in general. Finally, Manisha came into the program with the lowest subject matter knowledge and confidence for data literacy and grew in both knowledge and confidence over the course of the Workshop series. Mary is not presented as a case here because her implementation ended up being significantly different than the other three teachers due to choices and pivots she made mid-year. As a result of these choices, her implementation was misaligned from the other three teachers and she was not able to share a video of her teaching during the Extension Workshop Series.

One measure for how the teachers’ orientation to teaching data literacy shifted over the course of the full intervention is their implementation survey scores from pre-summer PD to post-summer PD to post-workshop. The implementation survey measured teachers’ disposition to teaching with and about different components of the PBL including use of mobile technologies, embedded socio-scientific issues, data literacy, STEM integration, and bioinformatics using a Likert-scale (1 = strongly disagree to 5 = strongly agree). The data literacy component asked participants to score their agreement on statements specific to their use of data in their classroom (e.g., I teach my students to recognize and understand patterns in data, and I teach my students to understand that all data comes with a level of uncertainty in how it represents the phenomenon). The STEM integration component probed for agreement with statements about teaching for STEM integration more generally (e.g., I consider connections to other disciplines when
designing and teaching lessons, and I feel comfortable teaching components of math and technology in my science classroom). Table 8 shows teachers scores for the data literacy component of the survey and Table 9 shows scores for the STEM integration component across the three measurements. As teaching data literacy explicitly is a form of STEM integration, knowledge of and disposition toward STEM integration generally can affect knowledge for data literacy. As such, both sets of scores are presented here.

**Table 8**

*Results of Likert-Scale Survey Questions on Data Literacy*

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Pre Summer PD</th>
<th>Post Summer PD</th>
<th>Post Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hallie</td>
<td>3.83</td>
<td>3.58</td>
<td>NA</td>
</tr>
<tr>
<td>Mary</td>
<td>3.58</td>
<td>4.00</td>
<td>3.83</td>
</tr>
<tr>
<td>Manisha</td>
<td>3.33</td>
<td>3.92</td>
<td>5.00</td>
</tr>
<tr>
<td>Will</td>
<td>4.08</td>
<td>4.83</td>
<td>5.00</td>
</tr>
</tbody>
</table>

Avg of entire cohort 4.04 (0.59) 4.33 (0.51) NA

*Note:* Most teachers from the full summer cohort of 9, including Hallie, had not completed implementation as of this writing.

**Table 9**

*Results of Likert-Scale Survey Questions on STEM Integration*

<table>
<thead>
<tr>
<th>Teacher</th>
<th>Pre Summer PD</th>
<th>Post Summer PD</th>
<th>Post Implementation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hallie</td>
<td>3.75</td>
<td>4.00</td>
<td>NA</td>
</tr>
<tr>
<td>Mary</td>
<td>4.50</td>
<td>4.50</td>
<td>4.00</td>
</tr>
<tr>
<td>Manisha</td>
<td>3.25</td>
<td>3.50</td>
<td>4.75</td>
</tr>
<tr>
<td>Will</td>
<td>4.00</td>
<td>4.25</td>
<td>4.75</td>
</tr>
</tbody>
</table>

Avg of entire cohort 4.06 (0.58) 4.53 (0.52) NA

*Note:* Most teachers from the full summer cohort of 9, including Hallie, had not completed implementation as of this writing.

Manisha and Will both displayed steady inclines in their scores for both data literacy and STEM integration, while Mary’s and Hallie’s were more unstable. The implementation survey asked questions about how participants currently taught, or in the case of the post-summer survey, how they were planning to teach in the coming year.
Will and Manisha both showed a marked increase in confidence for teaching data literacy as well as plans to include more data literacy in their teaching which was supported by their written responses. Hallie, as will be discussed later in this chapter had a shift in her orientation to teaching overall that led the data literacy component of her implementation plans to be less salient. Details for how each teacher changed in their knowledge of teaching with and about data literacy is presented in this chapter. Their subject matter knowledge for data literacy is presented first as it is a primary knowledge base upon which PCK is built as depicted in Figure 3, followed by their PCK for data literacy.

### 7.2.1 Will

Will came into the Bioinformatics PD with an open predisposition toward growing his subject matter knowledge and skills for data literacy. He also had the most stable school schedule and attendance rates of all the teachers as well as an elective class which gave him full control and freedom over his course design and goals. Additionally, he chose to begin the Bioinformatics PBL unit in September, which gave him ample time to complete it, reflect on it, and apply what he learned to future units and lessons. Finally, there were three other science teachers at Will’s school who had been through the Bioinformatics program in previous cohorts, one of whom served as a mentor to Will as he went through implementation. Given this context, it is unsurprising that Will showed growth in his subject matter knowledge and PCK for data literacy as well as his confidence and plans for future implementation of data literacy concepts and skills.

#### 7.2.1.1 Subject Matter Knowledge of Data Literacy – Will

Despite his decades of experience teaching science, Will’s background and original training in history left
him feeling inadequate in his content knowledge for science in general and data literacy in particular. In his written reflections during the first asynchronous assignment in which the teachers were asked to read about PCK and data literacy and write a personal reflection about their own knowledge in this area, Will wrote, “After reviewing the information presented, I feel like a fraud to some extent and barely competent to another. My background, undergraduate degree, is in History... yet I teach Environmental Science.” As an observer in his classroom, I was often impressed with Will’s knowledge of science content, such as when he engaged in a conversation with students during one of the lessons about airflow and geography and the role that mountains and bodies of water can play in air quality. However, Will’s sense of his own knowledge and his feeling of being an outsider allowed him to approach all of our conversations about data literacy with an open mind.

There was a noticeable expansion in Will’s knowledge about what data literacy can be from the beginning of the PD intervention though the end of the workshop. In his pre-survey, during the summer PD discussion forums, and in his survey after the summer PD he wrote about data literacy in a very procedural way, aligned mostly with a traditional “scientific method” approach, saying in his most expansive of these responses, a discussion forum post, “Data literacy usually encompasses designing an experiment, creating a set of data (quantitative/ qualitative), collecting data, analyzing the collected data, drawing conclusions based upon the background information and data collected.” While his pre-survey did add a brief note about “communicating” data as well, Will’s definition of data literacy was not only linear but also more aligned with the ‘in services of learning a science topic’ view of data usage in a science classroom.
This was in contrast to his definition of data literacy provided on the post-workshop survey in which he wrote, “Data literacy is the ability to translate complex data sets and find the key metrics. It is also the ability to interpret graphs to understand what is being shown and what is being manipulated.” In this response, Will defined data literacy more as a way of thinking and approaching all sorts of data, including complex data. Additionally, Will expanded his definition of data literacy to a more nuanced understanding of what it means to “analyze” data. This was also evident in his interview in which he responded to a question about defining data literacy by saying,

So, data literacy is the taking of information, numeric or qualitative information, and putting it in terms that are easily read and understood and can be deciphered and compared and contrasted with each other and having the ability to understand the differences between the data and between different data sets.

Whereas in his earlier definitions, analyzing data was just one component in a list of steps, his post workshop definition of data literacy is more focused on the analysis piece of working with data and contains a more specific understanding of what it means to analyze data and even why it’s important with his mention of manipulation.

Will’s comprehension of the role data literacy as a form of STEM integration as well as the role STEM integration might have in his own classroom also shifted from the beginning to the end of the PD intervention. Will’s written definition for STEM integration did not change much from pre to post survey over the summer PD, using a very standard definition, “STEM integration is making sure to incorporate technological research, engineering application while using math in a science classroom.” However, he did add to this definition in the post survey, writing in response to a different prompt
about the importance of teaching data literacy, “Data literacy will also strengthen their math skills and prepare them for educated discussions in the future.” While these responses show that the summer PD did potentially lead Will to think more about the role of data and math in science classrooms, his response to the question about the definition of STEM continued to be passive.

In the post workshop PD, Will struck a more active tone when asked the same question “What is STEM integration as applied to a high school science classroom?” After the workshop series he wrote, “STEM integration is vital in a High School classroom. Students need to engage, hands-on, in the connected fields of science and math to fully grasp the concepts and understand their importance/practicality.” While this post-workshop definition dropped the mention of technology and engineering to focus solely on math and science, Will talked not just about the definition of STEM integration but it’s important and the need for students to engage with it. As discussed in the previous section 5.1.3 making connections to math was a topic discussed by all the teachers and became more vital as they implemented and discovered student’s understanding of math and its role in science. These discussions also led Will to think about the connection between math and science more himself. He reflected in his interview,

The cross-curricular nature of math is that it's essentially everywhere, but that was huge, [the students] not knowing that. This is an area that needs to be addressed earlier. Having the ability to make those connections for the students between the classes, trying to work on that is definitely something that I'm focusing on.
While Will still had some hesitancy and humility around his own knowledge of Math and Science connections, this statement from his interview showed that he was hoping to focus on continuing to build that knowledge for himself.

Another area in which Will’s own subject matter knowledge of data literacy changed was through his own self-professed growth in knowledge of and skill with Google Sheets. In response to a question about how his knowledge of data literacy might have grown during the workshop series and implementation process, Will immediately responded by talking about Google Sheets. He said, “I definitely got stronger in using Sheets. Like I told you before, with you there as my backup buffer, when my brain would just shut down on: I know how to do this, why isn't that... And you go, ‘Oh, it's this. Okay.’ That was huge.” In this quote Will was diagnosing growth in his own knowledge of Google Sheets and attributing it in part to my presence as an observer in his classroom, but also as a resource when he forgot how to do something in Sheets, or a student made a choice that led to an outcome he was unfamiliar with. Will acknowledged the value of his growth with Google Sheets further when he continued,

I had never used Sheets in class before, I had never done large data sets, it was always, I did all the graphing for them. They gave me the data, I did it for them. So, having them, the students, use the formulas, use the insert tools, yeah, that was huge.

By going from talking about how his knowledge of Google Sheets got stronger to explaining how that increased knowledge translated into the classroom for his students Will was also providing evidence of his increase in confidence for teaching with Google Sheets and with data and for data literacy in general. This is supported by Will’s jump in
score on the implementation survey to a straight 5 on the data literacy section (Table 8), meaning he “strongly agreed” with every statement about teaching data literacy in his classroom. He also increased from a 4.25 post-summer PD, to a 4.75 post-workshop in the STEM integration section (Table 9). He supported this increase in confidence in his own words, saying succinctly during his interview in response to a question about how the Workshop Series helped him grow as a teacher, “Yeah, I mean, it was very confidence-boosting.” and following that up with, “So it's made me more reflective.” A lot of what the workshops made Will more reflective about were the goals for teaching data literacy and how to go about reaching them with his students which is discussed in the next section on his PCK development.

7.2.1.2 PCK for Data Literacy – Will. Along with all the other participants except Hallie, Will did not have an understanding of the concept of PCK going into the Workshop Series. However, despite not knowing the term PCK or the explicit definition for it, he came into the Workshop Series with significant knowledge about teaching with and about data. In his initial written reflection on his own PCK for science in general and data literacy specifically, he identified “knowledge of student understanding” as the component of PCK in which he felt the strongest and when on to identify some of the themes that would later surface in the Workshop discussions. Will wrote,

Students generally want to conduct an experiment, receive the results and have the results match their hypothesis. If the results differ from their hypothesis, students will usually stop the process and not repeat it. The acceptance that all Data is useful and provides a story is important to understand. Connecting how Data is everywhere and useful to their lives is the area that needs the most work for students.
In this reflection, Will demonstrated that he was already thinking about the themes of “students struggling with the ambiguity of real-world authentic data” and “students have a constrained conception of data.” Additionally, he was already thinking about data in terms of storytelling. The theme of storytelling with data did not arise in the Workshop Series until session 3 and was initiated by Mary, not Will, which adds additional credence to the themes developed in Chapter 5, as they were not driven by one teacher but surfaced what multiple teachers already knew and were reflecting on. Additionally, Will wrote in his initial reflection about how data does not have inherent interest for students, but he applied it to all of science writing, “Students (typically) are not interested in learning about science in general. They do not make the connections from the coursework into real world application.” He then added that one characteristic that made him strong in his knowledge of student understanding was that he also used to think science was boring, so he drew on his own past experience to connect with students.

In his initial reflection on his own PCK, Will identified “knowledge of strategies for teaching” as the component of PCK that he most wanted to work on and grow in. He wrote, “I feel that learning new and more relevant Instructional Strategies will allow me to keep my students engaged and interacting with the material.” True to this, Will was the least vocal in suggesting new strategies during the workshop session. When he did talk about strategies, it was often to add on to something another teacher had suggested, or to explain an activity he had already successfully tried in his classroom, such as when he talked about how he had made a connection for students to air quality by talking about the Covid pandemic. However, as he was ahead of the other three teachers in implementation, there were often connections between the discussion happening during
the Workshop sessions and activities or strategies Will had already done in his classroom. Often it took the facilitator calling out these connections to prompt Will to open up and share. During a conversation in session 4 about demonstrating mistakes, I as the facilitator called Will out because I had seen him do that in a previous lesson. I said,

I remember yesterday that there was at least one spot where you were like I know that this is something that students will probably make a mistake with so I'm just going to right off the bat, I'm going to do it so that they can see it so they can see that it's not a big deal.

This prompt encouraged Will to talk in some depth about how he modeled Google Sheets for the students in the classroom using repetition and his sense of where students are confused to support them in developing their skills.

In the concept map activity during the last workshop session, Will posted a number of ideas depicting his knowledge of student understanding and of strategies for teaching data literacy. His posts that clearly depict this knowledge have been gathered in Figure 16. From these posts, it can be determined that Will had a strong sense of the need to make connections to student’s lives, as well as knowledge of students’ fear of numbers and their constrained conception of data. He also added notes about explicit objectives to focus on when teaching Google Sheets that would allow students to tell a better story with their data.

Despite demonstrating many of the strategies discussed during the workshops and in his classroom throughout his implementation, Will ended the Workshop Series still feeling that he needed to work on his “knowledge of strategies for teaching data literacy.” He wrote in his post-reflection on his own PCK,
I am most motivated to grow my Knowledge of Instructional Strategies for Teaching With and About Data. I am constantly seeking out new methods and professional development opportunities to grow my instructional strategies database. I feel that the best teachers are the ones who are always learning new ways and approaches to teaching the content they teach. Collaboration with colleagues helps to strengthen my areas of weakness and support my areas of strength.

**Figure 16**

*Concept Map on PCK for Data Literacy - Will’s Posts*

Note: This collects the majority of Will’s posts from the concept map with their connections to each other maintained, but their external connections removed.

While Will still identified the strategies component of PCK as the one where he felt he needed the most growth at the end of the workshop intervention, his response does not demonstrate that he views himself at a deficit in this area, but rather that he views it as an
area where all teachers should continue to grow in order to improve their craft. Indeed, Will had many plans for strategies and activities he could use in the future to infuse more data literacy into his classroom.

7.2.1.3 Future Plans for Teaching Data Literacy. One benefit that Will had in starting and therefore completing his implementation during the first semester is that it gave him time and space to not only think about what he might do to include more data literacy in his teaching going forward, but to actually carry out some of those plans in the current school year. One example of this, during his post interview, Will spoke about the next unit his class was engaging in which involved raising Brook Trout in the classroom. The eggs had arrived they day we conducted his interview and Will described what he planned to do with the students saying,

And as part of that, we have to constantly monitor the water readings, the pH, the microbe levels. How many eggs did we receive that were viable? How many eggs were dead? We usually just report that information back without really analyzing it. But now that they have the skills, the students are already saying, "Oh, are we going to do more charts on this?" Absolutely. We didn't just learn those skills once and then we're done. So, we're going to start looking at how... what we're doing and charting changes in our water chemistry versus the development and what they need.

