

The Nature of Analyzing and Interpreting Data in 6-12 Science Education

A Dissertation Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Amanda Victoria Garner
August 2024

Acknowledgments

When I embarked on this journey, I had no clue what I was getting myself into. Thankfully, I've had an amazing support team to help me through it all. First and foremost, I want to acknowledge and thank my Mamaw Garner and Mamaw Wilhoite. Throughout my life, they staunchly supported my education and always told me to get all I could because it's the one thing no one can take away. They believed in the power and importance of education, especially for women. Though they are not here to share this milestone with me, with tears in my eyes, I say thank you for being my biggest fans and for allowing me to see my own potential through the love you had for me.

I also want to thank my parents, Kyle and Nancy Garner, for believing in me and supporting me throughout this journey. Without their support, I wouldn't have reached this moment. Thank you, Mom and Dad, for always being there to listen, give me pep talks, and, most importantly, feed me!

To my advisor, Joshua Rosenberg: thank you for your unwavering support, mentorship, opportunities for growth and experience, and the great chats about the Smokies. You believed in my potential even when I didn't. Thank you, Josh, for your endless positive affirmations!

Leia Cain, thank you for giving me the confidence to explore research in ways I never thought possible. You've been a mentor, a writing colleague, and a constant source of encouragement to put my writing out there.

Thank you, Kristin Rearden, for your unwavering support throughout this process. Your experience working with teachers has been invaluable, and your mentorship in teaching undergraduates has greatly enriched my journey. You provided the guidance I needed to understand this path and ensured that the teachers' voices and mine remained at the forefront.

Thank you, Elisabeth Schussler, for providing an outside perspective and reminding me to remain mindful of the dual roles I play and to be conscious of which hat I am wearing at any given time.

Kristen, your friendship, support, and encouragement have been invaluable. Thank you for listening to all my stresses and celebrating all my successes.

Ronan, you have been an amazing support, friend, and conference buddy. Thank you for allowing me to be myself, for sharing the ups and downs, and for laughing with me until we cried.

Marta, your support and help with writing this dissertation have been crucial. Your mentoring has kept me sane and allowed me to see the light at the end of the tunnel.

A special thank you to my dog, Beaker, my best friend, who should be getting this degree alongside me. You sat with me through every second of coursework and dissertation writing and protected me from various mail carriers.

Alice, thank you for always being there to meow at colleagues during Zoom meetings and for sitting on anything I needed at the most inconvenient times.

Finally, I want to thank all my participants for being so open and willing to share their experiences with me. Your voices are crucial in making positive changes in science education.

Abstract

This dissertation explores the nature of analyzing and interpreting data in 6-12 science education, focusing on how it's taught and how it can be improved to meet the demands of the 21st century. It examines instructional practices, challenges, and curriculum content to provide a comprehensive understanding of data literacy in secondary science education. It is structured as an introduction chapter, three research chapters, and a conclusion chapter, each addressing different aspects of data literacy. The second chapter investigates how 6-12 science teachers define and teach data analysis and interpretation. This study uses focus groups to gather insights into teachers' approaches and influencing contextual factors. It also examines the alignment of their practices with a synthesized model of best practices for data literacy in science education. The third chapter identifies the obstacles teachers face in teaching data literacy, such as insufficient training, lack of resources, and curriculum limitations. This investigation uses the same focus group methodology as Chapter 2 but with a different focus on perceived teacher challenges and their impact on instructional practices. The fourth chapter evaluates a 7th-grade unit from an open-source science curriculum to assess how it incorporates data literacy. The methodology involves a detailed content analysis of the curriculum materials, focusing on the types of data used, the data analysis methods taught, and their alignment with modern data practices. This approach allows for a thorough examination of the curriculum's strengths and weaknesses in promoting data literacy. The dissertation concludes with a fifth chapter that synthesizes the findings across the three studies and offers recommendations for improving data literacy instruction in 6-12 science education. This dissertation highlights the need for a comprehensive and integrated approach to teaching data literacy. It offers valuable insights and

actionable recommendations for educators, curriculum developers, and policymakers to enhance science education and better equip students with data-driven skills.

Table of Contents

Chapter 1. Introduction	1
1.1 Rationale	1
1.1.1 Statement of Problem.....	2
1.1.2 Purpose of Study	3
1.1.2 Significance of Study	4
1.2 Literature Review.....	6
1.2.1 What is Data Literacy?.....	7
1.2.2 21st Century Data Terms Defined	9
1.2.3 Science and Engineering Practice 4 in the Context of 21st-Century Use of Data	13
1.3 Conceptual Framework	20
1.3.1 Context in Education.....	21
1.3.2 Context and Teaching Practice	23
1.3.3 Synthesizing Frameworks for Best Practices.....	25
1.4 Positionality	29
1.5 Introduction to the Three Papers.....	30
Chapter 2. 6-12 Science Teacher Approaches to Teaching Analyzing and Interpreting Data	33
2.1 Introduction.....	33
2.2 Literature Review.....	34
2.2.1 What is Data Literacy?.....	34
2.2.3 An Examination of NGSS Practice 4- Analyzing and Interpreting Data.....	36
2.2.4 Data Literacy and SEP 4.....	37
2.2.5 Four Frameworks for Teaching Data Literacy.....	38
2.2.6 The Role of Textbooks in Science Education.....	39
2.3 Conceptual Framework	40
2.3.1 The Role of Contextual Factors in Teacher Practice	40
2.3.2 Synthesizing Best Practices for Teaching Data Literacy	42
2.4 The Study	43
2.4.1 Research Design.....	43
2.5. Data Collection and Analysis.....	47
2.5.1 Data Analysis	48
2.6 Findings.....	50
2.6.1 RQ 1-Data Literacy Defined.....	50
2.6.2. RQ 2- Approaches Used in Teaching Analyzing and Interpreting Data	51
2.6.3 RQ 2- How the Approaches Align with Recognized Best Practices	55
2.7 Discussion and Conclusions	63
2.7.1 Connection to Framework (2012).....	64
2.7.2 Connection to the SBPM	64
2.7.3 Conclusion and Limitations	66
Chapter 3. Exploring Data Literacy Challenges in Secondary Science Education.....	68
3.1 Introduction.....	68
3.2 Literature Review.....	69
3.2.1 What is Data Literacy?.....	69
3.2.3 An Examination of NGSS Practice 4- Analyzing and Interpreting Data.....	70
3.2.4 Data Literacy and SEP 4.....	72

3.2.5 Four Frameworks for Teaching Data Literacy.....	73
3.3 Conceptual Framework.....	74
3.3.1 The Role of Contextual Factors in Teacher Practice	74
3.4 The Study	76
3.4.1 Research Design.....	76
3.4.2 Participants.....	78
3.5 Data Collection and Analysis.....	79
3.5.1 Data Analysis	79
3.6 Findings.....	80
3.6.1 Insufficient Training	81
3.6.2 Curriculum	83
3.6.3 Time	85
3.6.4 Resource Availability.....	87
3.6.5 Student Prior Knowledge and Engagement	88
3.6.6 Summary of Findings.....	90
3.7 Discussion and Conclusions	91
3.7.1 Limitations	94
Chapter 4. The Nature of Data Literacy in a Middle Grades Science Curriculum Unit.....	96
4.1 Introduction.....	96
4.2 Literature Review.....	98
4.2.1 Discussing Data Literacy in Science and Engineering Practice 4	98
4.2.2 Current Models for Teaching Data Literacy	100
4.2.3 The Role and Impact of Science Textbooks	104
4.3 Conceptual Framework.....	105
4.3.1 Synthesizing Best Practices for Teaching Data Literacy in 6-12 Science	105
4.3.2 OpenSciEd Curriculum.....	106
4.4 Purpose and Research Questions	107
4.5 The Study	108
4.5.1 Research Design.....	108
4.5.2 Data Source.....	109
4.5.3 Coding Frame.....	109
4.6 Data Analysis and Inter-Rater Agreement.....	109
4.7 Findings.....	112
4.7.1 RQ1- The Nature of Data Used by Students in the 7 th -grade Curriculum	112
4.7.2 RQ 2A-Types of Analysis.....	113
4.7.3 RQ 2B- Frequency of Student Data Practices.....	114
4.7.4 RQ 2B How does this data analysis align with the SBPM?	115
4.8 Discussion and Conclusions	116
4.8.1 Future Directions	119
4.8.2 Limitations	119
Chapter 5. Overall Findings, Conclusions, and Implications	121
5.1 Key Findings	121
5.2 Conclusions Across Three Studies.....	128
5.3 Implications Across Three Studies	131
5.4 Future Directions	133
References.....	135

Appendix.....	150
Appendix A: How Curricula Affect Student Perspectives of Science Claims	150
Appendix B: Conceptual Framework Model.....	151
Appendix C: Frameworks for Teaching Data Literacy.....	152
Appendix D: Synthesized Best Practices Model	155
Appendix E: Participant Demographics for Chapters 2 and 3.....	157
Appendix F: Modified Thematic Analysis Process	158
Appendix G: Codebooks for Chapter 2	159
Appendix H: Codebook Chapter 3.....	163
Appendix I Themes for Chapters 2 and 3	166
Appendix J: Codes and Definitions for Chapter 4 Content Analysis	168
Appendix K: Occurrence of Each Code Chapter 4.....	170
Appendix L: Figures of Findings Chapter 4	172
Vita.....	175

List of Tables

Table 1.1 K-12 Science Education Goals (National Research Council, 2012)	14
Table C.1 Levels of Science Inquiry.....	152
Table C.2 Data-Informed Learning Principles.....	153
Table C.3 Workflow Through Data Science Tutorial Design	154
Table D.1 Synthesized Model of Best Practices for Teaching Data Literacy in 6-12 Science Education	155
Table D.2 Updated SBPM for Teaching Data Literacy in 6-12 Science Education.....	156
Table E.1 Teacher Participants, Demographics, Focus Group Attendance.....	157
Table G.1 Inductive Codebook Chapter 2	159
Table G.2 Deductive Codebook Chapter 2 Alignment with Framework (2012) Goals	160
Table G.3 Deductive Codebook Chapter 2 Alignment with SBPM	161
Table H.1 Inductive Codebook Chapter 3	163
Table I.1 Themes from RQ #2	166
Table I.2 Themes Defined.....	167
Table J.1 RQ 1 Codes and Definitions for Content Analysis	168
Table J.2 RQ 2 Codes and Definitions for Content Analysis	169
Table K.1 Occurrence of Each Code for RQ 1	170
Table K.2 Occurrences of each code for RQ 2	171

List of Figures

Figure 1.1 The Three Papers	32
Figure A.1 Flow Chart Demonstrating Deterministic vs Probabilistic Thinking in Science Education	150
Figure B.1 Conceptual Framework Model	151
Figure F.1 Modified Thematic Analysis Process for RQ 2 and 3.....	158
Figure L.1 Type of Data Included in Curriculum	172
Figure L.2 Type of Visualizations Included in Curriculum.....	173
Figure L.3. How Data is Used by Students.....	174

Chapter 1. Introduction

1.1 Rationale

The terms data and data literacy lack a universally accepted 21st-century definition (Gebre, 2020, p.1081). Modern iterations generally involve critically thinking about data, working with and understanding messy and big data, and considering data ethics and storage. However, since there is no consensus on a definition, this will be explored further in the literature review. However, the literature does indicate that the conventional definition- making sense of data, data collection, analysis, and interpretation, is narrow and limited, particularly evident in many standards-aligned science curricula (Bowen & Bartley, 2020). In *A Framework for K-12 Science Education*, the authors emphasize the importance of organizing raw data into visualizations yet fall short of explicitly detailing the process of cleaning raw data and defining a full data cycle (National Research Council, 2012). Science and Engineering Practice 4 - Analyzing and Interpreting Data (SEP 4) (Lead States, 2013), a practice within the Next Generation Science Standards, focuses on basic skill competencies in working with data. However, it lacks specificity regarding data size, the full data cycle, and visualization types. While SEP 4 emphasizes representing data to reveal patterns and relationships and communicating findings effectively, the specifics of achieving this remain unclear. Current standards for teaching, analyzing and interpreting data do not fully represent how working scientists work with data (Bowen & Bartley, 2020; Pangrazios & Sefton-Green, 2019; Wolff et al., 2019). The Next Generation Science Standards encourage the utilization of all Science and Engineering Practices collectively to achieve the overarching goal of "sense-making" in science (Kang et al., 2019; NGSS Lead States, 2013; Schwarz et al., 2017).

Existing research has investigated teachers' instructional strategies to engage students with data (Banilower et al., 2018) and educators' data sources/tools (Rosenberg et al., 2022). However, both studies heavily rely on extensive participant survey responses and may miss contextual factors that influence teacher practice. Capturing these factors through only survey methodologies can pose a challenge (Banilower et al., 2018). The Banilower et al. (2018) study explored teachers' beliefs, professional development, adherence to state standards, and usage of instructional materials related to NGSS practices. However, the study fell short of providing detailed insights into the nuances of teacher engagement with these practices. While addressing whether teachers engaged with these practices, the study failed to illuminate the manner and the specific tools through which such engagement occurs. Rosenberg et al.'s (2022) study extended foundational research by focusing on the sources and tools employed by teachers in handling data, establishing connections with data handling embraced by scientists, and exploring their translation into science education. Encompassing elementary, middle, and high school teachers, this study shed light on various aspects of analyzing and interpreting data, including data types, resources, and tools.

In this dissertation, I will build upon these studies, examining these areas in more detail by exploring teacher insights within the context of smaller participant focus groups and curriculum analysis. I will highlight how data literacy is currently used in 6-12 science education, providing insights directly from practicing educators about the specifics of working with data in 6-12 science classrooms.

1.1.1 Statement of Problem

There has been very little research directly focused on Science and Engineering Practice 4 (SEP4) resulting in a lack of attention that other practices have received. What literature is out

there about SEP 4 s only mentions it briefly along with other practices in a larger volume (Schwarz et al., 2017). In contrast, modeling (Campbell et al., 2010; Hernandez et al., 2013; Lehrer & Schauble, 2004; Passmore & Stewart, 2002; Schwarz et al., 2009) and argumentation (Berland & McNeill, 2010; Grooms et al., 2015; Manz, 2015; Rowland et al., 2017) are well represented in the literature. Science and Engineering Practice 4 mainly emphasizes basic skills and competencies and lacks specific guidance on the instructional methods for teaching this practice (Lead States, 2013). Most of the literature involving analyzing and interpreting data is based on basic skills competencies rather than comprehensive data literacy (Bowen & Bartley, 2020). Thus, current standards for teaching, analyzing and interpreting data do not fully represent how working scientists work with data (Bowen & Bartley, 2020; Pangrazios & Sefton-Green, 2019; Wolff et al., 2019). Without a standardized and comprehensive framework for teaching, analyzing, and interpreting data, textbook and curriculum publishers can interpret the concepts of analyzing and interpreting data in any manner they choose, resulting in varying levels of data engagement across different curricula (Morris et al., 2015). This dissertation suggests that a comprehensive revision of standards for data literacy in K-12 science education, oriented towards the contemporary demands and applications of data in the 21st century, is needed to help teachers provide students with the scientific skills they need for the future.

1.1.2 Purpose of Study

My research addresses the abovementioned challenges by focusing on teacher perspectives and curriculum analysis. Chapters 2 and 3 examine contextual factors influencing teachers' approaches and challenges in analyzing and interpreting data with students in a classroom setting. The goal is to center teachers' perspectives. In Chapter 4, the study conducts a detailed analysis of a science curriculum to assess how it presents working with data and its

alignment with contemporary practical working with data. This seeks to identify opportunities for curricular enhancement. This dissertation investigates the nature of analyzing and interpreting data in 6-12 science education, delving into teachers' perspectives and how a curriculum delivers analyzing and interpreting data.

1.1.2 Significance of Study

This research aims to contribute to the existing literature on data literacy in secondary science education by examining how analyzing and interpreting data manifest in teacher instructional practice and curriculum design. The existing research landscape in data literacy within the context of 6-12 science education is limited, leaving room for a greater body of work. The studies specific to teacher practice and analyzing and interpreting data conducted by Banilower et al. (2018) and Rosenberg et al. (2022) have made valuable contributions, but there are opportunities and needs to expand this topic. Their work provided a broad-scope view of analyzing and interpreting data but lacked detailed insights into data instruction within the science classroom, particularly in the context of science education. Also, the existing research focused on data within curricula is limited. Morris et al. (2015) examined data activities across 70 textbooks, explicitly assessing the availability and rigor of data activities within instructional materials. Their findings revealed that while half of the activities included data, very few presented opportunities for students to engage with data, and instructional support was rarely provided. However, their work lacked information on how students might engage with data in the classroom. Chapter 4 centers on analyzing a single unit within a curriculum to add a more in-depth examination of these data activities. This approach situates these activities within the contemporary landscape of working with data in educational settings.

My research seeks to bridge the gap in more detailed knowledge of practices by adopting a distinctive approach centered around the teacher's voice. By utilizing focus groups, I aim to delve deeply into the perspectives and instructional practices of 6-12 science teachers regarding data literacy. This shift to focus group methodology places teachers at the forefront, allowing their experiences and insights to guide the research process. Unlike the existing studies that offer a general idea of how often they include SEP4 in their practice and the types of tools they use, my research hopes to capture a more detailed view of classroom practices, adding examples specific to working with data in their classrooms and increasing our understanding of how data literacy is taught in 6-12 science education. This study aims to provide valuable insights and a new perspective on analyzing and interpreting data, highlighted in the following sections.

In Chapters 2 and 3, the research expands on prior work by Banilower et al. (2018) and Rosenberg et al. (2022), by focusing on 6-12 science teachers' personal practices and approaches in analyzing and interpreting data in their classrooms. Chapter 2 investigates explicitly how these teachers approach data practices, addressing a gap in existing research that often surveys large participant pools without delving into contextual factors within science education classrooms. This study also diverges from previous approaches by utilizing focus groups, providing a platform for in-depth discussions, and amplifying teachers' voices to foster a more personal understanding of data practices. The research investigates how teachers integrate data literacy into instructional practices by deliberately opting for a small participant pool and employing a qualitative methodology.

Chapter 3 explores obstacles teachers face in teaching SEP4. This chapter also builds upon the foundation of Banilower et al. (2018) and Rosenberg et al. (2022). However, it focuses on teachers' challenges implementing data analysis and interpretation within their classrooms.

Both chapters 2 and 3 aim to discern the current educational landscape and potential future directions in teaching data literacy by employing a conceptual framework that integrates beliefs, practices, and contextual factors. The works also employ a synthesized model outlining best practices for teaching data literacy to guide recommendations for change.

Chapter 4 investigates the nature of data and how data analysis is integrated into the 7th-grade OpenSciEd curriculum. By assessing the curriculum through the perspective of Science and Engineering Practice 4 and synthesizing best practices for teaching data, the study contributes to curriculum evaluation literature, addressing the gap in comprehensive data literacy practices in curricula as identified by Morris et al. (2015). This study focused on how the OpenSciEd curriculum aligns with SEP 4 standards and modern data practices in science education classrooms. It builds upon the previous chapters, recognizing that teachers predominantly rely on curriculum as a primary or secondary source in their teaching (Morris et al., 2015; Stern & Roseman, 2004). This work hopes to heighten awareness of the importance of analyzing and interpreting data within 6-12 science education, providing a model for evaluating curriculum against updated data literacy methods. science education, which reflects the practices of scientists working with data in the field.

1.2 Literature Review

The literature review section will examine the literature on data literacy, NGSS Science and Engineering Practice 4, existing frameworks for teaching data literacy, educational context, and teacher practices, and the conceptual framework guiding this study. The literature on data literacy, Science and Engineering Practice 4, and teaching frameworks is essential for contextualizing this study within science education, forming the foundational understanding of data literacy in secondary science education.

1.2.1 What is Data Literacy?

Data literacy holds significant relevance in 21st-century education. Given the proliferation of Big Data and its integration into daily life, students must cultivate proficiency in data literacy to become well-informed citizens (D'Ignazio & Bhargava, 2015). Data literacy generally refers to the ability to collect, analyze, and interpret data (Bowen & Bartley, 2020; NGSS Lead States, 2013; Skrikant, 2017). Modern data literacy involves working with messy data (Bowen & Bartley, 2020; D'Ignazio, 2022; Kjellvik & Schultheis, 2019; Rosenberg et al., 2020; Schultheis & Kjellvik, 2020), working with authentic data (D'Ignazio, 2022; D'Ignazio & Bhargava, 2015; Grimshaw, 2015), understanding security and storage of data (Maybee & Zilinski, 2015), and identifying data's potential and limitations (D'Ignazio, 2022; D'Ignazio & Bhargava, 2015) among other aspects. While there is agreement on the importance of teaching about data in the classroom, a standardized framework is yet to be established (Pangrazios & Sefton-Green, 2019; Wolff et al., 2019). Nevertheless, existing frameworks and models within the literature indicate commonalities that merit consideration in developing a standardized framework.

In recognizing the need to emphasize the study of data collection and analysis, data literacy is included in the Program for International Learner Assessment (PISA) 2024 Strategic Informatics (PISA 2024 Strategic Vision and Direction for Science, 2020). In today's data-driven world, data literacy skills are increasingly vital, as they allow individuals to access, analyze, and comprehend data, which has become more pervasive and influential in our daily lives over the past three decades (Wolff et al., 2019; Pangrazios & Sefton-Green, 2019). Proficiency in data literacy entails understanding data collection, visualization, analysis, interpretation, and ethical considerations regarding data usage. With the rise of data-driven policies impacting various areas

such as social policy, marketing, and social justice, along with the lack of individual control over personal data, the need for a pedagogical framework for data literacy has become more urgent (Pangrazios & Sefton-Green, 2019; Wolff et al., 2019).

In science education, working with data aligns with the nature of science. Data is incorporated into the Next Generation Science Standards (NGSS) as Practice 4: Analyzing and Interpreting Data (NGSS Lead States, 2013). Practice 4 emphasizes the connection between data and scientific investigations in comprehending phenomena. Data literacy entails using appropriate quantitative and analytical tools to analyze, interpret, and communicate findings derived from data (Gibson & Mourad, 2018). It resides at the intersection of quantitative reasoning and data science, rooted in authentic contexts (Kjelvik & Schultheis, 2019). This viewpoint emphasizes using real and meaningful data sets for teaching data literacy, moving beyond contrived data solely for graph interpretation.