This example showed not only how Will was already increasing his explicit teaching of data literacy in the classroom during a unit he had done previously without a data literacy component, but also how his students were understanding that the skill of analyzing data was going to be an integral part of their science class going forward.
During his post interview Will also spoke about a unit he had taught in previous years about energy efficiency in the context of making connection to students’ lives and how he often asked them to ask their guardian for their household energy bill. He said, “So that they can start to incorporate it in their own lives that it’s not just school, it’s: how can we make this relevant to themselves? And then, in turn, them saving energy, using the data, makes them more energy-efficient in the future, which ultimately slows down climate change.” I had the opportunity to connect with Will a month after his post-workshop interview as part of the larger Bioinformatics project and he shared that during the energy efficiency unit the previous week, some of the students had used Google Sheets to collect data and do calculations that led them to determine how much energy the school was using keeping the Smartboards on while not in use.

When we were doing energy conservation, we added using a kilowatt meter and [one student] used that on the smart board. And he determined that in a year just using it for 10 months if using the sleep cycle and leaving it on, it’s about $8,000 per smart board to power for the year. And he asked, who pays it? I said, well school districts. So, then he did a whole, we have to get everybody to turn them off on the weekends! He did that through Google Sheets and through basically, what we did in this unit translating it into the next phase. So, him and [another student] were like yes, this is what we’re going to do. So, taking the initial work from the bioinformatics and now, translating it to another area in my program they’ve made that connection.

This story supported not only how Will’s students learned from the data literacy component of the bioinformatics PBL unit, but also Will’s focus on continuing to find opportunities for his students to engage further with the data literacy skills they developed. The use of the kilowatt meter to collect data from appliances over time was a new component to the energy conservation unit that Will had not previously done. He
added it this year to build more explicit use of student-collected data and support students in continuing to grow their data literacy.

7.2.2 Hallie

Hallie came into the intervention with both the strongest content knowledge of data literacy and as the only one of the four teachers who had an understanding of the definition and purpose of PCK at the start of the Workshop Series. She was thoughtful and methodical in how she reflected on PCK, taking the concept seriously and putting effort into contemplating her own PCK throughout the entire Workshop series and not just when prompted to. However, despite, or perhaps because of, her deeply reflective nature, she went through a potentially life-changing shift in her orientation to teaching over the course of the eight months of the study.

7.2.2.1 Subject Matter Knowledge of Data Literacy – Hallie. Hallie’s content knowledge of data literacy was high entering the PD, including her familiarity with Google Sheets and her experience teaching with it. She wrote in her application to the program, “I have used Google Sheets exclusively for the last decade, or so. I have taught students to construct graphs from data, determine slope and intercept to make equations of a line, and to identify algebraic functions in their graphed data.” Additionally, Hallie self-identified strongly as someone who is knowledgeable in science and data. During session one of the Workshop Series, she remarked offhand, “I’ll sit around reading peer reviewed journals because that’s something I do for fun.” As such, there was no marked shift in her subject matter knowledge for data literacy over the course of the intervention. However, Hallie’s PCK journey during the Workshop Series was more noteworthy and
was possibly supported by her already high levels of subject matter knowledge for data literacy. Her knowledge of data literacy and her confidence in that knowledge allowed her to focus more fully on developing her PCK.

7.2.2.2 PCK for Data Literacy - Hallie. In the first session of the Workshop Series, I asked for a “thumbs up, thumbs down” on who had heard the phrase “pedagogical content knowledge or PCK for short” before. Hallie was the only teacher who gave a thumbs up. When asked what PCK meant to her, Hallie said,

There's the idea of content knowledge, which is to say those things that are specific to the particular subject in which you are teaching. And pedagogy is methodology in teaching that specific stuff. So, we're just talking about what is the methodology for teaching particularly biological science.

This was a good understanding of PCK and put Hallie at an advantage for the remainder of the Workshop Series because the concept was already familiar to her and so she had an easier time than the other teachers in thinking and reflecting through the lens of PCK and picking out specific ideas that would fall under knowledge of student understanding or knowledge of teaching strategies. In her initial PCK written reflection, which was completed asynchronously after the first workshop session, Hallie wrote, “From where I sit, personal PCK is something I began to develop long before I imagined I would be a teacher.” In this quote, Hallie suggested that personal PCK is knowledge that she had been building for a long time and thinking about already before joining this project.

In the same initial PCK written reflection, Hallie noted that the component she felt strongest in was “Knowledge of Teaching Strategies” and the one that she felt weakest in was “Knowledge of Student Understanding.” In her post-workshop written
PCK reflection she identified the same two components as her strength and weakness. This consistency framed her interactions in the workshop sessions as well as her own reflections. Her strength in teaching strategies was evidenced throughout the workshop. Hallie was the most prolific of all the teachers at coming up with specific and well-thought out strategies and activities for teaching data literacy, many of which were described in some detail in Chapter 5. Her over twenty years of experience teaching science and her predisposition towards focusing on data literacy meant that she came into the intervention with experience using different strategies to teach data literacy. In her initial reflection Hallie wrote,

> I use analogies and storytelling to make meaning of science concepts for my students and encourage them to do the same. Metaphors and incorporating everyday experience creates a process of reflexive thinking for multiple ways of knowing. My evidence comes from feedback from my students.

It was evidenced by this reflection that not only did Hallie have experience with strategies that were successful in her classroom, but she also had experience evaluating the effectiveness of specific strategies and not just general outcomes with her students. Hallie also shared a number of different strategies in the discussion forum posts during the summer PD including, “To combat bias, I have students look at data and graphs without the axes identified.” and “One opportunity I foresee is to have students explore different choices for visualization, and to evaluate how effective they are in meeting the objective.” The fact that Hallie was sharing these ideas during the summer PD meant that not only did she already have well developed knowledge of teaching strategies for data literacy, but that she was confident in that knowledge and willing to
speak out with her ideas. Despite coming in with strong knowledge of strategies for teaching with and about data, Hallie still found value in the Workshop series in pushing her to think more explicitly about the difference between pedagogical strategies and pedagogical content strategies to dig deeper into reflection around which strategies from her toolbox actually worked best for specific components of data literacy. During her interview, Hallie reflected on this saying,

So, it was trying to figure out for myself, like what was a pedagogical practice versus pedagogical content practice. That part, I think, was valuable, and that I had to keep thinking about what it meant to look for it. And even when I wasn't finding what I was looking for, the fact that I was going back and considering what I was looking for. And then, okay, then I noticed these other things that I wasn't looking for, but I recognized them on the way.

In this reflection, Hallie pointed out that thinking about the difference between a general strategy and one that was specific to data literacy was valuable for her. She also noted next that she thought it would have been helpful to “try and identify what some of those things were before going in and doing it.” By “things” she was referring to strategies for teaching with and about data and this comment led to her and I having a conversation about the goals of the research and the value of trying to define those strategies for teachers.

While Hallie identified “Knowledge of Student Understanding of Data” as her weakest component on both her initial PCK and her post-workshop written reflections, her identification of it as such led her to focus on it throughout the Workshop Series and her post-workshop reflections supported that. In her post-workshop written reflection she wrote,
My knowledge of student understanding of data is a bit weak. I find it very difficult to get an initial formative assessment, because my students don't stake any claim in understanding data, so they have a hard time articulating what they do and don't get. By the time I am able to assess understanding, we are deep into making meaning, and I don't know where we started.

In this passage Hallie described how she felt about her knowledge of student understanding and the fact that students’ lack of intrinsic interest in data literacy served as a barrier to assessing students’ preconceptions about data. When asked to expand on this during the interview, Hallie spoke about the need for a greater focus on science and data literacy at the middle school level to better prepare students for entering high school and lay a stronger and more consistent groundwork on which high school teachers could build.

Despite feeling weakest in knowledge of student understanding, Hallie was confident in her PCK for teaching data literacy overall. When asked to describe her current understanding of pedagogical content knowledge during the interview, Hallie said, “I would say intermediate. I'm certainly not new at this, but I am still occasionally stumped by how to guide students through a struggle around looking at data. It's like, what is stopping you from seeing the data that I see? I'm still unsure about that one.” This shows that Hallie is confident about PCK but yet continuing to push herself to identify student preconceptions and determine strategies for addressing them. Hallie’s predisposition toward deep reflection was also evident when asked during the interview whether the Workshop Series changed the way she was thinking about her own knowledge and learning of data literacy. She immediately responded by saying, “Well I think I ended with more questions than I had answers.” This supports the fact that Hallie
was constantly asking the next question and striving for deeper understanding and that participating in the project support her in doing so. She continued on to say,

There are things like, so what are the basic misunderstandings that students will have about this? And my answer is, I don't know. I haven't really seen big patterns, but I'm looking for them, which I don't know that I was before. So, I'm looking to find those patterns to figure out how to address them. I've been looking for a long time for meaningful way to do this and I'm getting closer. We're not there yet.

In this quote, Hallie identified that the Workshop Series opened her eyes to specifically looking for patterns in students’ preconceptions and misunderstandings about data literacy, which was not something she was actively reflecting on before the intervention. While she acknowledged that she felt she was getting closer to a meaningful way to teach data literacy she ended by saying that she didn’t think we were there yet. The use of “we” rather than “I” in her closing thought is notable in that it speaks to the way Hallie was orienting herself to the teaching professional as a whole throughout the project and school year. Whether as a direct result of participation in the Workshop Series, or not, there was not enough data to determine, Hallie experienced a noticeable shift in her orientation to teaching during the period of time the Workshop Series occurred.

7.2.2.3 Orientation to Teaching – Hallie. Along with her deep, measured reflection on PCK and her knowledge development and orientation for teaching data literacy, Hallie was also experiencing a disillusionment with teaching as a profession during the 2021-2022 school year, which though not explicitly stated for most of the year was evident as an outside observer and upon analyzing and coding her responses to
prompts and questions. It did finally become explicit in an email sent to me late in March which read simply, “I’m sort of having a professional existential crisis…”

The lead up to this recognition of a potentially major shift in her orientation to teaching started with her application to the program, which, as quoted in the Participants section, read in part, “I find much of what I am required to teach to be esoteric.” It is important to note here that Hallie taught Biology, which is the only high school science course tested in the state of Pennsylvania as a requirement for graduation. She followed her initial comment up in her written reflection on PCK at the beginning of the Workshop Series, reflecting,

When discussing the knowledge that teachers have about how to teach particular topics to particular groups of students for particular outcomes, I find myself in disagreement with the conventional science education community in defining desired outcomes. I see very little value in the vast majority of what I learned in school.

In this written reflection, Hallie again stated her disillusionment with the standard content and method of teaching high school science. She is connecting that disillusionment to PCK though not making any specify statements about what she views as the knowledge held by the “conventional science community” or how her goals or knowledge differ from it.

In her post-workshop reflection on her own PCK, Hallie highlighted Orientation to Teaching as the component of PCK that she most wanted to work on writing,

I am most motivated to increase my knowledge in my Orientation to Teaching Science. My answer to the question of what the purpose and goals of science education are has changed a lot over my lifetime, and right now, I have more questions than I have answers. I object to many of the
expectations of the profession, both content and process-wise. I am a little lost-and-searching.

Hallie had not mentioned Orientation to Teaching at all in her pre-workshop reflection on her own personal PCK, so this was a marked shift in the way she was thinking about her own role and goals as a teacher of science. Despite feeling lost and disillusioned with the process and context of teaching science, Hallie did find some cause for hope. Hallie was still working through the implementation of the Bioinformatics PBL unit when the Workshop ended and the post-workshop interview was conducted. She was frustrated that it was taking her so long to finish implementation but when asked how participating in the workshop had changed her perspective on teaching data literacy she said,

Well, I think it's premature because, like I haven't really had the opportunity to sit back and reflect on what I was doing and how it was and what I would do differently much at all because I'm still in the middle of doing it. My mind hasn't been able to rest enough to really say like, how has this changed my perspective on things? It's just changed my attitude toward the great experiment. And it makes me feel just more optimistic about giving it a shot.

So, despite her impending professional existential crisis, participating in the workshop series did add some weight to the positive side of the equation and shifted her attitude back towards being hopeful about “the great experiment” called public education, even if it potentially wasn’t enough to stave off the crisis. As will be discussed in more detail in the Limitations section of Chapter 8, the contextual factors of teaching in a low-resource urban public school during the Covid pandemic left little space for making an impact with a program like this one.
7.2.3 Manisha

In contrast to Hallie, Manisha found strength, confidence and enthusiasm while participating in the intervention and implementing the Bioinformatics PBL unit with her students. Manisha entered the intervention with the weakest background in and knowledge of data literacy. She had the lowest scores on the implementation pre-survey for both data literacy and STEM integration, and though her scores went up by the post-summer PD survey, they were still below average on both. However, by the time Manisha took the survey again post-implementation, her scores had grown to a straight 5 on the data literacy component (Table 8) and even 4s and 5s on the STEM integration component (Table 9). This shows a strong increase in her confidence and plans for continued teaching for data literacy. Likewise, she had no understanding of the concept of PCK coming into the intervention but developed a sense of what it was and how to use the framing of it to reflect on her teaching strategies and goals over the course of the Workshop Series. These gains in subject matter knowledge and PCK are presented in the next two sections.

7.2.3.1 Subject Matter Knowledge of Data Literacy – Manisha. Manisha had very little experience with data literacy coming into the program. She wrote in her application that she had never taught data literacy to her students before and that her only experience with Google Sheets was to use it sometimes to track students’ grades. She also had a constrained view of data literacy at the beginning of the program. When asked to define data literacy Manisha wrote on her pre-summer survey, “Data literacy serves a purpose to enable us to observe a given set of variables - numerically. It is the observation and analysis of the data, that can lead to decision making and argument.”
post-summer interview produced a similar but less prolific response. In these definitions 
Manisha identified data literacy as being related to numerical data only, which is a 
constrained view of data held by many students. It can be hypothesized that Manisha was 
referring to “engaging in argument from evidence” which is one of the science and 
engineering practices from the NGSS. Similarly to Will’s original conception of data 
literacy, Manisha was defining data literacy in a procedural way before engaging in the 
Workshop Series.

After the intervention, on her post-implementation survey, Manisha had a more 
expansive definition for data literacy that included many of the ideas discussed during the 
Workshop sessions. She wrote that data literacy is, “The ability to co construct a 
framework to collect, collate and analyze data with a degree of integrity and credibility. 
To interpret data in various forms and recognize the value and shortcomings of the data 
set towards making decisions.” While Manisha was still focused on data literacy as a tool 
for making decisions, she had a more expansive understanding of analysis and the role it 
could play in making those decisions. In her interview Manisha also expressed her 
expanded conception of data literacy. She said,

Data literacy is the interpretation of information. It could be qualitative, it 
could be quantitative, but how do you take out what's pertinent and what's 
not. And even in the process of doing that, what are you including and what 
are you excluding? And are you doing it consciously? So, to be aware of 
inherent biases or your framework in how you're collating data, and then 
how you're using it, a stronger word is manipulating it, to serve whatever 
purpose you want or objective you want.

In this quote, Manisha described a lot of the thinking around what data is and the role it 
can play that occurred during the workshops. Rather than thinking about data literacy as
procedural, she had begun to think about it as a way of think and a way of engaging with data that was more robust than her original conception. She followed that description with a statement about how the Workshop Series supported her in her own data literacy growth saying, “[the PD] has also made me re-question that framework, the biases inherent, or what it is. And it's made me ask those questions of myself and therefore it will translate in the classroom.” Manisha reflected in this statement that participating in the PD caused her to think more for herself about the role and purpose of data. She acknowledged that these shifts in her own knowledge would also translate into her teaching in the classroom.