Data literacy in science education includes both skill-based competencies and contextual experiences. It involves the ability to critically analyze and interpret data from various perspectives, recognizing that data can be a multifaceted source of information with cultural, political, social, and ethical dimensions (Bhargava et al., 2016). Data literacy should go beyond basic skills and graphing, aligning with 21st-century needs to empower learners (Bhargava et al., 2016). Understanding data literacy is a complex undertaking. Researchers have proposed different conceptualizations of data literacy, with a shared understanding that it encompasses multiple concepts rather than being limited to basic competencies (Wolff et al., 2019). One widely used way of defining data literacy is through skill competencies (Gebre, 2022). These skills aim to equip learners with the ability to work with and analyze data. In data science education, D'Iganzio and Bhargava (2015) define data literacy as the desire and ability to engage

constructively in society through and with data. In information and library science, the definition expands to include critical assessment, management, and ethical use of data (Koltay, 2016a; Koltay, 2016b; Prado & Marzal, 2013). Another approach, rooted in research on math education, involves integrating data literacy into an inquiry process, particularly concerning data-driven decision-making. This perspective emphasizes that learners should investigate real-world issues, utilize data for evidence-based thinking, employ appropriate data, tools, and representations, develop and evaluate data-based inferences and explanations, and effectively communicate solutions in their data literacy curriculum (Vahey et al., 2012). These perspectives go beyond basic skills and comprehensively conceptualize data literacy in K -12 education. Data literacy extends beyond specific disciplines and is interdisciplinary (Vahey et al., 2012). The significance of data literacy is evident in its potential to enhance the economic impact of data, facilitate problem-solving, and empower individuals to hold governments accountable (Bhargava et al., 2016). These listed societal needs highlight the need to reconsider how data literacy is taught in science education as it is necessary to be an informed citizen.

1.2.2 21st Century Data Terms Defined

In this section, I will discuss prominent terms in contemporary literature on data literacy. I aim to define and examine these terms within the framework of 21st-century data applications to make a case for establishing connections to their relevance within the context of 6-12 science education.

1.2.2.1 Big Data. The term big data emerged in the early 2000s to define the volume, velocity, and variety of data gathered (Bhargava et al., 2015). Big data is used to describe amounts of data that are so vast that traditional methods cannot be used to store, analyze, or interpret. Big data has become ubiquitous in modern life, with datasets expanding to billions,

necessitating advanced technological tools for practical analysis. Its pervasiveness extends across societal and professional domains, encompassing data from routine activities such as grocery store transactions and social media interactions. The complexity of big data necessitates literacy in understanding and managing data, as students now require an awareness of the real-world implications embedded in big data.

As Bhargava et al. (2015) defined, big data emphasizes its multifaceted nature as an ecosystem shaped by the interplay of three fundamental components, commonly called the three Cs of big data. The three Cs encapsulate big data as big data crumbs, capacities, and communities. Each component contributes distinctively to the ecosystem of big data.

The big data ecosystem is founded on three key components: digital crumbs, big data capacities, and big data communities. Digital crumbs are fragments of data generated by digital devices, forming expansive datasets. Big data capacities refer to the tools and techniques needed to process and analyze these datasets, including data visualization and machine learning. Big data communities encompass all stakeholders involved in generating, analyzing, and using big data. Understanding these components is essential for effectively navigating and utilizing big data.

Data are found in traditional and nontraditional sources, i.e., social media, emails, and texts. With big data, we need new tools and technology to help manage it. Society, businesses, and science need ways to deal with it in real-time. Technological innovation allows scientists to collect massive amounts of information (Rosenberg et al., 2022). One example is iNaturalist, which has more than 140,000,000 observations accessible through any web browser or smartphone (Garner & Rosenberg, 2023; Rosenberg et al., 2022). However, this is not an isolated example. Big data is found across science and engineering disciplines (Rosenberg et al.,

2022). Given the rapid accumulation of information and the necessity for quick decision-making, it is vital to swiftly assess the quality and trends in the data, which can shift rapidly. The idea of big data is far removed from our K-12 classrooms, where traditionally, learners work with datasets they gather through small investigations and labs or preset cleaned data sets (Bowen & Bartley, 2020; Rosenberg et al., 2020). There is a place for those in K-12 science classrooms. Learners need to know and understand the vastness that data has become and how it dictates their lives in many ways.

The importance of data literacy is on the rise due to the scientific community's and society's increasing dependence on extensive data sets (Kjelvik & Schultheis, 2019; Schultheis & Kjelvik, 2020). Given the strong interconnection between science and data, we have a unique chance to incorporate data into K-12 science education, enriching learners' access to this valuable resource (Kjelvik & Schultheis, 2019; Schultheis & Kjelvik, 2020).

1.2.2.2 Authentic Data and Messy Data. Contemporary literature addressing data literacy within K-12 science education emphasizes the significance of authentic and messy data. This section aims to clarify and establish definitions for these data types and subsequently explore their relevance to science education at primary and secondary science education.

Authentic data is collected from real-life phenomena (Grimshaw, 2015; Kjelvik & Schultheis, 2019; Schultheis & Kjelvik, 2020). It automatically makes data relevant and provides a real-world connection to learners (Grimshaw, 2015). Scientists rely on a variety of authentic data, including data they collect, data obtained from collaborators, and data stored in online repositories where they may not have direct connections to the data collectors or sensors (Kjelvik & Schultheis, 2019; Schultheis & Kjelvik, 2020). These data sources can be classified into two categories: primary data collected by the researchers and secondary data acquired from external

sources (Kjelvik & Schultheis, 2019). Similarly, learners can access a wide range of authentic data in the classroom, such as data generated through their inquiry projects, searches conducted in online databases, figures and tables presented in textbooks, or scientific publications (Kjelvik & Schultheis, 2019).

Utilizing authentic data experiences enhances the quality of course projects. Grimshaw's (2015) work found that when working with authentic data, students can gain a sense of empowerment as they formulate questions and actively seek relevant data instead of relying on pre-curated datasets from repositories or other sources for their analyses (Grimshaw, 2015). Authentic data holds significance in working with data, as relying on curated datasets impedes students from engaging with data in a more comprehensive manner (Grimshaw, 2015). In addition, working with authentic data improves data literacy in students (Kjelvik & Schultheis, 2019).

Messy data is disorganized, has errors, and requires preprocess cleaning (D'Ignazio, 2022; Kjelvik & Schultheis, 2019; Schultheis & Kjelvik, 2020). Authentic data can be messy, but not always. Authentic data generally refers to data from a real-world context, and messy data refers to the state of the data (Kjelvik & Schultheis, 2019). Scientists in the field work with both types and should be included in science education. Messy data sets can exhibit variability, encompassing natural variations resulting from possible systematic errors and the precision of data collection methods (Kjelvik & Schultheis, 2019; Schultheis & Kjelvik, 2020). These data sets may also be incomplete or have missing values. The unique qualities of messy data offer distinct learning opportunities not readily available in other classroom activities, such as interpreting textbook figures or polished graphs from published studies (Bowen & Bartley, 2014; D'Ignazio, 2022; Kjelvik & Schultheis, 2019; Schultheis & Kjelvik, 2020). Instructors can

leverage messy data sets to foster critical thinking skills during data exploration, presenting authentic student opportunities (Bowen & Bartley, 2014; D'Ignazio, 2022; Kjervik & Schultheis, 2019). Students must develop the capacity to interpret data relationships that are less than perfect, as this more closely aligns with the nuanced nature of scientific data analysis (Bowen & Bartley, 2014; Rosenberg et al., 2020).

1.2.3 Science and Engineering Practice 4 in the Context of 21st-Century Use of Data

This section will examine Science and Engineering Practice 4 and explore its alignment with 21st-century approaches to working with data. The aim is to assess whether students' engagement with data aligns with the practices employed by scientists. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas* (National Research Council, 2012), the foundation for Next Generation Science Standards, identifies performance expectation goals at each grade band with Science and Engineering Practice 4 (SEP4) analyzing and interpreting data (Table 1.1). This section of the framework outlines what students in each grade band should accomplish in reference to SEP 4. Elementary is included here because the framework is set up as each band, laying the foundation for the next. By the twelfth grade, students should be able to systematically analyze data to identify significant patterns or assess data consistency with initial hypotheses. They should recognize instances where data conflicts with expectations and can revise initial models accordingly. The presentation of collected data is emphasized, requiring students to convey patterns and relationships effectively for communication with others. Proficiency in utilizing spreadsheets, databases, tables, charts, graphs, statistics, mathematics, and information and computer technology should be used for collating, summarizing, and displaying data, especially in exploring relationships between variables, including input and output.

Table 1.1 *K-12 Science Education Goals (National Research Council, 2012)*

Grade Band	Performance Expectation Goals
Elementary	Recognize the importance of recording observations and sharing them with others. Collect categorical and numerical data. Present data in tables or graphs. Introduce computer and digital tools when feasible. In engineering tasks, recognize that data serves the purpose of validating or improving a design.
Middle	Learn standard techniques for displaying, analyzing, and interpreting data (various graph types, identifying outliers, and averaging to mitigate measurement errors). Explain the rationale behind these techniques. In engineering tasks, recognize that data serves the purpose of validating or improving a design.
High School	Master additional techniques for displaying and analyzing data (x-y scatter plots, cross-tabulations). Recognize the need to explore multiple data display methods to identify significant features. Use mathematics and statistics to work with a greater diversity of data samples. Use computers or digital tools for analysis. In engineering tasks, recognize that data serves the purpose of validating or improving a design.

Additionally, students should demonstrate the ability to evaluate the strength of conclusions drawn from data using grade-level mathematical and statistical techniques, recognize patterns indicating relationships worthy of further investigation, and differentiate between causal and correlational relationships. Students are encouraged to collect data from physical models and analyze design performance under various conditions, highlighting practical applications. The table below outlines the performance expectation goals for each grade level, representing the knowledge and skill building in each grade band concerning SEP 4.

1.2.3.1 How the NGSS Incorporates Data-Related Practices. This section will address how the NGSS (Next Generation Science Standards) incorporate data-related practices primarily on SEP 4: Analyzing and Interpreting Data. This section aims to identify problems within the standards, highlight literature that addresses them, and provide an evaluation with current lesson examples. In selecting lesson examples, my approach is like that of a typical classroom teacher, who uses the internet to find content aligned with the specified standard. The evaluation of these lessons serves the purpose of highlighting the lack of specificity within the standards. Furthermore, it emphasizes the pragmatic reality that teachers who do not always have a provided curriculum (Stern & Roseman, 2004), who deliver hundreds of lessons a year, frequently resort to online resources as supplementary aids. It prompts consideration of the prevalence of such accessible online lessons, albeit not officially endorsed by NGSS, but widely utilized by teachers in their instructional practices as many teachers do not have access to a provided curriculum and must seek their own resources (Morris et al, 2015).

Examining an instance of a performance indicator concerning the SEP4 practice, consider the example- "MS-PS1-2 - Analyze and interpret data on properties of substances before and after the substances interact to determine if a chemical reaction has occurred (Lead

States, 2013)." Aligned with SEP 4, this performance indicator entails analyzing and interpreting data to identify similarities and differences in findings. A clarification statement included with this standard specifies permissible substances for experimentation, including burning sugar or steel wool, fat reacting with sodium hydroxide, and mixing zinc with hydrogen chloride. The associated Disciplinary Core Ideas (DCIs) cover the structure and properties of matter, emphasizing the identification of substances based on characteristic properties and chemical reactions, highlighting the regrouping of atoms in distinct molecules during a chemical process. A corresponding NGSS-aligned assignment (*MS-PS1-2 — the Wonder of Science*, 2021) exhibits alignment with the standard, offering comprehensive tables, pictures, and observation data, emphasizing fundamental data analysis skills. It is important to note that all data within this lesson has prerecorded and cleaned data, and its purpose is for students to analyze and interpret it for claims and evidence.