Despite being one of the teachers most focused on her students’ math abilities and how they could create barriers to learning data literacy in the classroom, Manisha did not have a strong enough sense of what STEM integration entailed when she came into the PD to make the connection between her students’ math abilities and the role of STEM integration in her classroom. On the pre-summer PD survey, in response to the question “In your view what is STEM integration as applied to a high school science classroom?” Manisha wrote, “STEM integration is simply adults figuring out that nothing exists in isolation - everything is connected! Science, Technology, Engineering and Math are different 'specialized' categories established by historical figures.” Even though she had been explicitly prompted to consider STEM integration as it related to a high school classroom, she did not consider the connection. Instead, she equated STEM integration to something that adults were doing and the way it had structured the isolated STEM fields in the past.
By her post-implementation survey, Manisha had begun to realize the role that STEM integration could play in her own classroom writing,

STEM integration is the solution use to resolve the previous issue where teaching/learning is separated into silos of the subjects labelled Science, Technology, Engineering and Math. Now the STEM idea is to recognize it first and foremost as connect and whole, and next to teach/learn/experience the disciplines as integrated, holistic and interconnected forms/disciplines.

In her interview, Manisha also expanded on her developing view of STEM integration in her own classroom, reflecting on the need to be connected to students’ knowledge of both technology and math in order to successfully teach data literacy. She said,

Clearly our kids are more ... calculator, everything's about the calculator, they're not learning math the way my generation learned math. The calculator was a tool. It wasn't a must have like a laptop, right. So, we have to embrace that and not go back to what we and the generation and the time and expect, and then judge our kids for those gaps, right? So, that's become very clear to me. And this has given me much more comfort with that.

In this statement, Manisha expressed her acceptance that the way her students interact with technology and math, and the preconceptions they bring about both are perhaps different than what she used to think or expect. She discussed how the workshop discussions gave her a clearer understanding and more comfort with meeting her students where they are with their math and technology preconceptions as part of teaching science.

One component of subject matter knowledge development that allowed Manisha to develop a more nuanced approach to STEM integration was her own development with Google Sheets.
From the beginning of the summer PD when Google Sheets was first introduced as a major component of the Bioinformatics PBL Unit, Manisha expressed her misgivings and lack of confidence in her technology skills, writing in one of the Discussion Forum posts, “My knowledge is weak on technology and apps, and all things digital. I take longer and am prone to anxiety over that when familiarizing myself to new technology than my more adaptive or faster assimilating colleagues.” From classroom observations conducted in Manisha’s classes at the beginning of the project, her lack of confidence with Google Sheets was clear. When she attempted to demonstrate how to do some of the data analysis activities for the students on her smartboard she would get confused and make a number of errors and be unsure how to correct them. After reflecting with Manisha off record after the second lesson like this, we decided that I would teach a portion of the next class, which I did. I also conducted a few additional informal training sessions with Manisha either right before or right after class, or while students were working on an individual assignment or reflection quietly. Through this continued support, I was able to watch Manisha’s skills and confidence with Google Sheets grow over the next few months. In fact, Manisha felt so much more confident in her own Google Sheets abilities, that once the students had collected their data and were ready to analyze it, she paused and went back to some of the original Google Sheets introduction lessons and retaught them to the students. In a comment in my Classroom Observation field notes for one of these lessons I wrote, “[Manisha] remembered how to show the students the trick about pulling the formulas across! All the students are actually following along this time.” Whereas Manisha had lost many of the students’ attention during her first attempt to teach data aggregate calculations in Google Sheets, the second
time through she was able to hold the students interest, partly due to her increased
certainty and smoothness in modeling the process.

Manisha spoke extensively to her increased knowledge of and confidence with
Google Sheets in her post-workshop written reflection and interview. In her reflection she
wrote, “Google Sheets which is clearly an invaluable tool - has grown from this step by
step experience. My previous familiarity was limited - this process has enabled me to use
cells to formulate and translate to graphic representation, with and for students.” While
the abilities to write formulas into cells and create graphs can be considered basic Google
Sheets skills, Manisha had come a long way from where she was. In her interview,
Manisha reflected on the power of having me in the classroom to support her in her
learning. She said, “It's like when you're in the classroom and you've shown us a few
tricks on the Google Sheets, it's valuable and there's so much more, and then it's not just
the one time, but it's the comfort with it, for myself and the kids.” In this quote, Manisha
expressed how my showing her “tricks” in Google Sheets increased her comfort with it
and that not only supported her learning, but also her students’ learning.

With Manisha’s increased knowledge of and confidence with Google Sheets came
an increased belief in the importance of it. In her interview Manisha said about Google
Sheets, “It is a very contemporary tool that is not, it really became very clear to me, it's
not a maybe, it is, yes, it must be taught.” She then added to this later in the interview by
recognizing how her increased realization of the importance of Google Sheets as a tool
for students to learn helped her overcome some of her insecurities with it. She said,

My Google Sheets practice skills suck, but you know what, this is an
important skill, and [this program] brings it to light, and you confront it and
then you get past your own little insecurities and you’re like, hey, hold on. I can do this. And so, I think that's quite, it's wholesome.

While Manisha’s subject matter knowledge and skills for data literacy increased during this project, her knowledge of strategies for teaching with data also grew and shifted to match her new comfort with and focus on bringing data literacy into the classroom.

7.2.3.2 PCK for Data Literacy – Manisha. Manisha had a very low conception of PCK at the beginning of the Workshop Series. Though all of the teachers except Hallie said they had not heard of PCK upon initial query, Manisha was still struggling with the concept after the first asynchronous assignment when teachers were assigned reading and a video to watch introducing them to the concept of PCK. In her initial written reflections on PCK after engaging in the assignment, Manisha was still unclear about the difference between content knowledge and PCK and did not make a distinction between the two in her initial written reflection on PCK. By the end of the workshop sessions, Manisha had a clearer sense of the concept of PCK and how it applied to her own knowledge and growth. In her post-workshop reflection, she identified “Knowledge of Student Understanding” as her strongest component of PCK and “Knowledge of Strategies for Teaching Data Literacy” as her weakest and the one that she most hoped to continue to work on. However, during her interview she made clear that rather than breaking it down and thinking about the individual components of PCK, the idea that had the greatest effect on her learning was simply identifying the concept of PCK, reflecting on, and beginning to understand that it was a separate component of knowledge that she needed to focus on and develop.
When asked whether participating in the Workshop Series changed her perspective on teaching data literacy, Manisha responded,

Yeah, it did. Because some things that were sort of simmering in the background of your mind and they're there, but they're just thoughts that take flight really. And then when somebody else says it and I was like, okay. It brings the simmer to boil in your mental thought and then somebody else says it and then it's like, hold on, this is steaming up here. It's not a simmer at the background anymore. Now it's at the forefront in your face, hey, confront this.

Taking part in the conversations during the Workshop sessions led Manisha to grapple with concepts that hadn’t really formulated in her brain previously. By listening to the other teachers and interacting with them, it brought the ideas and concepts of PCK to the forefront for her as knowledge components that she felt she now needed to confront. She added to this later in the interview when asked what her greatest takeaway was from the Workshop for her learning. She said,

My greatest takeaway, I mean, I love the collaboration. I love the fact that you pushed us to think about things that we didn't realize we were doing at times or not doing, right? We do it because it becomes part of our thing. But it's like, you had mentioned once is to make visible what you are doing. And because of those collaborations, it allowed us to zoom into things that were happening, that you didn't realize that were happening. And then it starts now bringing up other questions on what you're doing. And I was like, even stopping after something and I was like, oh, that's what I was doing. Oh gee, I missed that.

In this passage, Manisha referenced how I had described one of the goals of the Workshop Series at the very beginning, which was to make their practice visible in a way that allowed them to discuss and refine it. She reflects on how the opportunity to collaborate with other teachers to surface knowledge about strategies for teaching led her
to look for those strategies in her own teaching and changed the way she reflected on her teaching process. The workshop developed for Manisha a new habit of questioning her own teaching and reflecting on the actions she was actually taking in the classroom.

As displayed in Chapter 5, Manisha held a strong view about the unreliability of data and the biases inherent in most data. However, that personal orientation to the nature of data had not translated into her classroom. In her interview, Manisha spoke about how that personal view of the nature of data had changed her orientation to teaching data and how she planned to approach teaching data literacy in her classroom. She wrote,

The questions I pose to my students are clearly going to be different. The part about the process is going to be so much deeper, I think, because the questioning is different. The process is not just, here's the data and it's presented, and then you're judging it, basically. It’s questioning and the quality of the questions. The quality of the question differs dramatically to what it was previous. The idea of cleaning data and framing data. But what's happened before to get that? What is the credibility of that data? What's framed and what's on the outside of that framing? What's the agenda? There's so many different questions you're asking in the cleaning process itself. And why are you leaving it out purposely? Why you not, why you're not addressing it? So, there's a lot that has come out of the process and I'm still not done.

In this quote, Manisha focused on the strategy of supporting students in questioning data and spoke extensively about all the ways she is now thinking about encouraging students to question data in ways that are “dramatically” different to how she would work with data in her classroom before participating in this intervention.

These three case studies offered insight into how the teachers were developing personal subject matter knowledge and PCK for data literacy. The next section returns to collective knowledge to explore how subject matter knowledge was developed.
collectively in the collaborative discussions, and themes for PCK changes across the cases of the individual teachers.

7.2.4 Changes in PCK for Data Literacy

While all of the teachers had different journeys with their PCK development over the course of the intervention, the one theme that was apparent across all of their development was the notion that explicitly thinking about and talking about PCK and the components of student understanding of data and strategies for teaching data shifted their ability and enthusiasm to reflect on their own knowledge and practice. As Hallie deftly pointed out in her interview, we were not working from an established corpus of knowledge components for teaching with and about data, so measuring growth in any one area was complicated as we were surfacing the knowledge we were attempting to measure growth in as we went. Despite this, the teachers were confident that participating in the workshop shifted their thinking on how to teach data literacy in their classroom in noticeable ways. The final chapter summarizes the findings on PCK for data literacy and the context and design which limited those findings, as well as how this research fits into the larger fields of learning sciences and teacher professional development.
CHAPTER 8. DISCUSSION

This study sought to better understand how PD can support teachers in surfacing their existing knowledge of teaching data literacy in a high school science classroom and use that identified knowledge to further develop PCK for data literacy in order to enhance STEM integration. The results from the previous three chapters presented the beginnings of a framework for PCK for data literacy and showed how the design of the Extension Workshop Series allowed space for the participating teachers to surface their existing knowledge and discuss data literacy and PCK in a way that led to a deeper more robust understanding of both. An important finding from this research is that experienced teachers already hold a lot of knowledge about teaching data literacy in a STEM integrated way and that creating space for them to highlight and reflect on that knowledge allows them to focus on and refine it.

While experienced teachers may already hold PCK for teaching data literacy effectively, without the space and structure to reflect on it, it does not always convert from personal PCK into enacted PCK through implementation in the classroom. Additionally, while expert teachers may hold that knowledge, less experienced teachers may need space to develop it, and developing an understanding of a framework for PCK for data literacy will allow for novice teachers to build that knowledge. This study provides an advancement towards both an understanding of what PCK for data literacy might look like and some methods that can be employed to allow teachers space to build it. Through the process of talking about their knowledge of teaching data literacy, the teachers in this study not only began to answer the question of what PCK for data literacy might compose, but also deepened their understanding of their own knowledge and
experience. This study asked: 1) What components of PCK for data literacy are surfaced by the participating teachers during the PD experience? 2) How the PD components appear to support teachers’ development of subject matter knowledge and PCK for data literacy? And 3) What subject matter knowledge and PCK for data literacy was developed by teachers during PD experiences that have been designed to explicitly focus on data? Each of these research questions has been answered in the previous chapters, and in the remainder of this chapter, the findings related to each are connected back to existing literature so that implications and recommendations for the future can be made.

8.1 The Components of PCK for Data Literacy Surfaced and Identified during the PD

The primary assumption on which this study was built is that there exists special PCK for teaching data literacy that falls outside standard PCK for teaching science. The findings presented in Chapters 5 support this assumption. Though the goals for teaching and learning data literacy, students’ understanding of data, and strategies for teaching data literacy have all been studied to some extent (e.g., Lee & Wilkerson, 2018) there have not yet been any studies conducted on what teachers know about these concepts and how they apply them in the classroom. Figure 8 demonstrates a preliminary collection of some of what teachers know about teaching data literacy. The following sections compare that PCK to the limited knowledge that the field has about teaching data literacy to high school students.
8.1.1 Orientation to Teaching with and About Data

The teachers surfaced three primary components of orientation to teaching with and about data that fell outside the framing of goals. Those were: a) data literacy must be grounded in relevance to students’ lives, b) the tools of data literacy, specifically Google Sheets, are powerful, and c) data has the power to manipulate. The first of those components of PCK is one that is widely discussed in the literature (Wilkerson & Polman, 2020). Lee and Wilkerson (2018) wrote that students engage differently with data that is situated in familiar and meaningful contexts and have more opportunities to grow their competencies for data literacy when they see connections to the data. Wolff and colleagues (2019) also noted the importance of engaging students with data that is familiar or relevant to their lives. While clearly an important component of knowledge for teaching data literacy, it is certainly not unique to data literacy as a topic, but rather a general orientation to teaching pretty much any topic or discipline to any age group. Where the unique nature of PCK for data literacy arises is in what teachers know about their students and how to engage them successfully in data literacy development.

The second component of PCK surfaced by teachers relayed a belief in the power of the tools of data literacy, in this case specifically Google Sheets, and the need to explicitly teach it to students. As discussed earlier in this chapter, the basic tools component has been underappreciated in most of the literature on data literacy for K-12 students. There has been a broader focus on developing tools specifically designed to teach data literacy (e.g., Erikson et al., 2019) or on using specific tools for modeling and visualizing data (e.g., Stornaiuolo, 2020) however fewer studies have focused on using basic widely accessible tools such as Google Sheets in connection with data literacy in K-
12 classrooms (Edwin, 2015). The use of Google Sheets does show up more frequently in practitioner journals and resources (e.g., Ridgway, 2019; Rivet & Ingber, 2016) which shows that teachers have a predisposition toward it as a tool that is simple and accessible but also powerful. The additional benefit of Google Sheets is that, as many of the teachers mentioned, students can be shown uses for it outside science such as tracking and calculating finances which may make it more inherently motivating to students than a custom tool which creates pretty graphs of specific data but doesn’t have use in their life outside of school.

Finally, the last component of data literacy surfaced by teachers in their orientation to data literacy, the belief that data has the power to manipulate is a concept that has been picked up by those studying data ethics and the way that a critical literacy approach can be applied to data literacy (Pangrazio & Selwyn, 2019; Pentland, 2013). Pangrazio and Selwyn apply a critical literacy approach to teaching about personal data that is designed to give students greater agency over the data that they produce and that is collected on them. The Personal Data Literacies frame they propose centers the importance of always considering data in context and being aware of the ways that personal data can be accessed and used by external forces for manipulation and control in order to regain agency over personal data and the stories it can tell. While the teachers were applying this approach to all data rather than just personal data, the critical data literacy framework proposed by Pangrazio and Selwyn could be modified and expanded to apply to the larger world of big data and its potential for less-savory outcomes identified by Pentland (2013).
As has been emphasized by other literature, data is constantly being used to make choices for us by social media apps, map app and economic choices that affect our health and access to resources (Dorsey & Finzer, 2017; Erikson, 2020). By centering the power that data potentially has over us as a major component of understanding and interacting with data, the teachers are focusing on the importance of this orientation to data, and in fact pull it into every other category of data literacy. While a focus on the procedural aspects of data literacy is important, this in another example of how addressing the conceptual components of data literacy, especially around data usage to influence students live is a way to potentially get students more invested in developing their own data literacy.