Consider another example: "MS-LS2-1 - Analyze and interpret data to provide evidence for the effects of resource availability on organisms and populations of organisms in an ecosystem (Lead States, 2013)." Accompanied by a clarification statement in this standard, highlighting cause-and-effect relationships between resource availability, organismal growth, and population dynamics, this performance indicator aligns with SEP 4's focus on analyzing and interpreting data to provide evidence of phenomena. Pertinent DCIs address interdependent relationships in ecosystems, emphasizing dependencies on environmental interactions and competition among organisms for limited resources. An NGSS-aligned assignment (*MS-LS2-1 — the Wonder of Science*, 2022) sourced from the same website and author mirrors the prior example, featuring pre-made tables and graphs for students to analyze and interpret, showing a consistent representation of SEP 4 in 6-12 science education. While not inherently problematic,

this uniformity may not fully capture the diverse applications of data in the contemporary era. Again, this lesson is prerecorded and cleaned, and its purpose is for students to analyze and interpret it for claims and evidence.

These examples highlight a primary issue with SEP 4: its emphasis on basic skill competencies, such as reading and interpreting graphs, rather than fostering a comprehensive understanding of data in the 21st century. While Practice 4 recognizes the importance of analyzing data and introduces tools like tables, graphs, data visualization, and statistical analysis, it falls short in addressing critical aspects of data analysis. As pointed out by Gebre (2022) and Bowen & Bartley (2014, 2020), current iterations of data literacy in curriculum lack discussions on understanding data's purpose, potential manipulation or bias in real-world applications, and fail to differentiate between quantitative and qualitative data, leaving out their respective purposes and analysis methods. The two examples above indicate that the data is already cleaned, organized, and presented linearly. Research methodologies are not delineated or explained. This practice presents data deterministically and lacks a probabilistic perspective (Bowen & Bartley, 2014; Bowen & Bartley, 2020; Dimopoulos & Karamanidou, 2013). Although NGSS Appendix F states that students will learn the difference between quantitative and qualitative data in elementary grades, it does not specify what should be taught (Lead States, 2013). One aspect to consider is that, particularly in the context of 6-8 education, science educators are more inclined to possess preservice training focused on education rather than science (Devetak & Vogrinc, 2013; Stern & Roseman, 2004). Consequently, they may have limited background knowledge in specific science skills, such as analyzing and interpreting data, and may require curated curriculum materials to enhance their own comprehension and instructional capabilities in imparting this skill (Devetak & Vogrinc, 2013; Morris et al., 2015).

The absence of explicit guidance and specificity within the Science and Engineering Practices may hinder teachers from effectively conveying this knowledge in the classroom.

Furthermore, another significant problem arises as SEP 4 inadequately covers the concept of "big data." While it involves working with graphical displays or large data sets, the definition of what qualifies as a "large set" remains subjective and unclear. This ambiguity leads to a discrepancy between the datasets encountered in curricula (e.g., interpreting 100 data points as a large set) and the massive datasets prevalent in actual 21st-century applications, potentially indicating a lack of trust in learners' ability to handle complex data (Rosenberg et al., 2022). In the two examples above, the data sets, though large, are presented in already cleaned and organized formats, which lessens the impact of what large data means. These lessons do not allow students to work with computers or digital tools to analyze data.

Additionally, SEP 4 lacks guidance on discriminating between tools used in data analysis. SEP 4 includes basic skill competencies related to data literacy, such as distinguishing causal and correlational relationships and using data as evidence, which the lesson examples cover. Although SEP 4 does mention utilizing concepts of statistics and probability, it fails to guide when and why learners should use specific tools or software for data analysis. The example lessons do not provide opportunities for students to use tools.

Lastly, SEP 4 falls short in providing explicit links to real-world applications. Real-world context is considered an important strategy when teaching data literacy skills (Bhargava et al., 2015; D'Ignazio & Bhargava, 2015; Gebre, 2022; Lee et al., 2021; Rosenberg et al., 2020; Schwarz et al., 2017). While it mentions considering limitations and errors, it lacks specificity in connecting these concepts to real-world scenarios, diminishing the practical applicability of the skills taught within the framework.

SEP 4 predominantly emphasizes the basic skills competencies associated with data manipulation. However, it fails to establish a comprehensive framework for fostering a thorough understanding of data. Data representation appears definite and absolute, lacking perspectives that encourage learners to scrutinize data critically and assess its validity. SEP 4 overlooks the broader and deeper dimensions of data literacy, which should ideally be regarded with the same significance as traditional reading literacy (Bhargava et al., 2015). Since the emergence of Big Data, the importance of citizens' capacity to critically analyze and question data has become necessary, as (Radinsky & Tabak, 2022) highlighted. Citizens need to discern whether data is employed manipulatively to evoke emotional responses (Radinsky & Tabak, 2022). A more comprehensive data literacy framework is necessary for K-12 science education to transcend basic skill competencies, encompassing the ability to engage with data discerning and critically.

1.2.3.1 Critiques and Opportunities within SEP4. The NSTA (National Science Teachers Association) book, *Helping Students Make Sense of the World Using Next-Generation Science and Engineering Practices*, recognizes the central role of Practice 4 in Science and Engineering activities (Schwarz et al., 2017). The book emphasizes key aspects of utilizing Science and Engineering Practice 4 for effective data engagement. Emphasizing the purpose of working with data to solve problems, the book advocates for employing tools such as Venn Diagrams, GIS images, and mathematical thinking. It stresses the interconnectedness of Practice 4 with other scientific practices and highlights the importance of collaboration in this context. The book validates the significance of analyzing observations as a legitimate form of data and encourages working with relevant data, emphasizing the value of local and community resources. However, it broadly defines data as facts, statistics, observations, or information from natural and built environments (Schwartz et al., 2017). This literature discusses the diversity of data

sources and types. However, it does not fully address the need for critical data literacy skills in the modern world or explicitly provides specific strategies for analyzing and interpreting data.

Research critiques how data is worked with in science classrooms. Often, the data students collect or work with are sanitized versions of work already done (Bowen & Bartley, 2014; 2020). Students are not working with data like scientists do in the field; they are replicating investigations or working with cleaned data sets and not getting the entire experience with the data cycle. This type of data is helpful for students to understand already known relationships but does little to prepare students for how scientists work with data in the real world (Bowen & Bartley, 2020; Dimopoulos & Karamanidou, 2013). Additionally, the representation implies that data is viewed as absolute or deterministic without delving into conceptualizing scientific outcomes more probabilistically. There is a disparity between scientific reasoning within the classroom and practical applications in the field. A figure that illustrates the potential impact of this disparity on students' perception of scientific investigations and their findings can be found in Appendix A.

In summary, NGSS Practice 4 primarily focuses on basic data analysis skills but lacks the depth and breadth required for comprehensive data literacy in the 21st century. Real-world data demands a more critical understanding of data's purpose, potential biases, and the ability to navigate and question data effectively. The current NGSS practices do not fully meet these requirements, emphasizing the need for a broader and more critical approach to data literacy education.

1.3 Conceptual Framework

My conceptual framework comprises three interconnected components: educational context, context and teacher practice, and a synthesized model of best practices for teaching data

literacy (SBPM). Each element serves a distinct purpose in my research. First, the context component allows for examining the contextual influences teachers encounter when teaching data literacy. The second component, context, and practices, allows for understanding how contextual factors impact practice. The third component of the framework involves a synthesized model called the Synthesized Best Practices for Teaching Data Literacy Model (SBPM). This model, derived from current literature, outlines best practices for teaching data literacy. It serves as a foundational lens, providing a clear perspective on what teaching data literacy should entail in line with current data literacy literature. The interconnectedness of these three components, detailed in Appendix B, allows for a comprehensive exploration of the factors influencing the teaching of data literacy in science education classrooms. In the following sections, I will review the literature relevant to the three components of my conceptual framework and explain how the component fits within my research.

1.3.1 Context in Education

In education, contexts are intricate interacting systems influenced by various factors that impact the learning process (Cobb et al., 2003; Jones & Carter, 2007; Lumpe et al., 2000; Luft et al., 2022). These factors include the educational environment, adherence to state standards, and teachers' and students' backgrounds (Cobb et al., 2003; Jones & Carter, 2007; Lumpe et al., 2000). While listing all contextual influences proves impractical due to their vast and diverse nature, identifying common contextual factors in studies can contribute to a more comprehensive theoretical framework for practitioners (Cobb et al., 2003). The contextualization of data in science education necessitates the consideration of multiple variables, such as teachers' perceptions of data importance, resource availability, curriculum employed, and the broader teaching environment (Tabak, 2004). Research design should incorporate exogenous, referring to

instructional materials and activity structures, and endogenous, involving existing materials and practices (Tabak, 2004). Broad theories often fall short in practical use within science education, as they may overlook specific contextual factors, emphasizing the need to identify commonalities across diverse contexts for more universal and practical understanding (Anderson & Shattuck, 2012; Berliner, 2002; Rosenberg & Koehler, 2015).

Within the scope of my research, the design encompasses an exploration of teacher discourse regarding data in their classrooms, an assessment of the accessibility and obstacles encountered in implementing data activities, and a critical examination of how the OpenSciEd curriculum presents data. Conducting research with a focus on contextual influences is pragmatic and beneficial for educational purposes, as it enhances validity and applicability to teachers (Anderson & Shattuck, 2012; Jones & Carter, 2007). The recognition of the significance of context in this research is essential. Identifying themes among these various contexts will contribute to a clearer understanding of teachers' approaches to analyzing and interpreting data and their challenges, potentially highlighting the need for further exploration in this area of study.

By engaging in discourse within focus groups, I intend to uncover the contextual factors influencing teachers' beliefs and subsequent actions in thinking about and incorporating analyzing and interpreting data in their classrooms in Chapter 2. For Chapter 3, I will use this approach to recognize the accessibility and obstacles teachers encounter when implementing the practices of analyzing and interpreting data. For Chapter 4, I will analyze a curriculum to understand its role as a contextual influence on instruction and teacher practice. Specifically, I will examine how the curriculum presents data and aligns with best practices for teaching data literacy.

1.3.2 Context and Teaching Practice

In science education, understanding the impact of contextual factors on teaching practices is crucial for fostering effective instructional strategies. This section highlights the interaction of various elements within educational environments and their influence on teacher practice. Factors such as adherence to state standards, the backgrounds of teachers and students, and the availability of instructional materials shape the educational landscape (Cobb et al., 2003; Jones & Carter, 2007; Lumpe et al., 2000; Luft et al., 2022). These factors can either enhance or hinder a teacher's ability to deliver effective instruction.

1.3.2.1 Teacher Self-Efficacy and Experience. Teacher self-efficacy is a significant influence, where teachers' confidence in their abilities directly affects their instructional practices (Dong et al., 2018; Thibaut et al., 2018b; Wright & Waxman, 2022). A strong sense of self-efficacy, supported by robust content knowledge often acquired through hands-on training, is vital for positive outcomes. Furthermore, professional development is important, with effective programs markedly improving how teachers implement classroom instruction (Dong et al., 2018; Thibaut et al., 2018b; Wright & Waxman, 2022). Additionally, both administrative support and collegial interactions are pivotal, facilitating the adoption of innovative teaching methods and the exchange of best practices.

The impact of teaching experience varies, with seasoned educators sometimes resisting new teaching methods. However, accumulated knowledge can also enrich a teacher's effectiveness (Thibaut et al., 2018b; Wright & Waxman, 2022). It is difficult for seasoned teachers to shift from more traditional ways of teaching to reform-based practices (Dong et al., 2018; Wright & Waxman, 2022). Resource constraints, particularly time and access to materials, emerge as significant barriers, restricting the ability to implement desired instructional methods

and engage in professional development (Dong et al., 2018; Thibaut et al., 2018b; Wright & Waxman, 2022).

1.3.2.2 District Standards and Contextual Factors. A school-level focus on content- and instruction-focused professional development results in high-quality teaching practices (Dong et al., 2018; Thibaut et al., 2017; Wright & Waxman, 2022). Time for planning, collaboration, and funding are contextual factors that impact practice most (Wright & Waxman, 2022; Dong et al., 2018). Ensuring adequate human, social, and physical resources and time available is crucial (Wright & Waxman, 2022). Teachers must be emotionally and pedagogically engaged (Dong et al., 2018). Insufficient access to training, curriculum, and resources negatively impacts the translation of concepts into teacher practice (Dong et al., 2018). Teachers know what can help improve their practice but often cannot implement these improvements due to unchanging contextual factors (Lumpe et al., 2000). Hands-on experience increases the translation into practice (Thibaut et al., 2017).

Curriculum affects teacher practice, and teachers will alter instructional practices for better or worse based on contextual factors (Dong et al., 2018; Rubie-Davies et al., 2011). Addressing these contextual factors is essential for developing supportive educational environments that enhance teaching practices and improve educational outcomes in science education. This approach acknowledges the complexity of educational contexts and highlights the need for a multi-faceted support system. Studying the context of practice is essential because teachers with the same background and training could teach vastly differently based on factors such as resources, time, and curricula (Fletcher & Luft, 2011).

Addressing these contextual factors is essential for developing supportive educational environments that enhance teaching practices and improve educational outcomes in science

education. This approach acknowledges the complexity of educational contexts and highlights the need for a multi-faceted support system. Studying the context of practice is essential because teachers with the same background and training could teach vastly differently based on factors such as resources, time, and curricula (Fletcher & Luft, 2011).

1.3.3 Synthesizing Frameworks for Best Practices

This section will describe my methods for synthesizing best practices for data literacy. Given the scarcity of frameworks designed explicitly for 6-12 science education, I sought adaptable frameworks. One such framework, Open Inquiry (Bowen & Bartley, 2014), caters specifically to K-12 science education, while Data Informed Learning (Maybee & Zilinski, 2015) pertains to science education in higher education. The Foundational Pillars of Data Literacy (D'Ignazio & Bhargava, 2015) and Workflow through Data Science Tutorial (Srikant & Aggarwal, 2017), designed for data science education, were also included in my selection. I draw upon these four frameworks to determine the best practices for teaching data literacy in 6-12 science education. These frameworks serve as guiding principles for instructing data literacy in the classroom and contribute significantly to the ongoing discourse within data literacy education research. By offering diverse perspectives on the practical analysis and interpretation of data, they align with contemporary practices in working with data.

1.3.3.1 The Frameworks. This section will explore four frameworks for teaching data literacy. These frameworks provide valuable insights into the importance of data literacy in education and its relevance in various contexts, including science education. They are instrumental in helping us identify what is essential to teach when it comes to analyzing and interpreting data.

Open inquiry (2014) provides a basic framework for allowing students to work with data in a more relevant way, with the full data cycle, and supports a shift in thinking from deterministic to probabilistic thinking in K-12 science education (Bowen & Bartley, 2014; 2020, Dimopoulos & Karamanidou, 2013). Bowen and Bartley (2014) define open inquiry (Appendix C) as investigating a problem that is ill-defined, allowing for student modification, allowing students to experience uncertainty and ambiguity, starting at the current knowledge state of the student, and is used to locate conceptual foundations (start where they are), allows materials and discursive practices and resources to be shared and developed within a community of learners, and allows for newcomers to draw on others' experiences and expertise (p. 6). Engaging in an investigation involving an ill-defined problem will enable students to modify the project as they progress. This approach encourages students to grapple with uncertainties and ambiguities when drawing conclusions, fostering a deeper understanding of the subject. By starting with the student's current knowledge in each project, the learning experience becomes more personalized and relevant.

Furthermore, such investigations can create a collaborative learning environment, allowing for the sharing of materials, discursive practices, and resources within a community of learners (Bowen & Bartley, 2014). This collaborative setting offers newcomers the chance to draw on the expertise of others and utilize any available resource that could enhance their learning, promoting a dynamic and interactive learning process (Bowen & Bartley, 2014). This inquiry better reflects how scientists work in the field, supporting data used in real-world contexts and higher-order thinking skills in K-12 education (Fontichiaro & Johnston, 2020). In Figure 4, the levels of inquiry are shown in relation to problems, procedures, and conclusions. Science textbooks and curricula generally stay around the inquiry levels of 0-1, occasionally

reaching 3, while scientists work typically in a 3 (Bowen & Bartley, 2014). Bowen and Bartley's model (2014) illustrates a need for a more comprehensive approach to working with data in K-12 science that reflects the nature of science in the field.

D'Ignazio and Bhargava's (2015) Foundational Pillars of Data Literacy framework challenges the conventional view of data literacy and focuses on empowering citizens and communities through data literacy (Raffaghelli et al., 2020). The first pillar emphasizes agility and adaptability, focusing on cultivating adaptive capacities and resilience. This approach prioritizes the development of skills rather than fixating on specific platforms and technologies, recognizing the inevitability of technological evolution.

The second pillar envisions building upon essential features across all subcategories of literacy, conceiving literacy as a continuum rather than isolated domains. The third pillar involves empowering individuals to navigate their current ecosystems and societies in ways that are personally significant to them. The fourth pillar advocates for providing diverse pathways to accommodate individuals with varying data literacy needs and capacities, acknowledging the complexity of the overarching system. The four pillars advocate for shifting from rigid technical competencies toward a holistic understanding of data literacy as essential for 21st-century life. Their framework emphasizes the significance of data literacy in fostering societal inclusion and citizenship and recognizing the relevance of data to learners' lives (D'Ignazio, 2022; Raffaghelli et al., 2020).

Maybee and Zielinski's (2015) Data-Informed Learning Principles framework (Appendix C.2) stresses the importance of building upon learners' prior experiences when teaching data literacy. They emphasize integrating new ways of using data into disciplinary core subjects so that learning leads to students developing new understandings of the subject being studied. This

framework highlights the necessity of context in data-informed learning, reinforcing that data literacy should not be taught in isolation but incorporated into all learning contexts (Koltay, 2016b; Raffaghelli, 2020).

This approach aligns with the concept of relevance, particularly in science education. Teaching data literacy through the lens of students' personal experiences and communities can engage students and deepen their understanding of the data cycle. While this framework initially aims at higher education, it can be adapted to K-12 science education.

Srikant and Aggarwal's (2017) Workflow through Data Science Tutorial framework is particularly relevant to middle-grade science education, as it directly involves K-12 learners in the data collection process and emphasizes the need for relatable data (Appendix C.2). Their framework promotes the full data cycle, relatable data sets (e.g., about themselves), and hands-on data collection and entry (Miao et al., 2022). By reducing problem complexity, this approach ensures learners can comfortably engage with data to solve real-world, contextual problems related to their experiences.

This framework is a strong foundation for building data literacy in science education (Wolff et al., 2019). It demonstrates that learners can be motivated and engaged with relatable and meaningful data. While this framework originated from a specific study on data literacy, it can be adapted for use in various contexts where data is a valuable learning resource.

In summary, these four frameworks provide valuable insights into the multifaceted nature of data literacy and its relevance in education. They highlight the importance of working with relevant data and integrating data literacy into learning contexts, particularly in science education. These frameworks offer flexible approaches that can be adapted to suit various

educational levels and settings, ultimately empowering learners to navigate the data-driven world effectively.

1.3.3.2 Synthesizing Best Practices for Teaching Data Literacy. I conducted an in-depth examination of the commonalities among these frameworks, synthesizing a modified version that comprehensively incorporates their shared aspects (Appendix D). This synthesized model enables a comparative analysis between the existing and ideal conceptualizations of teaching data literacy. This synthesized model aligns with what the literature on data literacy says is necessary for teaching data literacy.

In contrast to the initial two components of the conceptual framework, which focus on examining teacher practice and contextual factors within their practices, this component delineates a best practices model for data literacy in 6-12 science education based on current research. This synthesized framework will serve as the deductive codes for analysis for in Chapters 2 and 4, allowing for the comparison of existing practices with the synthesized best practices from the literature, thereby providing an evaluation of the alignment between current practices and an idealized framework for teaching data literacy in science education.

1.4 Positionality

As a researcher, I acknowledge that my background, experiences, and beliefs shape my approaches to these studies (Cain et al., 2023). I am a white Appalachian woman who has taught middle-grade science for 17 years in low-socioeconomic schools. My perspectives are influenced by my upbringing and my experience as a teacher. My interest in this research comes from my background and interest in making science education accessible for all students.

My educational background includes a bachelor's degree in Human Learning K-8 and a Master's in Curriculum and Instruction. These experiences have provided me with a strong

background in pedagogical skills. My teaching experience in diverse classroom settings has highlighted the disparities in science educational resources and opportunities, particularly in low-income areas.

In my paradigmatic beliefs, I lean toward critical realism (Gorski, 2013) and interpretivism (Alharahsheh1 & Pius, 2020). I deeply believe that we are constantly influenced by contextual factors that shift minute by minute and day by day. Understanding participants' stories is crucial to identifying the reasons behind their choices and responses. Therefore, it is essential to engage directly with participants, as survey methodologies alone do not provide the depth of understanding necessary for informed research.

I also embrace interpretivism, recognizing that the researcher interprets all data in a way unique to their perspective and interactions with participants. This process creates a shared reality and mutual understanding. Through discourse with participants, the researcher and participants become partners in the research journey.

Recognizing that my positionality may introduce biases in interpreting data, especially in interpreting approaches and challenges that teachers face in teaching data literacy in 6-12 science education, I am committed to transparency and accountability in my research. I also acknowledge that my knowledge and experience help me interpret data because, due to my background, I know what to look for. To mitigate biases, I have engaged in reflexive practices throughout this process, including journaling to reflect on my positionality. By explicitly stating my positionality, I aim to foster trust and reliability in my research.

1.5 Introduction to the Three Papers

This section will describe the three papers within this dissertation and their purpose. My research tells a story through these papers, and in Figure 1.2, I created a chart showing how they

connect. Chapter 2 examines teachers' approaches to teaching analyzing and interpreting data in their practice. I want to determine how the participants define data literacy and what approaches do teachers use to teach data analysis and interpretation. Additionally, how do these approaches align with recognized best practices in the field, such as Practice 4: Analyzing and Interpreting Data (National Research Council, 2012) and the Synthesized Best Practices Model (SBPM)? Chapter 3 focuses on the obstacles teachers face in teaching, analyzing, and interpreting data. I want to identify the contextual factors that impede effective data literacy teaching practices in 6-12 science education classrooms. Chapter 4 focuses on data in a science curriculum. It is an analysis of the OpenSciEd 7th-grade curriculum. This chapter focuses on a curriculum, which is a major contextual factor that influences teacher practice. It evaluates how the curriculum presents data to students in addition to how it matches the best practices model. I want to determine what characterizes the nature of the data used by students in the OpenSciEd 7th-grade curriculum, including the types of data, visualizations, the size of the datasets, sources of data, and variable types. Also, I want to determine how analyzing and interpreting data is incorporated within the OpenSciEd curriculum at the seventh-grade level. This includes the methods of data analysis students participate in and how this analysis aligns with modern data practices in the 21st century. The three chapters will provide a comprehensive look at how data literacy is contextualized in science education and guide future research.