Freidrechen and colleagues (2011) in their exploration of orientation to teaching and learning as a component of PCK for science, found that one important indicator for orientation to the nature of science was an epistemological belief about what counts as science. Indeed, the teachers in this study got into an extensive conversation trying to determine what counts as data. However, they didn’t come to any definite conclusions or definition other than to conclude that it is significantly more expansive than what their students think it is. As Hallie pointed out and was presented in Chapter 5, “If you don’t know what it is and that would be data in general, then that's already very alienating and it's going to be tough to develop a relationship to something that's undefinable for you.” The problem with this of course is that what counts as data changes across fields and disciplines but considering how to teach the expansiveness of data is definitely a component of PCK for data literacy that needs to be developed for teachers to successful engage students with data literacy.
8.1.1.1 Goals of Teaching Data Literacy. There is still disagreement in the field about the learning outcomes for data literacy (Frank et al., 2016; Kjelvik & Schultheis, 2019). As discussed in section 3.3.1, there is a range of ideas in the field for this orientation from procedural to more conceptual. The teachers, despite being more conceptual than procedural in much of their development of content big ideas for data literacy, the themes that emerged for their orientation to the goals and purpose of teaching data literacy were procedural and three fold, a) learning to interrogate and question data, b) learning to tell a compelling story with data, and c) making interdisciplinary connections. The first two of these goals are in line with many of the goals identified by the literature on data literacy. Erwin (2015) in designing and evaluating a data-based project curriculum for middle and high schools students, gave as one of his primary goals that students should be able to consider the story and guiding questions for the data. Many studies have emphasized the important of learning to develop inferences and scientific arguments from data (e.g., Lee & Wilkerson, 2018; McNeill & Krajcik, 2011; Rubin, 2020) which is similar to how the teachers were thinking about learning to tell a compelling story, but the storytelling goal is more expansive because it involves data communication as well, which was another objective of Erwin’s data literacy unit. Learning to interrogate and question data is another way of saying attend to the context which is a goal stressed by Rubin (2020), Hardy and colleagues (2020) and Pangrazio and Selwyn (2019).

These first two goals for teaching and learning data literacy identified by the teachers also build upon a history of the scientific method as a process to be taught in science classrooms (NRC, 2012) and the science practices in the current NGSS which
include asking questions, analyzing and interpreting data, engaging in argument from evidence, and communicating information (NGSS Lead Teachers, 2013). However, the goals identified by the teachers took those traditional science practices and applied them to data literacy by focusing them explicitly on data as the primary content rather than science. Yes, data literacy should be grounded in authentic context within science or engineering, but it should not be taught as just a tool to engage with and develop science concepts but as a content area in its own right and the first two goals for teaching data literacy surfaced by the teachers suggest that they held that orientation to teaching with and about data.

The final goal for teaching data literacy surfaced by the teachers was that of making interdisciplinary connections. As presented earlier in this chapter teachers did not make the connection between this goal and the role of STEM integration in their classroom. Teachers did not seem to think of data literacy as explicitly STEM integration but as a new topic to teach in science. However, they did still identify the need to make connections between the disciplines, especially to math, explicit to students. This is a goal identified as important in much of the literature on STEM integration (English, 2016; Gardner & Tillotson, 2019; Kelley & Knowles, 2016; NRC, 2014). There is clearly still work to do with this particular group of teachers to develop a stronger orientation in them towards using data literacy as a bridge into STEM integration in their classrooms. There is also still work to be done by the field to more solidly identify the goals of data literacy for K-12 students so that teachers can use those to grow their PCK for data literacy (Kjelvik & Schultheis, 2019.)
8.1.2 Students’ Understanding of Data

While research into teachers’ knowledge of student understanding of data is limited, some research has been done on student understanding of data. So, I can make a comparison between what this particular group of teachers surfaced as knowledge of how students think about and work with data to what the literature knows so far about how students think about and work with data. The teachers in this study identified six main themes in what they knew about students understanding of data: a) students are not inherently interested in data literacy, b) students have a fear of numbers, c) students’ conception of data is constrained, d) students are not used to questioning data, e) students struggle with the ambiguity of real-world authentic data, and f) students make and use assumptions rather than inferences when working with data.

In relation to the first of these components of PCK about students’ interest in data, Harris and colleagues (2020) conducted a study with youth participating in community and citizen science projects. They found that the process of data collection and analysis was not inherently interesting to students, but that some became more interested if they saw how it was relevant to them and their lives or if they truly believed it could actually be useful to people with power to make change. As discussed in the previous section, making content relevant for learners is a universal pedagogical tool, however it is slightly more nuanced for data literacy because it isn’t just about grounding the data in a concept that might peak students interest such as we attempted to do with the Bioinformatics PBL unit and asthma, but it is also about creating the motivation to learn the tools of data literacy, not just the underlying content. While students might be interested to know how the air quality affects their health, that doesn’t mean they are automatically interested in
being the data producer in order to gain that information. Harris and colleagues’ finding is important because it shows that students can be motivated to learn the tools if they believe that the data they produce will actually contribute to change making. The teachers in this study already had this knowledge. They understood that in order to make students interested in developing data literacy, they had to not only ground it in content they are interested in but convince them of the power of the tools as well, as Will did when explaining how Google Sheets could be made to create budgets. This motivation to develop the skills piece seems to be underrepresented in the literature on what students think about data literacy.

The second and third components of PCK surfaced by teachers should be considered together because while students’ fear of math is a widely documented barrier to development in math and science (e.g., Itter & Meyers, 2017; Mutodi & Ngirande, 2014) it takes on more relevancy to data literacy specifically when combined with the third component about students’ constrained conception of data. While some few studies that examine data literacy in K-12 classrooms attempt to define what they mean by data (e.g., Pangrazio & Selwyn with personal data), most do not. This is in my opinion, and the opinion of the teachers in this study, an oversight. In one study Gebre (2018) measured students’ understanding of data and its relevance to their lives external to any particular intervention. He found that students have a conception of data as survey and experiment related and having no connection to their daily lives. He was particularly surprised to find that students made no connection to social media in their conception of data or to the “abundant open data that is freely available from local and federal
institutions” and that this lack of connection should be “a wake-up call for educators” (p. 339).

I wonder if that wake-up call should in part be directed at researchers. The teachers in this study seemed to be already aware of this component of students’ thinking about data. While some literature on data literacy calls for changing students’ mindsets to make them more aware of the data in the world around them and how it connects to their lives (e.g., Doresy & Finzer, 2017; Hardy et al., 2020) few studies seem to be actively engaging with how students define what even counts as data and how that might affect their ability to engage with it.

The final three components of PCK surfaced by the teachers have had more traction in the literature as they are all aligned to some degree with findings from Lee and Wilkerson’s (2018) review of the literature on data literacy. Lee and Wilkerson found that one theme in the literature was that students sometimes view visualizations as illustrations rather than as tools to be used and interrogated. Another theme was that students can have a tendency to view data as “true” rather than containing uncertainty, and a third theme was that students are more likely to interpret data in ways that support their existing assumptions or theory. These three themes are more or less aligned with the final three themes identified by the teachers, though they added to and generalized the first two. Hardy and colleagues (2020) also found that students were inexperienced in interrogating data.

While the knowledge that the teachers in this study held about how students think about and interact with data was not new information to the field, what they chose to focus on as the most important things to know and consider about students’
understanding of data was different in some ways from what the field has been focused on for understanding how students think about data. While Gebre (2018) was the only study found that actually sought to measure and understand students’ conceptions about the nature of data, Lee and Wilkerson’s (2018) review highlighted a number of other themes that did not surface in the teachers’ knowledge including that students often bias “fair” treatment in choosing samples rather than using true randomness, and that students struggle with viewing data sets in the aggregate rather than as a collection of single points. While these findings about the way students think about and interact with data were part of the foundation on which the six components of data literacy were built and identified, it is unclear whether or not the teachers in this study held this knowledge about students, or whether they simply didn’t consider it a priority. While researchers and external parties in the learning sciences field can continue to study students’ understanding of data, it is often the teachers who hold the most knowledge about their students and what they know. Perhaps the components of student understanding of data that these and other teachers find most relevant should hold higher priority among future research in this area.

8.2.3 Strategies for Teaching With and About Data

Again, while few if any studies have attempted to measure what teachers know about strategies for teaching with and about data a number of theoretical suggestions for strategies have been made in the literature (Lee & Wilkerson, 2018; Wilkerson & Polman, 2020, Wise, 2020; Wolff et al., 2019). This section seeks to compare the strategies surfaced by the teachers to the ones suggested in the literature. The five
strategies surfaced by the teachers, discussed in Chapter 5 and displayed in Figure 8 were a) making connections to students’ lives, b) making data tangible, c) using storytelling to build inferences, d) interrogating data by questioning assumptions, and e) scaffolding the tools of data literacy through repetitive modeling including intentional modeling of common mistakes.

The first strategy surfaced by the teachers, making connections to students’ lives, is again ubiquitous in the literature. One of the two core pedagogical commitments provided by Wilkerson and Polman (2020) was that data literacy should be grounded in meaningful contexts. Wolff and colleagues (2019) list activities that include data that is familiar and relevant to students as one of their design principles for supporting the development of data literacy and Lee and Wilkerson (2018) suggest that the context for data pursuits should be familiar and meaningful. The question here is not really with the general strategy but with the specifics of it, which include matching this strategy to knowledge about the particular students you are trying to engage and their interests and experiences. What entails a “familiar, relevant, or meaningful” context will potentially be different for every student. While generalized vocabulary is useful in theoretical frameworks for teaching, it is less useful for actual teachers attempting to teach. While the teachers in this study acknowledged this as a general strategy for teaching data literacy, they struggled with the specifics of it. As Mary said in session 8 of the Workshop, “Drilling down a little more specifically on how to get our students engaged in this work would help me and probably help other teachers.” So, more research into the specific of how to get students at certain ages and in certain demographics interested in data literacy would be helpful for teachers.
The second strategy surfaced by the teachers of making data tangible is not one that appears in the literature on teaching very often. Wolff and colleagues (2016) in their suggestions for future work suggest that a useful future topic of investigation might be “how to make data more salient in order to help learning, for example through tangible data” (p. 24). I and the teachers in this study agree with Wolff and colleagues that this is an under-researched area of data literacy education, how to make not only the collection of data, but the analysis of it more hands-on than information on a computer screen. One parallel but complimentary field that has taken a much more prominent focus to the idea of making data tangible is the field of computer programming and the maker movement (citation). This also connects to the third strategy surfaced by teachers of using storytelling as a framework for building inferences. There has been a strong focus in the programming and makerspace fields on supporting students in telling stories through programming tangible artifacts (Shaw et al., 2021; Stornaiuolo, 2020). Based on the knowledge that teachers in this study were building for strategies to engage students in data literacy that focus on building tangible artifacts so prevalent in the programming literature could benefit the study of strategies for teaching data literacy.

The idea of building stories does surface in the literature on data literacy to some extent. Rubin (2020) discusses the need to focus on the context of data in order to understand the full story of the data. Wise (2020) suggests that one strategy for teaching data literacy should be to support students in understanding that the same set of data can be put to multiple purposes depending on the audience it is trying to reach. She suggests that one method for developing this skill in students is to have them, after completing a data analysis, go back to the beginning and see if they can tell a completely different
story with the same data. This strategy is very similar to one of the examples provided in Chapter 5 for the theme of using storytelling to build inferences in which Mary suggested having students all work from the same data set to create visualizations based on a different question they were trying to answer, and then compare the different visualizations that resulted.

The fourth strategy surfaced by the teachers takes a broad strategy stemming from a broad goal and makes it more specific and attainable. As mentioned earlier the goal of supporting students in interrogating data is common throughout the literature on data literacy (e.g., Hardy et al., 2020; Lee & Wilkerson, 2018; Rubin, 2020; Wise, 2020). However how to reach that goal is not well articulated. Wise (2020) suggests that teachers build a bridge from the process of collecting data to the process of interrogating data by having student reflect on their own process and then apply that to what they do and don’t know about external data. Lee and Wilkerson (2018) provide a suggestion for correcting student misconceptions about data visualizations through “reflection on how a given data representation works and corresponds to the situation being modeled” (p. 8). This suggestion is somewhat aligned with the more generalizable strategy that the teachers surfaced which was to directly engage students assumptions in order to lead them to ask questions. By allowing students to first identify their assumptions about data and then asking probing questions that lead them to either defend or rethink those assumptions teachers can create cognitive dissonance for students which is a well-established tool for supporting people in changing their views (Cooper, 2011).

Finally, the last strategy surfaced by the teachers was scaffolding the tools of data literacy through repetitive modeling and explicitly demonstrating common mistakes and
how to fix them. While the under-representation of focus on basic tools for data literacy such as Google Sheets in the literature has already been discussed. This particular strategy is also an example of STEM connections in data literacy work. In the research on uptake of new technologies in classrooms and the TPCK that teachers need to be successful at integrating new technologies, the components of modeling discussed by the teachers are already established methods (Niess, 2014; 2017). This suggests that a greater focus on the role that data literacy plays in STEM integration and therefore more integration between the fields of technology education and data literacy might be beneficial to teacher and student uptake of data literacy development in classrooms.

8.2 How the PD Components Supported Teachers’ Knowledge Development

While the exploratory nature of this study produced an unclear answer to the third research question about the extent to which teachers actually developed additional PCK as a result of this intervention, it was clear that they became more aware of the concept of PCK and more inclined to reflect on their practice within the framework of knowledge for student understanding of data and strategies for teaching data literacy. It was also clear that the teachers gained a lot of confidence in teaching data literacy as a result of participating in the intervention. The findings in Chapter 6 suggest that participating in the video reflections was the biggest source of this confidence, but that developing the learning community also contributed to it. Previous research has shown both of these strategies to be beneficial to teacher learning (Donnelly & Hume, 2015; Loughran et al., 2012; Wilson et al., 2018).
Additionally, engaging in development of big ideas for data literacy and subsequent completion of CoRes for those big ideas supported teachers in developing a better sense of the strategies they were using, language to describe them, and tools to more productively reflect on them. Unpacking the components of data literacy in order to determine the big ideas led teachers to, if not increase their subject matter knowledge of data literacy, certainly approach it with a different perspective. Nilsson and Loughran (2012) had a similar finding in their work with CoRes to develop PCK with pre-service teachers writing, “Identifying Big Ideas for the topic offered access to the way in which the student teachers conceptualized the topic as a whole” (pg. 708). Additional previous research has shown that CoRe development can support teachers knowledge growth (Aydin-Gunbatar et al., 2020; Carpendale & Hume, 2019; Cooper et al., 2015; Loughran et al., 2012; Vossen et al., 2020) However, most of those studies were conducted within disciplines that had well established big ideas and frameworks for PCK that the teachers could use to build off of. The field of data literacy does not have that. So, while one implication of this research is to suggest that these tools continued to be used to help teachers develop knowledge and confidence for teaching data literacy, another implication is that the field needs to engage further with teacher knowledge for teaching data literacy. As Hallie reflected in her interview it would be nice if teachers had an established set of knowledge about students’ preconceptions for data and teaching strategies to engage those preconceptions, but the field is not there yet, partly because there has been so little research conducted on teacher knowledge.