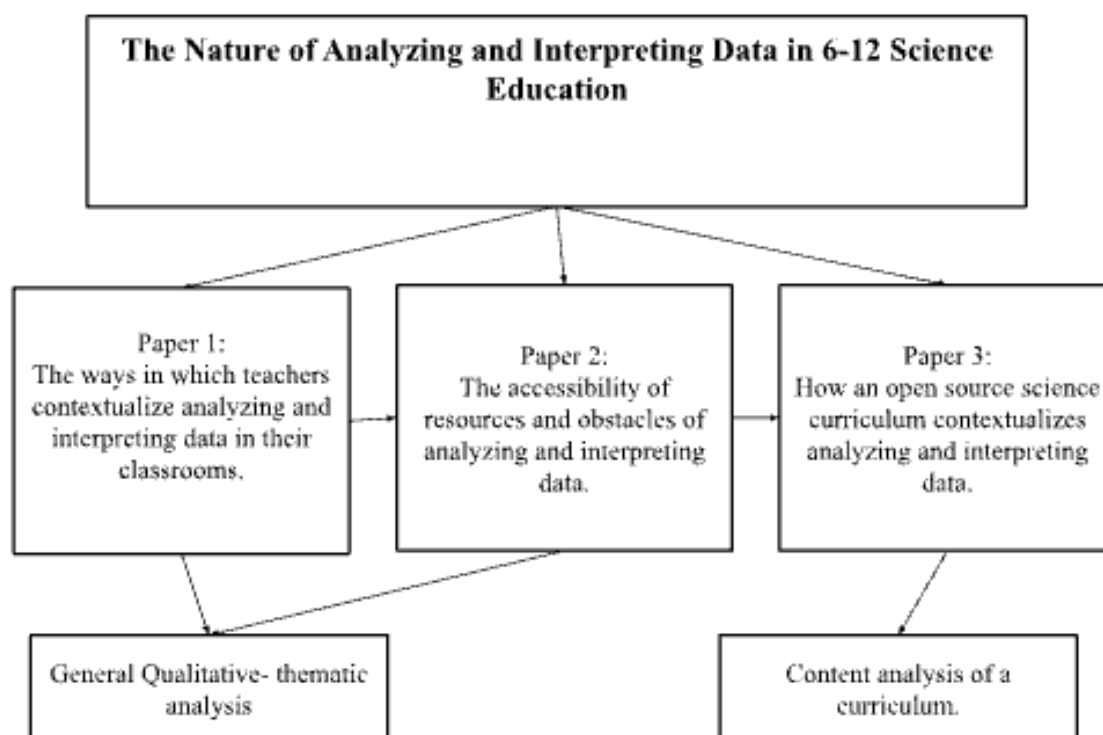


Figure 1.1 *The Three Papers*

Chapter 2. 6-12 Science Teacher Approaches to Teaching Analyzing and Interpreting Data

2.1 Introduction

In education, the terms "data" and "data literacy" still lack a universally accepted definition suitable for the 21st century, with existing definitions often described as narrow and limited to basic aspects like data collection, analysis, and interpretation (Gebre, 2020). A *Framework for K-12 Science Education* (National Research Council, 2012) emphasizes data organization into visualizations but fails to provide details on aspects such as cleaning raw data and defining a full data cycle (Bowen & Bartley, 2014). This limitation has carried through to the Next Generation Science Standards (Lead States, 2013). Furthermore, the Next Generation Science Standards (NGSS) Science and Engineering Practice 4 (SEP 4)-Analyzing and Interpreting Data focuses primarily on basic skill competencies, lacking specificity in data size, the full data cycle, and types of visualizations. This practice emphasizes data representation to reveal patterns and relationships and effectively communicate findings (Lead States, 2013; Schwarz et al., 2017), yet it falls short of providing explicit guidelines for integrating data literacy into a modern relevant educational framework (Bowen & Bartley, 2020; Kang et al., 2019).

Despite the critical role of SEP 4 within NGSS, research on its implementation in educational contexts remains sparse. Existing literature reveals that teacher engagement in data analysis and interpretation is limited, with studies like Banilower (2018) and Rosenberg et al. (2022) indicating that while teachers often employ traditional methods such as paper and pen for organizing data through tables, charts, and graphs, there is a general lack of comprehensive insights into the practices they use and the challenges they face. Approximately 11% of NGSS Performance Expectations involve data-related tasks (Kastens, 2014), yet teachers struggle to

implement it due to unclear instructional guidance (Morris et al., 2015; Stern & Roseman, 2004). Furthermore, while educators frequently use student-collected data (Rosenberg et al., 2022), there is a noticeable decline in engagement as the complexity of these tasks increases (Banilower, 2018) showing a need for further research into data-handling practices in 6-12 science education. This study seeks to bridge this gap by exploring how teachers teach data analysis and interpretation and determining whether these methods align with SBPM. This research study directly examines what teachers report about their practice rather than relying on survey data. It accounts for diverse factors that can influence teacher practices and instructional methodologies (Fischer et al., 2018; Lumpe et al., 2000), by allowing teachers to speak about their own experiences of how they implement data analysis and interpretation in various educational contexts, such as time and training. This study aims to provide insights into the approaches teachers use in teaching SEP 4 and possibly create a foundation for best practices in effective data literacy instruction in classrooms.

2.2 Literature Review

2.2.1 What is Data Literacy?

The definitions of data literacy within science education are evolving as the capacity to collect, analyze, and interpret data in society changes. This section explores the evolving definitions of data literacy within science education, emphasizing the necessity for a modernized interpretation that aligns with the complexities of the 21st-century data landscape. Current scholarly works highlight various facets of data literacy, including the ability to handle messy and authentic data, to understand data security and storage, and to recognize data's potential and limitations (Bowen & Bartley, 2020; D'Ignazio, 2022; Kjelvik & Schultheis, 2019; Rosenberg et al., 2020; Schultheis & Kjelvik, 2020; Maybee & Zilinski, 2015; D'Ignazio & Bhargava, 2015;

Grimshaw, 2015). Despite the widespread acknowledgment of its importance in classrooms, a framework for teaching data literacy remains underdeveloped (Pangrazios & Sefton-Green, 2019; Wolff et al., 2019). This section explores the evolving definitions of data literacy within science education.

Data literacy is crucial in contemporary education due to data's pervasive influence across various sectors and its central role in STEM workforce development (Wolff et al., 2019). Data literacy is also essential for making informed decisions in an increasingly complex world (Wolff et al., 2019; Pangrazios & Sefton-Green, 2019). Within science education, data literacy intersects with scientific inquiry as encapsulated in Practice 4 of the Next Generation Science Standards (NGSS): Analyzing and Interpreting Data (NGSS Lead States, 2013). However, current implementations in educational settings do not fully capture the breadth of skills required to navigate and utilize data effectively.

Scholars across various fields have proposed richer, more comprehensive conceptualizations of data literacy. These include not only basic skills such as visualization creation and interpretation but also extend to critical assessment, ethical data usage, and the ability to engage with data constructively in societal contexts (D'Ignazio & Bhargava, 2015; Koltay, 2016a; Koltay, 2016b; Prado & Marzal, 2013; Gebre, 2022). This broader view incorporates using data for evidence-based decision-making, employing appropriate tools, and effectively communicating solutions, highlighting the interdisciplinary nature of data literacy that transcends traditional subject boundaries like mathematics (Vahey et al., 2012).

Ultimately, data literacy should be conceptualized as encompassing technical competencies and the critical analysis and interpretation of data from diverse perspectives—cultural, political, historical, social, emotional, critical, and ethical—acknowledging that data is

more than mere numbers but a powerful tool that can drive social change and impact historical narratives (Bhargava et al., 2016). The ability to access and work with data is power (D'Ignazio & Bhargava, 2015). As such, defining data literacy in a more comprehensive way in science education is vital for empowering students to navigate and leverage data meaningfully in the 21st century.

2.2.3 An Examination of NGSS Practice 4- Analyzing and Interpreting Data

The NGSS sets comprehensive performance expectations across grade levels for Science and Engineering Practice 4 (SEP 4) - Analyzing and Interpreting Data (Lead States, 2013). This framework progressively builds data analysis skills from elementary through high school, detailing what students should achieve at each grade band, and is based on *A Framework for K-12 Science Education* (National Research Council, 2012). Starting in grades K-2, students begin by collecting, recording, and sharing observations through various methods such as pictures, drawings, and writings, using these to describe patterns and relationships to address scientific questions. By grades 3-5, they employ quantitative methods for data collection and start representing data in graphical displays like bar graphs and pie charts to analyze and interpret phenomena.

As students advance to grades 6-8, the focus shifts to more formal quantitative analyses, distinguishing between correlation and causation and using basic statistical techniques to enhance precision and accuracy. By high school, they engage in detailed statistical analysis using advanced tools and models to make scientific claims and evaluate the impact of new data on existing models. This framework emphasizes proficiency in utilizing digital tools and various data representation methods, including spreadsheets, databases, and graphical displays, to convey data effectively.

However, despite this structured approach, SEP 4 faces challenges that hinder comprehensive data literacy education. It primarily focuses on basic skills, failing to expose students to the complexities of big data, messy data, and authentic data sets, which could foster a more profound understanding of data's real-world applications and nuances (Bhargava et al., 2015). The current standards often neglect essential aspects such as exploring data's purpose, potential manipulation, and the differentiation between quantitative and qualitative data. This oversight contributes to a deterministic view of data presentation and lacks a probabilistic perspective (Bowen & Bartley, 2014). Consequently, there is a pressing need for a more holistic data literacy framework that transcends basic skill competencies and fosters critical thinking and discernment in data analysis. This enhanced approach, as advocated by D'Ignazio, should equip students with the skills to critically question and evaluate data, aligning with the evolving demands of a data-driven world. Doing so ensures that education in data analysis remains grounded in current standards and practices as outlined by the NGSS Lead States (2012), effectively preparing students for the complexities of modern data environments.

2.2.4 Data Literacy and SEP 4

This section highlights two key scholarly works exploring data analysis and interpretation as delineated in Science and Engineering Practice 4 (SEP 4) of the Next Generation Science Standards (NGSS). The goal is to deepen the understanding of how SEP 4 is represented in educational resources and to critique existing pedagogical methodologies.

Schwarz et al. (2017) emphasized the critical role of SEP 4 in Science and Engineering, particularly how it integrates with other practices and the importance of collaboration. This work discussed key elements essential for effective data engagement, such as using Venn Diagrams, GIS images, and mathematical thinking. However, while the book acknowledged the use of

varied data sources and the importance of contextually relevant information, it is noted that the discussion does not extensively cover critical data literacy skills or offer detailed strategies for data analysis and interpretation.

On the other hand, Bowen and Bartley (2014) provided a focused examination of data literacy within SEP 4, offering specific strategies for educators to foster a deeper and more critical engagement with data. Their critique centered on the conventional teaching approaches in science classrooms, which often involve sanitized or replicated data, thereby limiting students' exposure to the full data cycle experienced by field scientists. They argued that this practice does not adequately prepare students for engaging with authentic scientific data and tends to portray data in a deterministic rather than probabilistic manner. This discrepancy underscores a significant gap between classroom learning and real-world scientific practices, impacting students' understanding of scientific investigations.