This suggests that there may be a potential refinement necessary to the field-accepted best practices for PD presented most recently in a review by Darling-Hammond
and colleagues (2017). While these widely accepted best practices were utilized throughout the Bioinformatics PD as a whole, and the practices of enabling collaboration among teachers, and dedicated time for feedback and reflection on practice were especially present in the Extension Workshop, the list of best practices provided includes a focus on disciplinary content through concepts and pedagogies without considering the amalgam of the two. While best practices for PD focus on teacher learning, making it active, sustained, and grounded in prior knowledge, they mostly ignore the explicit engagement of PCK. Perhaps, if engaging teachers’ PCK for teaching the specific content focus of a PD became a more prominent component of designing highly effective PD, it would create motivation in the field of teacher education and professional development to build more comprehensive and robust models for PCK that were specific to particular topics and goals which would support teachers in identifying and growing best practices for teaching those specific topics.

The use of video reflections and CoRes in PD provide tools to guide teachers towards best practices for using authentic complex data in their classroom, but first a set of best practices needs to be established so that teachers can be supported in using those best practices to teach data literacy. Previous research has found that teachers often lack the content knowledge to engage in integrated STEM activities such as the use of complex authentic data (Aslam et al., 2018; Kelley et al., 2020) and this study supports that research specifically for the component of complexity in data sets which the teachers need additional support to develop knowledge and confidence in. However, the larger struggle for the teachers in this study was not the subject matter knowledge of data literacy but the strategies for how to teach it. While the teachers already had knowledge
of strategies for teaching science and engaging with data in the context of a science class and were able to reframe that knowledge in a way that applied to data literacy, most of them still felt at the end of the intervention that they were in need of additional support to grow their strategies for teaching with and about data.

8.3 Changes in Teacher Knowledge for Data Literacy

This study focused on the definition of data literacy provided by Kjelvik and Schultheis (2019) in Figure 1 as existing in the overlap between quantitative reasoning, data science, and an authentic science content in context. They defined data literacy simply as the ability to “understand and evaluate information obtained from authentic data.” As demonstrated in Chapter 7, teachers came in with this basic procedural understanding of data literacy, though without the authentic context component. Through the intervention teachers’ conception of data literacy expanded to include more conceptual understandings which actually put them more in line with the definition developed by Wolff and colleagues (2016) in their review of the literature on data literacy which included a greater focus on asking questions as well as considerations of ethics in data and creative skills. Another distinction between the way the teachers’ were defining data literacy and the way Kjelvik and Schultheis (2019) present it is that despite developing robust understandings of STEM integration and despite discussing explicit connections between data literacy and math and technology, teachers did not seem to explicitly make the connection between data literacy and the role it plays in STEM integration.
Understanding data literacy as a form of STEM integration was a primary goal of this study. The teachers did talk often throughout the workshops about explicit connections between math, technology, and the data literacy content they were engaging with in the project, so the connections were there (See section 5.1.3.3 for examples of this.) However, they spoke about the technology and the math as being part of data literacy without making the intellectual jump to the bigger picture of how these components of data literacy, the math skills, and the need for technology, made data literacy into a conduit for STEM integration. This suggests that greater emphasis should be placed on data literacy as an example of STEM integration, though it is unclear how making sure the teachers see that connection would actually change the way they think about or teach data literacy in their classrooms as they were already considering the components of data literacy in relation to the connecting math and technology skills.

Teachers in this study had the opportunity through the five CoRes built during the Workshop sessions and the big idea brainstorming that led up to them to make visible their thinking on and subject matter knowledge of five of the six components of data literacy. The big ideas surfaced by the teachers for each component of data literacy are displayed in Table 7 in Chapter 7. In this section those big ideas developed by the teachers will be compared to the indicators of mastery from the literature and the level of alignment with previous research will be discussed.

8.3.1 Understanding Data in Context Primarily as Consumers not Producers

Three indications for mastery of the component of data in context from the literature are a) asking probing questions about the origins of external data; b) providing
context for personally collected data; and c) considering how bias stemming from intention, existing theory, and the tools of data collection can affect data (Hardy et al., 2020; Rubin, 2020; Wise, 2020). While the teachers were considering the big ideas of data literacy more cognitively than the procedural way they are more often presented in the literature, there is some alignment between what the teachers came up with and what the literature deems is important. Teachers were focused mainly on the first indicator in their development of big ideas for data in context, their recognition that “Interrogating data is useful and necessary” speaks to the first indicator of mastery which is asking probing questions about the origins of data. In fact, most of the other big ideas teachers developed for this component are examples of that first indicator in more nuanced ways. “Data is often manipulated for non-scientific reasons” suggests the importance of the questions about who collected the data and what is the reason for the collection. “People make decisions about when and where to collect data” is an indicator that those questions need to be asked as well.

The big idea about data being manipulated also suggests that the teachers were thinking about bias within the component of data in context, but interestingly, this indicator appeared within the other components as well with “Complex data includes a human factor” as a big idea within the complexity of data component, “Decisions about the visualization affect the story that is being told” within the data visualization component and “Data is biased, there is no perfect data” as a primary big idea within inference with data. While these big ideas suggest that teachers are thinking a lot about the biases stemming from human intention, the big idea from inference with data “The tools (and their capabilities and biases) that are used can impact our inferences” suggest
that teachers were also thinking about how the tools of data collection can affect data (Hardy et al., 2020). However, the indicator of mastery that was mostly missing from teachers’ discussions about data in context, was the need to provide context for personally collected data.

While classroom observations showed that teachers did encourage students to provide context when they were collecting data with their sensors, and Will shared an anecdote during one of the session meetings about how a student got a high CO reading in one of the classrooms and through discussion with the student they discovered that the class that had just ended in that room had been using Bunsen burners, the teachers did not focus on students as data producers during their brainstorming sessions on big ideas. Hardy and colleagues (2020) suggest that giving students agency to view themselves as producers of data is an important component of developing data literacy. While the Bioinformatics PBL Unit is designed to do that, and the teachers saw and spoke often about the power of having students collect their own data, the absence of a big idea connected to this process in their brainstorming on data in context is noticeable. For example, a big idea such as “It is important to record the context in which you collect your data” would have suggested a focus on this indicator. The literature shows that students are rarely placed in the role of data producers in high school classrooms, and when they are it is usually in highly controlled scenarios where the context will have little to no effect (Hardy et al., 2020; Kjelvik & Schultheis, 2019; Lee & Wilkerson, 2018). The result of this seems to be for this group of teachers, that even though they know the important of providing context for personally collected data, they don’t think of that as a big idea of data in context because they are predisposed to think of students as only
consumers of complex data. Additionally, the teachers were more predisposed to think of data literacy in a conceptual and intellectual way rather than procedural, perhaps again due to their lack of experience actually having their students collect messy data for which the procedural ideas matter more.

8.3.2 Variability in Data Deemed a Low Priority by Teachers

The teachers did not spend time discussing big ideas for variability in data. Since the teachers were given autonomy over choosing which components of data literacy they wanted to discuss, the lack of engagement with this particular component of data literacy is important to note. It can be assumed that teachers view variability in data as a low priority focus in their classrooms. There are a number of factors that could contribute to this including teachers’ impressions of their students interests and abilities, teachers’ beliefs about the practicality of variability in data skills, or their own knowledge and comfort with the content of data in variability. Rubin (2020) acknowledged that uncertainty and variability are particularly difficult concepts for people of any age and experience to comprehend. Our own experience with previous cohorts in this project has supported this (Miller et al., 2021). So, it is likely that part of why teachers place a lower priority on variability in data is because consciously or subconsciously they feel less comfortable with in themselves. This suggests that either more focus should be placed on developing this component of subject matter knowledge for data literacy in teachers, or there is some other explanation as to why teachers do not prioritize it which might shift it’s priority among the literature in the field if better understood.
8.3.3 Data in Aggregate, the What and the So What

The two indicators for mastery of the data in aggregate component of data literacy highlighted in the literacy are a) calculating measures of centrality for the entire data set; and b) applying those measures of centrality to describe patterns and themes evidenced across the entire data set. The first of these is implied by the big ideas suggested by the teachers that “Averages can be useful for comparing between data sets in a way individual data points are not” and “Different aggregate measures support different data visualizations.” In fact, the teachers go beyond the procedural indicator most often identified by the literature to include reasoning about why it is important to calculate aggregate measures. This focus on making comparisons, creating different visualizations, and identifying the interesting story in the outliers as the reasons why finding aggregate measures is important, shows that not only do teachers have the content mastery of this component, they are also making the jump to PCK to think about to explain the “so what” piece to their students to make learning data in aggregate relevant to them.

Teachers are also displaying mastery of the second indicator of data in aggregate implicitly in their big idea of “Finding patterns and trends can allow for connections between different variables” and again are identifying not just the procedural indicator but taking it a step further to suggest one reason why looking for patterns is an important skill to focus on. In the same theme of focusing on the conceptual over the procedural, the teachers identified a big idea for data in aggregate that I have not seen in the literacy on content goals for data literacy which is recognizing when aggregate measures are not appropriate. There is so much focus in the literature on ensuring that students know how to apply an aggregate lens (Konold et al., 2015; Lee & Wilkerson, 2018; Wise, 2020) that
the next step of knowing when to apply it and when not to has not yet drawn as much attention. While on one hand the teachers’ predisposition toward thinking about big ideas as conceptual rather than procedural could be seen as an indicator that they have less mastery over the procedural content, here in the case of data in aggregate it was clearly an asset that allowed them to expand beyond the procedural to think about not just the “what” as big ideas, but also the “so what”.

Interestingly, “Sample size matters” and “Multiple trials ‘smooth’ out variability from human error and other variables and uncertainty” both made it into the data in aggregate component of data literacy as big ideas despite typically being considered part of the variability in data component in the literature (Lee & Wilkerson, 2018; Rubin, 2020). As this was the last component explored in the Workshop and the teachers knew we would not have time to dig into variability in data, perhaps they were bringing up big ideas for data literacy that they wanted to discuss but hadn’t had a chance to despite them not fitting into the theme of the last session. Or, perhaps data in aggregate and variability in data are more intricately connected in teachers’ minds than many of the other components and thus maybe should be addressed jointly when assessing or developing teachers’ knowledge for data literacy.

8.3.4 Mastery of Data Visualization

Data visualizations are perhaps the mostly widely studied component of data literacy (e.g., Gebre & Polman, 2016) so there is a wider range of indicators of mastery that have been identified by the literature. Most of these indicators can be categorized broadly into two primary objectives a) attending to design features of data visualizations
(i.e. axes scales, titles, color choices, type of graph or chart) when both interpreting external visualizations and building your own (Gebre & Polman, 2016; Hardy et al., 2020; Lee & Wilkerson, 2018) and b) drawing inferences and building scientific arguments from both external and internally created visualizations (Finzer & Reichman, 2018; Lee & Wilkerson, 2018). When considering that the teachers grew to use the word “story” as a way to talk about inferences, the big ideas developed by the teachers show near complete mastery of this component of data literacy. The big ideas “The categories/axes/titles with which you design the visualization matter for how it’s interpreted” and “What you exclude in a data visualization matters as much as what you included” and “Data visualizations can be misleading and can/should be examined for integrity” all indicate a focus on attending to the design features when both creating and interrogating data visualizations. The big ideas “Data visualizations tell a particular story about a set of data at a particular point in time” and “Decisions about the visualization affect the story that is being told” point to a focus on drawing the scientific argument or “the story” out of data visualizations and also making decisions about how to create a data visualization that will make the argument or story clear.

This component of data literacy was the one in which the agency of the students as data producers was most clear (Hardy et al., 2020). Through the language choices it is clear that teachers were considering big ideas related not just to students interrogating data visualizations that had been created externally, but also related to students building their own visualizations. Again, this may be a result of the experience teachers have of teaching this aspect of data literacy in their classrooms. Three out of four of the teachers indicated on their pre-surveys or applications that creating graphs was one of the things
they already did in their classrooms, though on closer questioning, both Mary and Will shared that they usually created graphics for their students when working in Google Sheets and then simply had students type their numbers into a pre-created template. However, the creation of graphs from data has long been an accepted component of science classes and science standards (NGSS Lead States, 2013; NRC, 2014).

8.3.5 Inference with Data as Conceptual Rather the Procedural

The primary indicator in the literature for mastery of the component of data literacy is simply the ability to build a cohesive and sound scientific argument from data. One common way of scaffolding that skill is through the use of CERs (McNeill & Krajcik, 2011). This is a purely procedural indicator that is a skill built off of other understandings about data that are needed in order to succeed at that skill. While teachers did discuss CERs as a strategy for supporting students in learning to build scientific arguments, it was likely that they thought of it as a skill or tool rather than a big idea for inferences with data. Their discussion of the component of inference with data was purely conceptual. They focused more on the conceptual ideas that students needed to have about the nature and purpose of data in order to not only be able to build a scientific argument, but also understand why they need to.

The conceptual ideas that the teachers highlighted as big ideas for inference with data were indicators more of their PCK for students’ orientation to data than of their own subject matter knowledge for building inferences from data. “Anyone can make meaning of data” is an indicator of their knowledge that students do not feel connected to data or see its relevance in their lives. “Data is biased, there is no perfect data” is an indicator of
their own orientation to the nature of data and their interest in supporting their students to develop this same orientation. “Not limited to one conclusion, there could be numerous interpretations” is an indicator of their knowledge that students always want there to be one right answer. As a result of this conceptual focus on addressing students’ preconceptions toward data that might create barriers to successfully building an inference from data, it is hard to determine teachers’ own subject matter knowledge for building inferences from data. However, it does suggest that the literature may want to expand its understanding of this component of data literacy to recognize that there are a number of conceptual big ideas that act as building blocks and therefore can be barriers to achieving procedural mastery of inference with data.

8.3.6 Complexity of Data Sets Requires Knowledge of Tools to Master

As discussed in Chapter 7, this component was the one for which growth was visible in teachers’ implementation intertwined with growth in use of Google Sheets to scaffold complexity of data with students. The sole indicator for this component of data literacy in the literature can be inferred from Kjelvik and Schultheis (2019) and it is once again procedural. Mastery simply entails being able to work with and apply the other five components of data literacy to a complex data set. This indicator is deeply entwined with the ability to use a spreadsheet software such as Google Sheets. While the use of spreadsheet software and other digital data analysis tools has been studied some in the fields of statistics and data science education (e.g., Dichev & Dicheva, 2017) it has not made many inroads into the study of science education at the K-12 level. However, mastery of complexity of datasets needs to go hand in hand with mastery of a spreadsheet
software because there is no other way to apply the other components of data literacy such as data in aggregate, data visualization, and variability in data to a complex data set without the use of such software.

As such, it is necessary that studies of teacher and student learning on data literacy include skills with data analysis software as an indicator of mastery. If one of the goals of data literacy is to learn to work with authentic data sets which are by nature complex, more attention needs to be paid to the skills required to work with that data. Research on statistics literacy has shown that science teachers often don’t have the skills to engage with large complex data sets in a meaningful way (Chick & Pierce, 2012; Zieffler et al., 2018). Additionally, previous research within this project found that skills with Google Sheets was a barrier for teacher successful implementation of the Bioinformatics PBL Unit (Miller et al., 2021). However, in this study, development in knowledge of and comfort with Google Sheets was the primary component of subject matter knowledge development that teachers display and identified growth in.

Despite this procedural growth, their discussion on big ideas for complexity of data was mostly unconnected to the way the literature defines this component. Other than the big idea “Messy data needs to be cleaned up and summarized in order to be analyzed” the rest of the big ideas were less related to complexity of data and more related to conceptions of data in general, complex or otherwise. This was the component of data literacy that sparked the conversation on the definition of data for the teachers, so they were predisposed to be in that head space when building the big ideas for this component. However, they were right in identifying “data can take many different forms” and “data exists within students’ experiences” as big ideas for teaching and learning data literacy.
with high school students, and it is unclear where these big ideas would fit within the existing framework of critical components of data literacy.

The tendency of the teachers to consider more conceptual rather than procedural big ideas when discussing the components of data literacy suggests the perhaps more attention needs to be paid in the literature to these conceptual components and how they might be barriers to students developing data literacy as well as hooks to engage students in the process of engaging with authentic real world data in meaningful and robust ways.

8.3.7 Developing PCK for Data Literacy

Due to the exploratory nature of this study, it ended up being difficult to separate the process of surfacing the PCK that teachers already held from changes in their PCK. While this is obviously a limitation which will be discussed in more detail in the limitations section (8.4), it allowed for teachers to engage in the process of discovery and contribute to a collective PCK for data literacy that didn’t exist before the Workshop Series. As mentioned in chapter 7, since we were not working from an established set of knowledge components for teaching with and about data, measuring growth in any one area was complicated as we were surfacing the knowledge we were attempting to measure growth in as we went. Despite this, the teachers were confident that participating in the workshop shifted their thinking on how to teach data literacy in their classroom in noticeable ways. There were a few examples of shifts in how teachers were using language to express their knowledge that support teachers’ self-reported growth in PCK for data literacy.
As the components of data literacy were surface, specifically the strategies for teaching data literacy, the language teachers used to talk about their teaching practice also shifted and they began to move toward a more collective language for describing specific strategies. One example of this was the use of “storytelling” as a way to describe the process of building a scientific argument. The use of references to “storytelling” and “using data to tell a story” or determining the “story a visualization is trying to tell” increased over the course of the Workshop Series as the teachers focused on this way of talking about inferences and science argumentation as a collective way of making those concepts more familiar to both themselves and their students. So, while it seems clear that something about the teachers’ PCK shifted as a result of participating in the Workshop Series it is unclear whether that shift was an actual increase in their PCK or simply and increased awareness of and ability to communicate about PCK that they already had.

Next I will discuss some of the limitations of this study before exploring some of the implications of this research to practice and future research.

8.4 Limitations

The limitations on this study were significant. As any qualitative study which relies on case analysis, the findings are deeply bound with the context in which the study was conducted (Ravitch & Carl, 2016). As such, it is important to acknowledge two important contextual factors that played a significant role in limiting this study, those are the 2021-2022 school year during the Covid pandemic, and the Bioinformatics PBL Unit curriculum itself.
8.4.1 Challenges with the context of the 2021-2022 school year

The 2021-2022 school year fell during the second year of the Covid-19 pandemic. The School District of Philadelphia had been remote for high school students for the entirety of the 2020-2021 school year as well as the final quarter of the 2019-2020 school year, so the majority of students that the teachers in this study were working with had not physically been in classrooms between March 2020 and September 2021. As has been studied and will likely continue to be studied for years to come the learning loss caused by the Covid pandemic was huge (Donnelly & Patrinos, 2021; Engzell et al., 2021). The learning loss incurred was not only academic but also social and emotional (Loades et al., 2020). The teachers in this study commented on the anecdotal evidence of this learning loss throughout the intervention, often casually as if it was just another fact of teaching that everyone was aware of as exampled by a comment Will made during session 6, “Coming off of Covid and virtual teaching and all that the students are not, as you all know, the easiest to work with in the classrooms.” Mary also comment specifically on students’ struggle with motivation saying, “The motivation, I mean the pandemic, in this virtual universe and this trend where everything's a transition, has really shifted people's ability to be motivated.” The social and emotional learning loss was rolled into the other challenges and accepted by the teachers the same way they excepted the low math and literacy skills of their students.

As an observer in their classrooms, I was acutely aware of the shift in student attention span and desire to focus or engage. All four teachers mostly gave up on convincing students to put down their mobile devices because often requests to do so ranged from incredibly short lived compliance to open defiance. When I asked teachers
off record about this tendency they all said it was noticeably worse since before the pandemic and remote learning. As we are still immersed in the Covid pandemic it is hard yet to fully comprehend the trauma of isolation and loneliness that many of these students likely faced during the year and a half of remote learning and the degree to which their phones likely felt like the only lifeline they had. But however understandable the trauma response of detachment from the outside world and connection primarily through a mobile device is, the fact remains that this generation of students, even more so than previous ones, have an addiction to their phone which inhibits established methods of teaching and engaging. It is unclear still to what extent future classes of students will heal from the collective trauma and be able to engage more fully in in-person learning again or whether teaching styles and strategies will need to shift to accommodate for the need to connect with students in a different way.

Compounded on the learning loss of a year and half of remote teaching the contexts of teaching in an urban school district were ever present. School A had a shooting occur a block from the school that put them into lockdown for weeks, unable to take their students outside and causing a spike in attendance issues. School B had a rash of students pulling fire alarms that, as they were on an alternating day block schedule, meant that Manisha didn’t see her students for a normal class period for almost two weeks in a row. All three schools were plagued by attendance volatility, though Will less so than the other three teachers. Additionally, all three schools went back to virtual learning for over two weeks after winter break without warning. Though Will had wrapped up implementation by then, the other three had been in the middle of data collection with their students. The forced extended break from the unit as a result of not
having access to the sensors derailed all three of them as it was difficult to get the students to reengage with the project upon finally retuning to in-person school.

These changes to the schedules and constant protocol shifts due to the mutating response to the pandemic put a lot of additional pressure on teachers. Hallie and Will both had young children in schools within the School District of Philadelphia and had to call out from their own teaching job multiple times to care for children who had been sent home into remote learning because of Covid exposure, or even on one occasion due to so many teachers calling out sick that the school had to close for the day because there were not enough adults to legally keep the school open. Manisha was a special education case worker on top of her teaching responsibilities and spent most of the year feeling overwhelmed and unsupported. At one point she complained to the cohort during one of the sessions,

I’m just like so bummed out because every time it's all last minute change and suddenly you get dumped with a whole lot. Every time there's a shift, we have to send these surveys out to parents and students and then there's something else. And it's your whole caseload and it's not like we've got manageable caseload they just loading on it's just, we don't know, it's like I'm so numb right now it's like… yeah.

That sentiment of feeling numb and overwhelmed expressed by Manisha in that quote was pervasive among the teachers for the duration of the intervention. It often made it difficult for reflection on PCK to break through as the teachers were in survival mode nearly all the time.
8.4.2 Challenges with the Bioinformatics PBL Unit

Additionally, the Bioinformatics PBL Unit was a source of challenges for the teachers. The Bioinformatics PBL Unit had been designed with an ideal population in mind and had been piloted with an ideal population of students as some of the highest achieving schools in the District of Philadelphia. During the second year of the study the Covid pandemic caused most teachers to either forgo implementing completely or make major changes and pivots to accommodate for online learning. As a result, this was the first time the curriculum had been tested in-person with non-ideal populations. This resulted in large numbers of complaints and feedback from teachers about the shortcomings of the curriculum and supporting resources for implementing with non-ideal populations, and though teachers had been strongly encouraged to modify the curriculum to meet the needs of their own classroom, the “survival mode” mentioned in the previous section meant that the teachers did not have time or energy to devote to making massive modifications to the material, though they all made small modifications throughout the lessons.

One primary suggestion for the curriculum was to make it more hands on and tangible for the students. Though the data collection was hands-on and kinesthetic, most of the rest of the curriculum involved computer work and classroom discussions. Hallie suggested during her interview that she would like to see more hands-on application in the curriculum saying,

Well, I think that what I would've liked more of and feel like was lacking in the original set of materials is like, and then how would we apply that? How would we stop with the stand and deliver lecture and discussion and do
something? And then what could that be? Because the first time where you stop and do an active activity is lesson four and that's way late.

Sometimes curriculum developers, myself included think of high school students as not needing as much physical activity and kinesthetic learning, but they do, especially sometime when you are working with a high percentage of students with IEPs. Will supported Hallie’s suggestion during his interview, providing the feedback that all of the background lessons during which the students were not actively engaged with the sensors or Google Sheets were, “very dry and boring and not something that excites me to learn about or to teacher about” and followed that up with the suggestion, “as you're going through the background information, incorporate more hands-on activities so that the students grasp it.”

Another major piece of feedback that the teachers provided was on the length of the Bioinformatics PBL Unit. Will spent over three months implementing it and while he had the space for that in his class as an elective, Hallie and Manisha, as biology teachers, did not. Manisha spoke to this in her interview saying, “Streamline it. Streamline it a lot, because it's a turnoff for teachers, and make it very nugget point concept based. There was so much put in front that was a total shutdown. It's a total overwhelm. So, keep it shorter, snappier, more succinct, kept to the big ideas.” Even Mary, who also had total control over her curriculum and learning goals provided feedback that the unit was too long saying in her interview, “I mean, there's just too much there easily could be three separate 10-lesson units from this whole thing.”

Manisha was also concerned with the level of the vocabulary used throughout the unit. She reflected during her interview, “It’s also very verbiage dense. You can’t have the
verbiage dense pieces because they shut down like crazy.” This was actually something that I had observed in her (and the other teachers’) classrooms. All of their students had limited vocabulary and reading and writing skills compared to the ideal student population the curriculum had been designed for. The pre and post surveys were a big struggle for the students not because of lack of interest or motivation, but because it took them so long to read the questions, and then they needed support with comprehension, and then they struggled to put their thinking on the question into writing.

Finally, as indicated by many of the comments supplied throughout chapter 5 in relationship to relevancy to students’ lives, some of the teachers found the premise of the PBL unit boring and unengaging for students. When asked in the post interview about the problem on which the unit was based (Figure 6), Will responded,

One issue, I would say, is asthma is not as... Everybody knows what asthma is but not the precursors to it so I'm trying to tie that all in with just the carbon dioxide levels and the PM 2.5 is a little bit difficult. It's a stretch for some of them to understand how this data correlates to asthma rates. I think that's where the connection was lost a little bit, for making the connection to specific action.

As noted earlier, data literacy is not inherent engaging for students, so if the connection between the air quality data and asthma rates were lost to them and they weren’t able to understand the relevance of the data to their lives, they lost interest and motivation for engaging in the whole project. This is a particularly notable limitation because the asthma problem was chosen specifically for its perceived interest to students in line with the framing component of culturally relevant pedagogy. The disconnect between the
relevance of asthmas to students’ lives and their disinterest in collecting and analyzing air quality data is important to explore further in the larger bioinformatics project.

So, the context of the students and schools and school year combined with the limitations of the curriculum not being designed for that context made implementation very difficult for the teachers. Will was the only one of the four teachers who completed implementation of the unit by the time the Workshop Series ended. Manisha and Mary had finally completed implementation by March of 2022 and Hallie was still wrapping up in April. Additionally, the fidelity of the implementation after the unexpected revert to virtual learning for a large portion of January was poor as the teachers were in maximum survival mode and just trying to get through each day.

8.3.3 Developing PCK for Data Literacy vs Understanding PCK for Data Literacy

Additionally to those two contextual limitations to the implementation, the exploratory nature of this study created limitations as it was designed to engage teachers in surfacing the existing knowledge they had and through that process analyze how their knowledge changed. As the study relied on the surfacing of teachers’ existing PCK, the small sample size limited the scope of the PCK surfaced and the components that could be identified to a narrow context and knowledge base. All four teachers in the study taught in the same school district, so, though they came from different knowledge backgrounds and their respective schools represented a range of contexts within that district, they were limited by their experiences. A study with a larger group of teachers, or with teachers from more diverse teaching contexts would potentially add to the PCK
that could be surfaced and therefore expand the components of PCK which could be identified and named.

As PCK for data literacy and teacher knowledge for data literacy have not been widely studied, there is a lack of existing frameworks or measurement tools against which to measure teachers’ knowledge and thus determine growth. As such, we were building the boat as we sailed it and it was difficult to determine whether as mentioned earlier, the shift in the language teachers used to talk about PCK was an actual increase in their PCK or simply an increased awareness of and ability to communicate about PCK they already had before the Workshop but were maybe unaware of. While this is a limitation for this particular study, the surfacing of components of PCK for data literacy opens the door to future research that will better be able to measure change in teachers’ PCK for data literacy as there is now a roadmap for doing so.

In addition to lacking an existing framework or measurement tools for PCK for data literacy, there is also still disagreement in the field about what the priorities and goals of data literacy education should be at the K-12 level. As Kjelvik and Schultheis (2019) discussed, it is difficult to determine best practices for teaching if there is not yet agreement on what the big ideas should be. Previous studies conducted on PCK for subjects that are outside the typical classroom focus have also encountered this limitation. For example, Vossen and colleagues (2020) found that their participants struggled to complete a CoRe for the topic of Design and Technology because they “had difficulty choos[ing] and stick[ing] to one particular big idea, as they saw all big ideas as connected to each other” (pg. 314). This is an accurate description of one of the challenges faced by the teachers in this current study. Though the teachers were provided
with a framework for data literacy in the form of the six components of subject matter knowledge for data literacy and were given some suggested big ideas within the descriptions of those components, they were provided a lot of autonomy to determine the big ideas they wanted to put into the CoRe framework and these conversations about big ideas took up a large percentage of the time spent in the workshop sessions. Similar to Vossen and colleagues’ findings about participants having difficulty delineating big ideas, as can be seen in Table 7, the teachers in this study also produced overlapping and non-distinct big ideas. Determining the big ideas of data literacy was not the focus of this study but lacking a framework for the big ideas of the subject matter on which PCK relied was a limitation of the study.

8.3.4 Making ‘Data Literacy as a Form of STEM Integration’ Explicit

The framework for this study placed a lot of emphasis on STEM integration and the connection between data literacy and STEM integration. While the way that teachers thought and talked about data literacy showed that they were considering how math and technology affected teaching data literacy, the connect to data literacy as a form of STEM integration was not made. This likely arose as a result of a flaw in the Workshop Series design which placed a lot of focus and emphasis on data literacy without making the connection to STEM integration clear.

8.5 Implications for Practice

One very obvious implication of this research that is of no surprise to anyone in the field of teacher education is an add-on to the widely established fact that teachers
benefit from extended support and PD ((Darling-Hammond et al., 2017; Desimone, 2009). However, beyond that, the content and format of the support matters. The hypotheses of this study were that there exists PCK that is unique to teaching data literacy in the context of a high school science classroom and in order to effectively teach data literacy teachers need support not only in developing their subject matter knowledge for data literacy but also in developing that PCK. This study has provided substantial support for the first hypothesis and initial support for the second, though a lot more research is needed to be able to make any sort of causal claim about teachers’ PCK on classroom development of data literacy in students.

While this study was unable to determine whether participation in the Extension Workshop Series increased teachers’ PCK there is ample support that it encouraged them to reflect on their practice in new ways that they found helpful and supportive of their growth as teachers. Their overwhelming positive reflections on the process of the Workshop suggests that adding more time for community building and discussion of teaching practice into the Bioinformatics PD would benefit teachers in future cohorts of the project. As discussed in Chapter 6, there were a number of specific strategies that were used that benefitted the teachers in this study. The video reflections were mentioned by all the teachers as being beneficial to their practice and their confidence. Video reflections allow teachers to make their practice visible and, in the case of a new form of PCK, develop a language to describe that knowledge. The CoRes, while beneficial to the participants as a framework to guide their discussions, were less concretely supportive of growth than they might be for content with more well-established big ideas. However, as the big ideas of data literacy are developed into a stronger canon, the CoRe framework
will be a useful tool for building PCK off those big ideas. Additionally, as Loughran and colleagues have written (Bertram & Loughran, 2012; Loughran et al., 2012; Nilsson & Loughran, 2012), the CoRe template, though simple, needs to be practiced and scaffolded to provide the most benefit to teachers.