Together, these works demonstrate a need for a shift towards more probabilistic thinking in science education. This review contextualizes the instructional resources for SEP 4 and highlights the need for pedagogical advancements that embrace the complexities of real-world data engagement.

2.2.5 Four Frameworks for Teaching Data Literacy

This section synthesizes four influential frameworks that shape contemporary approaches to teaching data literacy within science education, emphasizing relevance and empowerment for learners across various educational contexts. The Open Inquiry approach, advocated by Bowen & Bartley (2014, 2020) and Dimopoulos & Karamanidou (2013), promotes a shift from deterministic to probabilistic thinking in science education by engaging students with the full data cycle—collection, analysis, and interpretation. This framework encourages investigation

into ill-defined problems, fostering personalized learning and collaboration within a community of learners. The Four Core Pillars of Data Literacy, as outlined by D'Ignazio & Bhargava (2015) and Raffaghelli et al. (2020), redefine data literacy to include agility and adaptability, focusing on empowering individuals to navigate their social ecosystems effectively. This framework emphasizes developing adaptable skills that transcend specific technologies and fosters a holistic understanding of data literacy as essential for societal inclusion and active citizenship. The Data-Informed Learning framework by Maybee & Zielinski (2015) and further discussed by Koltay (2016b) and Raffaghelli (2020), integrates data literacy within disciplinary subjects by building on learners' prior experiences, making it contextually relevant and a part of everyday learning. Lastly, the Workflow Through Data Science Tutorial Design by Srikant & Aggarwal (2017), supported by Miao et al. (2022), involves middle-grade students directly in the data cycle with relatable data sets, emphasizing hands-on collection and practical application. Together, these frameworks advocate for a shift in educational paradigms where data literacy is not only about technical skills but also about engaging learners in ways that are meaningful to their personal lives and community involvement, thus establishing data literacy as a critical component of modern science education and necessary for developing informed, capable citizens in a data-driven world.

2.2.6 The Role of Textbooks in Science Education

Textbooks play a significant role in science education, serving as the primary or secondary instructional resource (Morris et al., 2015) and are a key aspect of curricula (Stern & Roseman, 2004). They provide essential information and pedagogical guidance, especially in K-8 settings where teachers often have backgrounds in education rather than science (Stern & Roseman, 2004). Using textbooks that include textual and pictorial content enhances teaching

and learning experiences (Stern & Roseman, 2004). However, how textbooks are written can influence perceptions of science, often presenting it as absolute and leading to misconceptions (Bowen & Bartley, 2014; Lee, 2010).

Textbooks are generally seen as synonymous with curricula, meaning they provide content and shape the instructional framework and pedagogical approaches teachers use in their classrooms. This synonymy emphasizes the importance of high-quality textbooks in ensuring effective science education. When assessing science textbooks, content alignment with scientific principles is crucial, along with effective instructional support that considers students' prior knowledge (Devetak & Vogrinc, 2013; Stern & Roseman, 2004).

Textbook authors have significant autonomy, resulting in diversity among textbooks and curricula (Devetak & Vogrinc, 2013). Therefore, critically examining these resources is essential, as teachers rely heavily on them to shape their teaching practices.

2.3 Conceptual Framework

2.3.1 The Role of Contextual Factors in Teacher Practice

In science education, understanding the impact of contextual factors on teaching practices is crucial for fostering effective instructional strategies. This section highlights the interaction of various elements within educational environments and their influence on teacher effectiveness. Contextual factors are important to my research design because they can influence teacher practice in analyzing and interpreting data. A teacher may want to include more comprehensive data activities but can't be due to time or resource availability. However, contextual factors are so vast it is not possible to name them all (Cobb et al., 2003; Jones & Carter, 2007; Lumpe et al., 2000). This section discusses the research on contextual factors and how they influence teacher practice (Cobb et al., 2003; Jones & Carter, 2007; Lumpe et al., 2000; Luft et al., 2022). These

factors provide a lens through which my research is designed and can either enhance or hinder a teacher's ability to deliver effective instruction.

Teacher self-efficacy is a significant influence, where teachers' confidence in their abilities in a specific area directly affects their instructional practices (Dong et al., 2018; Thibaut et al., 2018b; Wright & Waxman, 2022). A strong sense of self-efficacy related to teaching about data literacy, supported by robust content knowledge often acquired through professional development, is vital for positive outcomes (Bandura et al., 1999). Indeed, effective professional development has been shown to improve how teachers implement classroom instruction markedly (Dong et al., 2018; Thibaut et al., 2018b; Wright & Waxman, 2022). Additionally, both administrative support and collegial interactions are crucial, facilitating the adoption of innovative teaching methods and the exchange of best practices.

The impact of teaching experience varies, with extended tenure sometimes fostering resistance to new teaching methods, but accumulated knowledge can also enrich a teacher's effectiveness (Thibaut et al., 2018b; Wright & Waxman, 2022). Resource constraints, particularly time and access to materials, also emerge as significant barriers, restricting the ability to implement desired instructional methods and engage in professional development (Dong et al., 2018; Thibaut et al., 2018b; Wright & Waxman, 2022).

Addressing these contextual factors is essential for developing supportive educational environments that enhance teaching practices and improve educational outcomes in science education. This approach acknowledges the complexity of educational contexts and underscores the need for a multi-faceted support system.

2.3.2 Synthesizing Best Practices for Teaching Data Literacy

For this study, I decided to synthesize frameworks because there is no commonly used framework for teaching data literacy in science. I wanted to have a modern framework that focused on data literacy, as opposed to SEP 4, to compare the approaches to determine if teachers are teaching in a way that is reflective of more comprehensive 21st century ways of working with data. I chose to examine four frameworks, three from data science and one from science education, that came from different grade bands, elementary to higher education, to identify best practices and determine how they could be adapted for science education. I examined the commonalities among the data literacy frameworks reviewed in section 2.2.5 of the literature review, synthesizing a modified version incorporating their shared concepts. Appendix D illustrates the model's components and which frameworks match each. The frameworks in the literature did not use the same terminology, but I identified themes within them to construct this model. This synthesized model enables a comparative analysis between the existing and ideal conceptualizations of teaching data literacy. By analyzing these frameworks, I identified commonalities and integrated them into a unified model. The synthesized best practices model (SBPM) aligns with the distinctive needs and requirements of teaching data literacy within the 6-12 science education. In contrast to the first part of the conceptual framework, which focuses on examining contextual factors within their practices, this aspect identifies a best practices model for data literacy in 6-12 science education based on current research. The insights derived from this synthesis will serve as a foundational basis for comparative analysis to answer research question 3b, allowing for the comparison of existing practices with the envisioned practices, thereby providing an evaluation of the alignment between current practices and the idealized framework for teaching data literacy in science education.

2.4 The Study

2.4.1 Research Design

The first objective of this study is to identify the approaches that 6-12 science teachers take in analyzing and interpreting data to determine what teachers are doing concerning data literacy in their classrooms. The goal is to identify practices in their teaching methodologies and contextual factors that aid in these approaches, such as curriculum and resources. Then, these will be used to determine how their approaches align with the goals outlined in *A Framework of K-12 Science Education* (National Research Council, 2012) and compare these with the best practices using the Synthesized Best Practices Model (SBPM), as identified in Appendix D, for teaching data literacy. I chose these goals over the Next Generation Science Standards (NGSS) (Lead States, 2013) because they accommodate the varied grade levels of the participating teachers, offering a comprehensive assessment of how their data literacy approaches align with foundational science standards. Through these objectives, the study aims to contribute to a deeper understanding of how data literacy is taught in secondary science education.

This investigation utilizes a general qualitative methodology (Ezzy, 2002; Silverman, 1999), adhering to the methodological principles delineated by Creswell and Báez (2020) and guided by the design framework proposed by Hazzan and Nutov (2014). The selected methodology uses inductive and deductive coding. Inductive coding is well-suited for this study as it involves the identification of codes and categories directly from the data without relying on preconceived frameworks or theories (Thomas, 2006). This coding approach resonates effectively with the objectives of the current investigation, facilitating a comprehensive exploration of teacher practices (Thomas, 2006). By amplifying the voices of teachers, this method enables a thorough examination of their instructional approaches concerning data

analysis and interpretation, thereby affording the flexibility to uncover emergent themes that may be pertinent to the research question. Deductive codes derived from the SBPM were used to compare teacher responses to best practices for teaching data literacy. This approach was chosen to determine how their methods align with recognized effective teaching strategies and to identify if they are already implementing practices deemed effective by research.

I chose focus groups as the most suitable method for data collection in this study, and several vital rationales underpin this decision (Arksey & Knight, 1999). Firstly, the focus group method facilitated participant comfort during discussions, creating a conducive environment for open dialogue (Arksey & Knight, 1999; Rabiee, 2004). A sense of camaraderie within the cohort was created by discussing the grade levels everyone taught and sharing brief insights into respective content areas. As a fellow teacher, I actively participated in this dialogue, sharing my experiences to underscore my role as a teacher and researcher, promoting mutual understanding among participants (Wilkinson, 1998). Using groups was indispensable for cultivating diverse discourse surrounding analyzing and interpreting data (Arksey & Knight, 1999; Rabiee, 2004; Wilkinson, 1998). This methodology facilitated the incorporation of insights from educators with diverse backgrounds while also enabling the exploration of group dynamics. Consequently, participants could leverage each other's experiences, identify commonalities, and contribute to a more enriched dataset than individual interviews. Additionally, focus groups can effectively gather insights, opinions, and interpretations within a social context, specifically within the classroom (Busetto et al., 2020; Wilkinson, 1998), which is ideal for this study. This approach aligned with the study's goals of promoting participant comfort, fostering diverse perspectives, and identifying commonalities of experience and the social context of the participants. The focus

group followed a semi-structured format (Busetto et al., 2020; Creswell & Báez, 2020; Wilkinson, 1998) with the following prompting questions-

- How do you define data literacy?
- Can you describe how you use SEP 4 in your classroom?
- Can you tell me how comfortable you feel using data in your classroom?
- Can you give me an example of a lesson or activity you did that included working with data?

I chose this interview style because it closely relates to oral history types of interviews, aiming for participants to speak about their own experiences and practice in their own words (Arksey & Knight, 1999). These questions guided the discussion without imposing constraints, fostering an open and expansive discourse. While I couldn't be sure teachers would consistently adhere to the designated topic, I was prepared to use prompts and questions to guide the discussion. I remained open to the possibility of the research taking an alternative approach to exploring what teachers deemed significant within this topic (Busetto et al., 2020; Roulston, 2013; Wilkinson, 1998).

The second objective of this study involved a comparative analysis where these pedagogical approaches are evaluated against established best practices in the field. Specifically, this comparison focused on how these teaching methods align with the practice's six goals outlined in *A Framework for K-12 Science Education* (National Research Council, 2012) and the Synthesized Best Practices Model (SBPM). This objective aims to understand teaching practices for analyzing and interpreting data comprehensively. To address the alignment with the goals and best practices, I juxtaposed the data derived from the inductive coding phase with the deductive codes outlined in the SBPM. This comparative analysis highlighted how teachers'

practices aligned with established goals and best practices and offered insights into potential areas for further research to enhance the relevance of data literacy instruction in the 21st-century educational landscape. Using inductive and deductive coding methodologies provided a comprehensive perspective, allowing for an examination of alignment with predefined practices while also permitting the exploration of novel insights emerging from teacher narratives.