For future iterations of the Bioinformatics PD, and for other future PD projects which seek to support teachers in implementing STEM integration, whether through a data literacy focus, or not, I suggest the use of both video reflections and the CoRe template to support teachers’ development of their PCK.

8.6 Implications for Future Research

A major implication of this study on future research is the need for further research on both PCK for data literacy and on solidifying learning goals for data literacy in K-12 classrooms, specifically classrooms which seek to use data literacy to build STEM integration. In their study published a few years ago, Kjelvik and Schultheis (2019) called for the field to determine “authentic data best practices” by which they meant best practices for teaching with and about authentic, messy data. They then acknowledged that additional research is needed to solidify desired learning outcomes for the use of complex datasets in classrooms. While some progress has been made since that paper was published, for example the special issue of the Journal of the Learning Sciences focused on data science (Wilkerson & Polman, 2020). However, the framework presented by Rubin in that issue was theoretical, and the empirical work conducted on data literacy in K-12 classrooms has not coalesced around common goals. More research
is needed to provide a stronger framework for the learning outcomes of data literacy (Kjelvik & Schultheis, 2019; Wolff et al., 2016).

As part of that research, more attention should be paid to what teachers know and already do. Teachers across STEM subjects engage with data in different ways and grounded in different contexts, which is why data literacy is an ideal conduit for developing stronger STEM integration. Research on data literacy in K-12 has been primarily focused on what students know and how they understand and interact with data, and more of that research is still needed, but also needed is to bring teacher knowledge into the conversation. The bulk of research on what teachers know about working with data has been conducted within the lens of data-driven assessment with teachers learning to use data but not necessarily to teach it as content to their students (Mandinach & Gummer, 2016; Wolff et al., 2019). This feels like a giant hole in the research on data literacy in K-12 education. As this study showed, science teachers already have a wealth of knowledge about how to teach data literacy. Additional work needs to be conducted with expert teachers across the science disciplines as well as in math, technology, and engineering to bring that existing knowledge into the research on best practices for teaching data literacy.

Once best practices for teaching data literacy have been established, methods for PCK development should be applied to support teachers, and especially pre-service science and math teachers in growing their ability to teach data literacy with authentic, complex data by applying those best practices. However, as Frank and colleagues (2016) concluded we also need to be careful not to constrain our definition of data literacy and the best practices for teaching it to the point where it ceases to be universally useful, but
instead perhaps begin to expand our understanding of and research on data literacy to ask data literacy for what and for who. This study offers an extremely early, exploratory research into data literacy as a form of STEM integration for high school science teachers and students. More research is needed to understand what this form of data literacy is, what it’s goals are, and what PCK teachers need to successfully implement it in the classroom. A good place to start for that research is to continue to determine, at a wider and larger scale, what teachers already know about teaching data literacy in this context and begin to codify it.

8.7 Conclusion

In conclusion, data literacy is becoming increasingly vital for pursuing careers in STEM fields as well as for simply navigating daily life in a world awash in data (Dorsey & Finzer, 2017; Erikson, 2020; Gebre, 2018; Gould et al., 2016; Kjelvik & Schultheis, 2019; Wilkerson & Polman, 2020; Wolff et al., 2016). Data literacy, which includes the ability to work with and interpret data from authentic, complex contexts, is inherently a form of STEM integration (Kjelvik & Schultheis, 2019) as it requires knowledge of mathematics, technologies, and the science and engineering contexts in which it is grounded. As the STEM disciplines become more integrated, there is a need to align teaching of these topics in K-12 education with the way they are experienced in the real world (Nadelson & Seifert, 2017). Data literacy is one of the ways that this can begin to be accomplished (Aydin-Gunbatar, et al., 2020; Kelley & Knowles, 2016; Wise, 2020). However, in order to bring authentic data into classrooms, more research is needed to understand how students not only learn the skills of data literacy, but how they
understand data as a concept (Gebre, 2018). Many experienced teachers already hold some knowledge about not only how students conceive of data as a concept, but also how to teach it effectively, however, that knowledge has not been explicitly studied or codified and so it is inaccessible to less experienced or novice teachers.

In order to affectively teach data literacy in science classrooms, teachers need access to PCK for data literacy, which may lie outside their previous training and experience. The process for making that knowledge accessible could include working with expert teachers and researchers in partnership to surface PCK for data literacy and build the knowledge of the field in this area and develop a framework. It could then include using that framework to support novice teachers in growing their PCK for data literacy in order to address the need for increased data literacy teaching in K-12 classrooms. While this current study has attempted a preliminary exploration of the concept of PCK for data literacy with a strong group of expert teachers, there is still a long way to go to expand access to the vast world of data that exists all around us and ensure that future generations have the knowledge and tools to not only survive a data-rich world but to thrive and use that knowledge to enact change.
APPENDICES

Appendix A - Content Representation (CoRe) Template
From Loughran et al., 2012

**Topic:**

<table>
<thead>
<tr>
<th>Big Ideas</th>
<th>A:</th>
<th>B:</th>
<th>C:</th>
</tr>
</thead>
<tbody>
<tr>
<td>What you intend the <strong>students</strong> to learn from this idea</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Why it is important for students to know this.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>What else you know about this idea (that you do not intend students to know yet).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Difficulties/limitations connected with teaching this idea.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Knowledge about students’ thinking which influences your teaching of this idea.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other factors that influence your teaching of this idea.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Teaching procedures (and particular reasons for using these to engage with this idea).</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific ways of ascertaining students’ understanding or confusion around this idea (including likely range of responses)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B – Video Reflection Framework for PCK for Data Literacy

Data Literacy Content Knowledge
- What big ideas of data literacy are being focused on? (add as many lines as needed)

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Big Idea</th>
<th>Thoughts/comments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Engaging Students’ Understanding of Data
- What strategies are being used to engage students’ understanding of data?
- What aspects of teachers’ knowledge of student understanding are displayed?
  - How is the teacher displaying this knowledge?
- What aspects of students’ understanding of data are surfaced?

(Add as many lines as needed)

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Strategy being used</th>
<th>Display of teacher knowledge</th>
<th>Aspect of student understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- Were there any missed opportunities to engage with student understanding?

Instructional strategies for teaching with and about data
- What strategy is being used? (e.g., visual representation, model, vocabulary)
- How is the strategy matched to the learning goal?
- What evidence is there that the strategy used was helpful or confusing to students?

(Add as many lines as needed)

<table>
<thead>
<tr>
<th>Timestamp</th>
<th>Strategy</th>
<th>Connection to Learning goal</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Other Thoughts/Comments/Questions/Feedback
Appendix C - Surveys

Teacher Content Knowledge Survey (Data Literacy Questions)

1. In your view, what is data literacy?
2. In your view, is data literacy important to teach in a science class, why or why not?
3. In your view, what are the most important skills students need to have in order to work with data?
4. Do you teach data literacy in your classroom? (Y/N) (Pre-survey only)
   a. If yes chosen: How do you teach data literacy?
   b. If no chosen: Why have you not taught data literacy?

Teacher Instruction Survey (Data Literacy & STEM Integration Questions)

1. This section asks about classroom practices related to data literacy. Please select one answer per statement. Likert scale response options: Strongly disagree, Disagree, Neutral, Agree, Strongly agree. For post-survey, “I teach” is replaced with “I will teach”.
   a. I teach my students how to make and record observations
   b. I teach my students how to gather and use data to solve problems.
   c. I teach my students to represent data in graphs or charts.
   d. I teach my students to recognize and understand patterns in data.
   e. I teach my students to understand sampling variability.
   f. I teach my students to understand that all data comes with a level of uncertainty in how it represents the phenomenon.
   g. I teach my students to recognize relationships between variables.
   h. I teach my students how to evaluate the legitimacy, accuracy, and purpose of data sources.
   i. I teach my students that the context in which the data was collected is important.
   j. I teach my students to build scientific arguments using evidence and reasoning.
   k. I teach my students to be critical of data, for example that they represent only one particular snapshot of a particular moment.
   l. I use technology tools in my curriculum to help students problem-solve issues in their everyday lives.

2. This section asks about important aspects of STEM integration. Please select one answer per statement. Likert scale response options: Strongly disagree, Disagree, Neutral, Agree, Strongly agree.
a. I consider connections to other disciplines when designing and teaching lessons
b. I make connections between math, technology, and science explicit for my students
c. I use data as a way to connect math and technology to science
d. I feel comfortable teaching components of math and technology in my science classroom
Appendix D - Interviews

Classroom Observation Debrief (Adapted from Henze & van Driel, 2015 and Aydin-Gunbatar et al., 2020)

1. What were your main objectives for this lesson?
2. What was your role as a teacher during this lesson?
3. Did your students need any specific previous knowledge going into this lesson?
4. What was successful for your students? What is the evidence to support this?
5. What difficulties did you see? Why do you think they had these difficulties? What is the evidence to support this?
6. During this lesson what were the moments of STEM integration? Were there other moments where integration could have been made explicit?

End of Workshop Interview

The first set of questions are to see how your understanding of data literacy has evolved through this process.

1. How would you describe what data literacy is?
2. In your view, is data literacy important to teach in a science class, why or why not?
3. What would you say are the most important skills students need to have in order to work with data?
4. How do you plan to teach data literacy in the future in your classroom?

This next set of questions are to probe your pedagogical content knowledge for data literacy.

5. How would you describe pedagogical content knowledge?
6. How would you say that understanding has evolved over the course of this workshop series?
7. In your post workshop reflection comments, you said that you felt that _____ was your strongest… (tailor this to participants response to encourage them to expand on it)
8. Similarly, in your reflection comments you said that you felt that _____ was your weakest point… (tailor this to participants response to encourage them to expand on it)

The next set of questions are about the workshop series itself.

9. What were some of your biggest takeaways from the workshops?
10. Did you find that participating in the workshop changed your perspective on teaching data literacy? How? Probe for: confidence, PCK
11. What parts of the workshop series did you find most supportive of your growth as a teacher?
12. There were two primary structured components of the workshop: CoRes and Video reflections. I’m going to ask you about each one separately.

**CoRes**

a. What did you think of the Content Representations?
b. Did you find them helpful? Why or why not?
c. What was the most helpful format for completing them? (individual vs group)
d. How could they have been more effective?

**Video Reflections**

e. What did you think of doing the video reflections?
f. Did you find them helpful? Why or why not?
g. What was the most helpful format for completing them? (individual vs group)
h. How could they have been more effective?

13. What suggestions do you have for how the workshop could be changed/improved for the next set of teachers?
Appendix E - Classroom Observation Protocol Template

Date:  
Observer Name:  
School:  
Teacher:  
Grade:  
Class title (e.g., Honor Biology):  
Class time:  
Number of students:  
Special conditions (e.g., SPED, IEP, Honors, other important demographic information about the school)

Content/Topic (What is the focus of the class?)

Observation Strategies:

1. We are interested in understanding how teachers and students work with the K12Bioinformatics activities and tools as well as how teachers translate what they learned about the project in the professional development course.
2. We also want to understand what the barriers are to implementing this in a classroom. What are the modifications to curriculum plans and why?
   a. What are the supports teachers needed to deliver Bioinformatics units? What type of adaptation did they make?
3. How teachers are addressing critical pedagogy and students responding
4. How did you see critical pedagogy support your work?
5. Please note that there is a difference between observation (facts) and inference (interpretations of the facts). We are primarily interested in observation, e.g., what happened. But inferences are important too.
<table>
<thead>
<tr>
<th>Observations (What happened?)</th>
<th>Inferences/Observer Comments, Questions, &amp; Impressions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example: A group of three students (use their names) discussed their plan for outdoor data collection and their proposed route for about 5 minutes. One student said several times, “I don’t think we should go that way.”</td>
<td>Example: Students were on task; one student (Wendy) didn’t want to collect data by the cafeteria because it was too hot.</td>
</tr>
<tr>
<td>Additional things to include:</td>
<td>Things to include:</td>
</tr>
<tr>
<td>- direct quotes from teachers and students (used to substantiate claims in reports, for professional development workshops, and journal articles)</td>
<td>- interpretations &amp; impressions of what is happening</td>
</tr>
<tr>
<td>- demographic characteristics of the speaker, for example, African American female who was quiet throughout most of the lesson said...</td>
<td>- observer’s emotional response (e.g., “I was frustrated that the teacher did not encourage discussion amongst the students despite this being an integral part of the curriculum and project.”)</td>
</tr>
<tr>
<td>- amount of class time spent on the activity (or the time the observation was recorded)</td>
<td>- questions that observer may have for the teacher/K12Bioinformatics team</td>
</tr>
<tr>
<td></td>
<td>- comments or discussion points that might be important but were not part of the formal lesson.</td>
</tr>
</tbody>
</table>

**Brief Teacher Check-In (both formal and informal discussion notes)**

Check in with teacher about the topic and goal of the class before he/she teaches. Also collect any handouts, worksheets, or other resources that may contribute to our understanding of the project’s goals.

At the end of the class, ask the teacher about anything you were unsure of or anything you found particularly interesting. For example, a teacher may have spent a lot of time on explaining a certain concept that didn’t appear to be related to the lesson. If you ask him/her later why, you may find out that he/she was trying to make connections to a concept taught earlier in the year.

Also ask the questions, “How did you prepare your students before this class?” and “How will you debrief it?” (These are important questions in order to understand the context).

Look for continuity between the last day and today.
Write down any contextual information that will help to explain what happened in the classroom.

**Major Inferences**

What was your overall impression of the class? Include impressions on student engagement, teacher comfort, interest and ownership, and major difficulties or challenges you think may have influenced the implementation, e.g., lack of working computers, classroom management issues. Also write down modifications or adaptations the teacher may have made in order to meet the needs of the student population.

Using your observations above, briefly comment on how the lesson was executed, and student and teacher actions, according to each of the 4 lenses:

**Lens 1: Content (Bioinformatics)**
- Is there any evidence that student’s understanding of bioinformatics is enhanced through project activities?
- Are students understanding the big data supports the study of bioinformatics?
- How are students understanding the relationship between asthma and air pollution? Do they understand that the environment interacts with the human body to impact disease?
- How are teachers scaffolding student understanding? Are they making explicit connections? How? What are the strategies?

**Lens 2: Data Literacy**
- How are students interacting with data?
- Are they cognitively engaged?
- Evidence of how our custom websites enable learning.
- How is the teacher working with students to interpret, analyze, and visualize data? (e.g., walking around asking good questions to small groups)
- Refer to the 6 data literacy issues that we are trying to address: data in context; data variability; aggregate views of data; data visualizations; making inferences with data as evidence; data complexity. Is there evidence that students are learning any of these issues?

**Lens 3: Computational Literacy**
- How are students interacting with technology?
- Are they cognitively engaged?
- Evidence of how working with mobile devices and sensors enable learning.
- How is the teacher working with students to collect real-world data?
- Refer to the 3 computational literacy components: material; social; cognitive and comment on any of these aspects.
**Lens 4: Scientific Agency/Action**

- How are students interacting with the final project?
- Are they cognitively engaged and excited to take action? Did students actually take action?
- Evidence of how working on the PBL unit enables learning.
- How is the teacher working with students to solve problems and take action in their community?

**Additional Questions for this research**

**Lens 5: Pedagogical Content Knowledge (PCK)**

- Is there any evidence of teachers’ beliefs about the nature of data and the purpose of developing data literacy?
- How did the teacher explicitly address student misconceptions?
- What activities were used to explicitly address and scaffold student difficulties?
- What examples and representations did the teacher use? How were they received by the students?

**Lens 6: STEM Integration**

- What are examples of STEM integration in the lesson?
- How were these moments addressed or highlighted by the teacher?
Appendix F – Final Coding Manual

<table>
<thead>
<tr>
<th>Codes</th>
<th>Description</th>
<th>Example(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conceptions of Teaching and Learning With and About Data</td>
<td>Demonstration of an idea about the role of data literacy in the classroom. These ideas are more abstract than specific instructional strategies but represent dispositions to teaching data literacy. Two subcodes surfaced.</td>
<td>I think there’s a relevance. Like, why does this matter to students because really if we can't make it matter, what is it, why would they care? So, I guess I wouldn't have standalone data literacy. I would want that to be a term that they heard and that they understood and that students thought of themselves as being data literate. But this is a group of students who if you can't make it relevant, they're just not... So, if you can't find a hook. If it's not somewhat an authentic hook, then what’s the point?</td>
</tr>
<tr>
<td>Relevance</td>
<td>Demonstration of an orientation toward relevance as being fundamental to the teaching of data literacy. Though some statements about specific examples of making data relevant were co-coded with this, this code represents the disposition toward relevance as fundamental.</td>
<td></td>
</tr>
<tr>
<td>Data Tools are Powerful</td>
<td>This code is a combination of first round codes about “Tools are a Driver” and “Data Literacy is Powerful” because most of the instances in “Data</td>
<td>I mean it is a really powerful tool, so I feel like this idea of how versatile some tools are and that you need to be shown how. The visualizing of data that Google sheets can do is the twist in this particular unit for me. Google sheets, is a very, is a through line.</td>
</tr>
<tr>
<td>Conceptions About the Nature of Data</td>
<td>Demonstration of a belief or conception about the characteristics of data or what counts as data. Two subcodes emerged.</td>
<td></td>
</tr>
<tr>
<td>Defining Data</td>
<td>Demonstrates thinking about the characteristics of data or what counts as data</td>
<td>Well, information that's not verifiable wouldn't be data, it would be just, I mean again so there's so much that people believe that isn't actually data, so data is possibly reproducible or... Data can be in the form of facts, numbers, characters, symbols, or also anything that can be processed by a computer. Once the data has been interpreted, it can be considered to be information.</td>
</tr>
</tbody>
</table>
Data is Manipulative | Specific demonstration of a conception that data is or can be used to manipulate | I think that sometimes people use data to tell a story that's a marketing story, not a science story and sometimes people are sometimes people are have malicious intent so, I mean data can be manipulated, so how good is that and how different is that from belief it's just to reinforce the belief so it's really tricky.

Goals of Teaching Data Literacy | Explicitly stated goals or takeaways for what they believe is important for students to know or learn about data and the skills needed to interpret and work with it. Three subcodes emerged. | 

Learning to Question/Avoiding Manipulation | Statements about questioning data either as a way to simply learn more about it (as in the second example) or specifically to avoid manipulation (as in the first example) | Well, it's important for students to know that data can be manipulated. Data is evaluated by humans, and data sometimes is misinterpreted, and that there's always an opportunity for them to look at the raw data and reinterpret the data in a different way. You have to question where it came from. The process of collection is critical. The process of screening is critical and that impacts it. You have to question what happened before to have that data. So now whatever is served to you in terms of data is questioned in a different sort of way.

Communicating/Storytelling | Statements about communication as a goal of learning to work with data. Storytelling is one way that many teachers talked about the ability to communicate about data | I want them to like the, how do you talk about like, how do you talk about visualizing telling a story with data and that's the part that I want them to learn.

Interdisciplinarity/STEM Connections | Demonstration of connections to other STEM disciplines or explicit discussion of “interdisciplinarity” | I mean there's, I feel like the interdisciplinary opportunities for all of this stuff and we really, for anything, is so, they're so powerful they're huge.

Knowledge of Student Understanding | Demonstration of knowledge about students and how they think about and interact with data. | 

Student Motivation/Interest | Knowledge about what will cause students to lose interest and about what will engage students or motivate them to become engaged. | They wonder why are we learning this: I don't care about air quality, I don't care about… half the kids don't care about whatever it is we're doing.
<table>
<thead>
<tr>
<th>Fear of Numbers/Data</th>
<th>Demonstration of knowledge that some students are hesitant about engaging with numbers or math.</th>
<th>Eventually, when they start doing things that are fun like walking around school getting free passes to school for two whole days with sensors and blessing from the principal and it was like oh wow, this is a lot of fun. We've done such a bad job as a human race of helping people feel like they have power over their environment. I feel like numbers, like the fact that anybody is afraid of numbers is such a shame. They actually do understand the math its arithmetic that they get hung up on.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meaning/Significance of Data</td>
<td>Demonstration of knowledge about the different ways that students conceive of and define data.</td>
<td>They don't realize how much they already have been using data to make decisions they think of data as this other thing that scientists use, or we talk about in school If you think it’s just numbers and you’re not that wild about numbers, then it's you know once again going to be alienating.</td>
</tr>
<tr>
<td>Assumptions</td>
<td>Knowledge about how students engage their assumptions and opinions when working with data.</td>
<td>The assumptions that we talked about earlier, you know that they are they're driving a lot of conclusions We really need to get our kids away from this idea of opinions You know, especially in this world with too much information that is the thing it's like if you're, if our lens is clearly on just opinions we're not going to dig underneath it to see the credibility of it</td>
</tr>
<tr>
<td>Questioning Data</td>
<td>Demonstration of knowledge about how students do or don’t question data</td>
<td>It's like to ask those questions, just think about it, because once they think about it it might steer them in a different direction to where they thought before everything is handed to them, but they aren’t questioning</td>
</tr>
<tr>
<td>Need for Right Answers</td>
<td>Demonstration of knowledge about how students engage with “correct” answers in science class</td>
<td>We are always talking about results, and this idea of an end. There is no end; there's constant innovation and</td>
</tr>
</tbody>
</table>
and how this engagement relates to their understanding of data interaction and things like that. So, first of all it's like moving our kids through that “is this right?”

One thing my students always want is, is it right or wrong, and that’s not the answer that you're looking for. There is no right or wrong, and they have to know that.

<table>
<thead>
<tr>
<th>Instructional Strategies</th>
<th>Demonstration of thinking about instructional strategies for addressing students’ understanding of data and engaging them with the process and context of building data literacy. Both specific and general strategies are coded in this section.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interrogating Data</td>
<td>Example of how to interrogate data either as a suggestion for future use or as an identification of current or past use.</td>
</tr>
<tr>
<td></td>
<td>What I enjoyed is [Will’s] questioning. He kept on squeezing the kids to think through their responses and they dug into it. That was great to see, and he kept on kind of like getting them to think through it.</td>
</tr>
<tr>
<td></td>
<td>It's part of the probing for that interrogation is like okay what can we find from this, what can we discover about this context and then, what do you still wonder, and it's not so specific that it's even like what do you still wonder about this dataset it's about what do you still wonder.</td>
</tr>
<tr>
<td>Scaffolding/Modeling the Tools</td>
<td>Example of how to scaffold the tools through modeling either as a suggestion for future use or as an identification of current or past use.</td>
</tr>
<tr>
<td></td>
<td>Once I hit two or three students that are having the same issue, I redo it. Then, at the end of the lesson they had to share their sheet with me and then I had to go in and then, make sure that they actually use the formulas, not just typing in the numbers.</td>
</tr>
<tr>
<td></td>
<td>I think that demonstration of you know you're asking questions, he was just sort of riffing on like oh I’m looking at this, oh, I noticed this, oh, I noticed this now. What do you think? and showing that you know it's a process.</td>
</tr>
<tr>
<td>Demonstrating Mistakes</td>
<td>Example of how to scaffold the tools through modeling specifically by demonstrating mistakes either as a suggestion for</td>
</tr>
<tr>
<td></td>
<td>You know, you can always go back to that if you make a mistake and I made several mistakes in it, especially the first couple days, and</td>
</tr>
<tr>
<td>Relevance to Students</td>
<td>Example of strategies for how to make data literacy relevant to students either as a suggestion for future use or as an identification of current or past use.</td>
</tr>
<tr>
<td>-----------------------</td>
<td>---------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>So, they figured out it turned out video games and phone usage was really high and taking up a lot of time and then we got to talk about how some of that video game time overlapped with school didn't it, and like it was really interesting to sort of think about the visualization part.</td>
</tr>
<tr>
<td></td>
<td>Then in their small groups each person got the sensor for a night and they took readings on their way home and on their way back to school, the next day, so that there is some more of a personalized approach to it, and you know it's more relevant to what they're experiencing, and they really enjoyed doing that and having the sensors.</td>
</tr>
<tr>
<td>Making Data Tangible</td>
<td>Example of how to make data tangible either as a suggestion for future use or as an identification of current or past use.</td>
</tr>
<tr>
<td></td>
<td>We can do felt white felt we stable to poles in the neighborhood to see which ones turn gray faster like that would be a really easy but tangible like air quality assessment that wasn't technical but would be very visually cool.</td>
</tr>
<tr>
<td>Storytelling</td>
<td>Example of how to engage students with telling stories with data as a suggestion for future use or as an identification of current or past use.</td>
</tr>
<tr>
<td></td>
<td>I think that there's at least two lessons here and the first really just being like let's look at relatively raw data that's not.. We have to pull the story out of it and figure out the part about how a person had to decide that these were the survey questions right. Someone had to determine Where data was being collected and which kinds of things are being looked for.</td>
</tr>
<tr>
<td></td>
<td>I almost wonder, would there be a way to like assign students, like they have the same data, and they could tell, we could ask them to tell completely different stories that weren't wrong that weren't lies, using</td>
</tr>
<tr>
<td><strong>Using CER</strong></td>
<td>Demonstration of knowledge of CER as a tool for building a story out of data</td>
</tr>
<tr>
<td><strong>Teacher Learning/Development</strong></td>
<td>Demonstration of teachers’ meta knowledge of their understanding of their own PCK and Data Literacy as big ideas as well as how their perception of them may have changed over time.</td>
</tr>
<tr>
<td><strong>Defining/Identifying PCK</strong></td>
<td>Instances of teachers displaying an understanding of what PCK is and why it’s important to pay attention to and develop.</td>
</tr>
<tr>
<td><strong>Defining/Identifying Data Literacy</strong></td>
<td>Instances of teachers displaying an understanding of what Data Literacy is and why it’s important to pay attention to and develop.</td>
</tr>
<tr>
<td><strong>Ideas for Teaching/Planned Changes</strong></td>
<td>Demonstration of teachers’ changed perspective on teaching data literacy or plans to make changes to their practice.</td>
</tr>
</tbody>
</table>
Well, I think one of the big takeaways is earlier and more often, and any excuse to collect data and any excuse to manipulate it and create visualization of any kind, like starting right off the bat.

| Confidence | Demonstration of teachers’ confidence levels teaching with and about data. | A sense of comfort and then to be able to say Oh well, got this platform, but I can tweak this or that and be able to play around with it yeah I mean I learned a lot from it.
Yeah, I mean, it was very confidence-boosting. |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Workshop Reflections</strong></td>
<td>Demonstration of teachers’ reflection on the structure and value of the Workshop Series and how they engaged with the different components</td>
<td>So, I get to see what she's doing. That's really helpful and so I feel like these ideas are helpful because we'll be able to incorporate them, but I think the part that's really exciting about all of this is the exchange of what's working for students, how are we going to engage them to care about any of this.</td>
</tr>
<tr>
<td><strong>Sharing Ideas</strong></td>
<td>Demonstration of the sharing of ideas as a component of the Workshop Series that either supported or hindered learning and how.</td>
<td>The free talk. Being able to not be tasked with: answer this question, but: here's a prompt, let's discuss this out loud. And it felt like all of us took something from another previous speaker and added to it, and then it just kind of became a bigger end goal or end thing than if we worked individually on it</td>
</tr>
</tbody>
</table>
| **Free Talk** | Demonstration of specifically free talk as a component of the Workshop Series that either supported or hindered learning and how. | I’m just amazed that you guys got all of that from my video because when I’m teaching it I don't know if you feel the same way when you guys are teaching it it feels like it is a massive train wreck, and I am just trying to survive. Is that how you all feel and when you're teaching it as well?
Being able to see each other teach their videos and then talk about how to make it better for next time, how to take What you're doing and how to make it even better take ideas that |
| **Video Reflections** | Demonstration of the video reflections as a component of the Workshop Series that either supported or hindered learning and how. |  |
| CoRes | Demonstration of the CoRes as a component of the Workshop Series that either supported or hindered learning and how. | somebody else has done and improved that's what teaching is supposed to be. | But it's like emptying your brain into it, onto paper. Like you made it visible and it's like okay, this is what I'm thinking, and this is why what you're doing intuitively

So that helped me focus on, when seeing what other... I was more looking at what the other teachers were talking about, so that if there's things... Obviously, I know what I don't know but seeing what other teachers were saying about what they were focusing on and what they knew opened my eyes: okay, oh, I hadn't thought of it that way. |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Big Ideas</td>
<td>Demonstration of teachers identifying and clarifying big ideas for data literacy. They may or may not explicitly identify them as big ideas.</td>
<td>So, I feel like there's something, there should be something about a big idea about sources matter. That um, and which is sort of about the background info I guess right because you, you have to know that there's a reliable source. Also, not just what the information is, but what I'm intending to do with it, or about it, like not just like okay so that's what that looks like but also like so what? so you don't just see the what you see the so what.</td>
<td></td>
</tr>
<tr>
<td>Classroom Context</td>
<td>Describes the context of the classroom, school, and/or students</td>
<td>Some of this bumpiness is kind of throwing them off. We've got the masking mandate. As soon as the kids mask is down they get sent home, so they're really enforcing that a lot.</td>
<td></td>
</tr>
<tr>
<td>Personal Context</td>
<td>Context about how teachers are feeling, their emotions, health, or other external factors affecting their lives.</td>
<td>Something I'm looking forward to? oh gosh, just sleeping. I'm really looking forward to getting like… I feel like I haven't slept in like 10 years because I have children, but it's even more so lately. I haven't had a kitchen for six months, still coordinating it. I didn't</td>
<td></td>
</tr>
</tbody>
</table>
have a working refrigerator. The new one came today.
REFERENCES


Ayoubi, Z., El Takach, S., & Rawas, M. (2017). Improving the pedagogical content knowledge (PCK) among cycle 3 in-service chemistry teachers attending the training program at the faculty of education, Lebanese University. *Journal of


Chan, K. K. H., & Hume, A. (2019). Towards a consensus model: Literature review of how science teachers’ pedagogical content knowledge is investigated in empirical studies. In A. Hume, R. Cooper, A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers’ knowledge for teaching science* (pp. 3-75). Springer, Singapore.


Henze, I., & Van Driel, J. H. (2015). Toward a more comprehensive way to capture PCK in its complexity. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-
examining pedagogical content knowledge in science education (pp. 120–134). Routledge.


STEM road map: A framework for integrated STEM education. (pp 3-12).
Routledge.


Park, S., & Suh, J. K. (2019). The PCK map approach to capturing the complexity of enacted PCK (ePCK) and pedagogical reasoning in science teaching. In A. Hume, R. Cooper, & A. Borowski (Eds.), Repositioning pedagogical content knowledge in teachers’ knowledge for teaching science (pp. 185-197). Springer, Singapore.

Demonstration of an evaluative framework. *Journal of Research on Technology in Education, 47*(2), 105–121. DOI:10.1080/15391523.2015.999640


Ravitch, S. M. & Carl, N. M. (2016). *Qualitative research: Bridging the conceptual, theoretical, and methodological*. Sage Publishing.