The findings of this study are presented through a thematic analysis of the focus group narratives, addressing the following central research questions:

1. How do the participants define data literacy?
2. What approaches do teachers use to teach data analysis and interpretation?
3. How do these approaches align with recognized best practices in the field?
 - a: Practice 4: Analyzing and Interpreting Data (National Research Council, 2012)
 - b: Synthesized Best Practices Model (SBPM)

This analysis sought to contribute to the academic discourse by providing concrete examples of pedagogical practices in data analysis and interpretation and how they compare to established best practices.

2.4.1.1. Participants. This study was approved by the university's review ethics board. It involved seven science educators recruited through a hybrid Convenience-Purposeful sampling strategy (Creswell & Báez, 2020). The recruitment utilized a survey distributed to 6-12 grade educators across six East Tennessee school districts and various online Facebook groups. The survey focused on educators' familiarity with data literacy and Science and Engineering Practice 4. Interested participants indicated their willingness to join a 1-hour focus group by providing contact details at the end of the survey. The participant demographics can be seen in Appendix E.

The recruitment utilized a survey distributed across six school districts and various online Facebook groups, focusing on educators' familiarity with data literacy and Science and Engineering Practice 4. Interested participants indicated their willingness to join a 1-hour focus group by providing contact details at the end of the survey. Six of the participants taught in public schools and one in a private school. Four of the participants were middle grades teachers and three were high school. The participant demographics can be seen in Appendix E.

2.5. Data Collection and Analysis

The primary data source for this analysis was a series of semi-structured focus groups of participants teaching science in one or more grades 6-12. I sought to allow participants to share their experiences and clarify their specific approaches to teaching, analyzing, and interpreting data. The broad and open-ended questions allowed teachers to construct their understanding of this and share their experiences. It also allowed teachers to build off each other's answers to get a complete picture of how they teach and an opportunity to identify support and commonalities with the other participants.

Based on availability, participants were grouped into small focus groups of 2-3 to foster in-depth discussions. Focus group sessions were held via Zoom, recorded (audio and visual), and captioned to create transcripts for analysis, ensuring broad accessibility and facilitating diverse geographical participation. In each focus group, at least two participants were present, except for the final session, which only one attended.

I designed the focus group to start broadly and become progressively more specific. I began with the broad question, "How do you define data literacy?" and then asked about their approaches to teaching, analyzing, and interpreting data. I conducted the focus groups. It is important to note that I have had previous interactions with four of the participants through their

jobs as an educator. These interactions were professional and included professional development and professional organization meetings. It is also important to note that the author has 17 years of science teaching in middle grades. This experience influenced the questions for the interview in addition to the inductive coding for the analysis. I conducted three focus groups on Zoom, then transcribed and cleaned the transcripts (1 hour each).

2.5.1 Data Analysis

For RQ 1: How do the participants define data literacy? I separated the responses, from each teacher to this question and put them into the codebook. The codebook with examples can be found in Appendix G. I highlighted text that was common across all the responses. From there, I identified standard terms and concepts from their answers. This synthesized definition will answer this research question.

For RQ 2: What approaches do teachers use to teach data analysis and interpretation? I performed a thematic analysis using inductive coding to explore the answer. This analytical approach was consistent with my background knowledge and expertise as a science educator, indicating that my personal experiences influenced my data interpretation (Boyatzis, 1998; Harper & Thompson, 2011). For analysis, I reviewed the recordings of each focus group session, cleaned the transcripts, and cross-referenced the transcripts to ensure an accurate representation of participants' responses (Boyatzis, 1998). The process (Appendix F), I reviewed the video along with the transcription and engaged in the observation and documentation of nonverbal cues displayed by the participants. Initially, these cues were incorporated into the transcript and distinguished by highlighting them in orange to signify their nonverbal nature. The cue analysis was necessary for discerning nonverbal indicators of agreement or dissatisfaction. For instance,

an individual might have refrained from agreeing verbally but could convey concurrence by nodding.

Following this, I analyzed the transcripts over several iterations (Boyatzis, 1998; Harper & Thompson, 2011) to identify teachers' approaches to teaching, analyzing, and interpreting data. During multiple iterations of the transcript review, I highlighted pertinent elements (inductive coding) in the transcript document related to the strategies employed in teaching data literacy. Subsequently, I categorized similar items by assigning distinct colors. I highlighted with different colors to quickly identify critical passages during subsequent analysis.

The first round of coding generated 19 unique codes from the three focus group transcripts. During the second and third rounds of coding, the codes were refined into six codes. These codes were added to the codebook with examples from the participants' responses. The codebook with examples can be found in Appendix G. Analysis of these codes revealed that the most prevalent codes included relevance, basic skills, data activities, curated experiences, data literacy, and technology. From here, I identified the patterns that occurred among participant responses and used this to identify three major themes (Boyatzis, 1998; Harper & Thompson, 2011). The themes and their definitions are listed in Appendix I.

For research question 3a, the codes and responses derived from the analysis of the first and second research question were utilized. These were compared against the goals outlined in the *A Framework for K-12 Science Education* (National Research Council, 2012). This comparison aimed to assess the alignment of participants' practices with the goals. The goals were placed in a separate tab in the codebook spreadsheet, and examples of the codes from research question 2 were compared against the goals to determine if they aligned. The codebook with examples can be found in Appendix G. These alignments were used to answer question 3a.

For research question 3b, the codes and responses from the analysis of the second research question were used to compare against the Synthesized Best Practices Model (SBPM). This comparison was conducted to identify alignments of teacher responses. These were outlined in a separate tab in the codebook spreadsheet. The codebook with examples can be found in Appendix G. This step was done to determine how the participants' methodologies align with the best practices in the model. It resulted in identifying connections to best practices from the literature. These alignments were used to answer research question 3b.

By comparing the responses to goals provided by the *Framework for K-12 Science Education* and the SBPM, I can comprehensively understand how current practices measure up to standard and best practices models in the field.

2.6 Findings

2.6.1 RQ 1-Data Literacy Defined

This section addresses the response to RQ 1 to identify how teachers define data literacy. The first question posed to each participant was, "How do you define data literacy?" setting the stage for exploring how these definitions frame their instructional approaches. This exploration facilitates positioning their responses in context and examining the alignment between the teachers' approaches and their definitions of data literacy.

The general consensus of their definitions was the ability to make sense of, analyze, and interpret data. Kelly indicated that being able to read data and making sense of it was necessary by stating:

They can look at any set of data, graph or otherwise, and make sense of it to further their thinking about some topic that I'm trying to explore with them. Data literacy, I guess,

would be them feeling comfortable looking at a data table graph making sense of it, and understanding what they're looking at.

This indicates that making sense of data is an essential aspect of data literacy. Jes, a middle school STEM teacher, mirrored this by also stating that being able to read data and make sense of it. “Data literacy, I would say, means being able to read data, analyze data, understand and comprehend data is being literate.”

These responses align with the definitions provided by the teachers and collectively highlight the necessity for students to interpret and understand data effectively. These definitions frame the teachers' views on data analysis and interpretation, emphasizing the aspects they consider essential for their students' education.

2.6.2. RQ 2- Approaches Used in Teaching Analyzing and Interpreting Data

This section discusses the themes that emerged from the data to address research question 2. Specifically, it outlines teachers' various approaches to teaching data analysis and interpretation.

2.6.2.1 Community-Based Activities and Relevance. A substantial portion of the participants actively engage community resources to enhance the analysis and interpretation of data within their practice. Participants reported that the integration of local resources significantly increased student engagement and interest, thereby enhancing the relevance of the activities. For instance, Jes, a middle school STEM teacher, observed:

The kids are definitely more engaged, obviously, when it is local. Any time I have a unit that talks about something that I feel like can be applied to East Tennessee. I try to change the context to East Tennessee.

A majority (five out of seven) of the participants noted that they deliberately connect data-related activities to their local community contexts, thereby enriching the learning experience. Molly, a 6th-grade science teacher in Idaho who employs the GeoInquiry process, articulated how local contexts influenced her students' inquiry. "The kids' question was, how is climate change affecting the biodiversity of our area where we live, and we were finding data on climate change and Idaho."

Additionally, two participants highlighted their collaborations with local organizations to further enrich their teaching and provide students with access to expert knowledge. Dora, a sixth-grade teacher in East Tennessee, shared her experience integrating community resources into her curriculum. "We're doing the trout in the classroom, and you know, the guys (from the Little River Watershed Association) just live right there in the community, and that's been a really big help." Similarly, Jes described her partnership with local university scientists to enhance her environmental science curriculum. "I've got UT coming out to help us with water quality tests. We're going to look at Macro-invertebrates, and we're going to tie that into a big, essential question."

Two participants highlighted the use of specific local data in their teaching strategies. Val, a high school STEM teacher, integrates local GIS (Geographic Information System) data into her curriculum to ground students' learning in real-world data and connect it to broader, conceptual projects such as urban planning and Mars colonization:

One thing that I do like to do is integrate local data that's already there. So, we take deep dives into KGIS (Geographic Information System), looking at the topography and where lines are drawn for different school zones. I teach them how to look up information about their homes and properties. This year is a new curriculum that we developed ourselves.

We are doing urban planning in the fall, and then we will segue into planning Mars colonization in the spring. So, they go from planning a city on Earth to planning a city on Mars.

Kelly, a high school Biology teacher, also integrates local GIS data into her analyzing and interpreting data activities focused on urban planning:

We've done urban planning experiments, where they build something for our city or our school. They have to make sure that their data supports their research and portray that in their final presentation, which is like a science fair presentation.

These findings demonstrate that integrating local and community resources into educational activities significantly enhances student engagement and the relevance of data analysis and interpretation tasks. The teachers indicated utilizing local data and collaborating with community organizations to contextualize their teaching and make the learning experience more tangible and meaningful for students. Examples include adapting curriculum content to reflect local environmental issues and partnering with local experts to bring real-world expertise into the classroom. They believed this approach deepens students' understanding and helps them connect more profoundly with the material by linking it to familiar, local contexts. Overall, according to the teachers, using community resources in teaching data-related activities greatly increased student interest and engagement.

2.6.2.2 Basic Skills and Curated Experiences. The findings revealed that educators emphasized basic data skills and curated experiences (heavily scaffolded activities) in their pedagogical approaches to teaching data analysis and interpretation. Participants unanimously recognized the importance of foundational skills in data handling, such as graphing, finding, and

collecting data, with a predominant focus (five out of seven participants) on graph analysis and interpretation.

Val, a high school STEM teacher, highlighted the routine use of pre-designed data sets to help students develop proficiency in data manipulation. “We're doing a lot of charts and graph analysis and map analysis and then bringing all those together.” Molly, a middle school science teacher, discussed the importance of familiarizing students with various graphical representations to enhance their understanding and analytical skills:

I like to give them various forms of data that are already out there and visualized so that they can see, for example, the data in map format, the data in bar graph format, or the data in a number like a dot plot or a scatter plot.

Kelly, a high school Biology teacher, also mentioned the integration of graph analysis in her curriculum, especially to make sense of ecological or climate change data. “We're doing a lot of chart and graph analysis and map analysis and then combining all those to make sense of them in some sort of ecological or climate change way.” Fay, a middle school science teacher, expressed a gradual approach to integrating data visualization skills. “I will occasionally give them charts to read or interpret on class work, homework, or quizzes. I have not yet had them make one from a lab, but it's time to do that.” She indicated that while using pre-made data sets for interpretation based on classroom activities, she recognizes the need to advance students' skills by having them create visual representations from their collected data.

These findings demonstrate that educators prioritize foundational data skills and structured learning experiences in data analysis and interpretation. All participants emphasized the necessity for students to acquire basic competencies in aspects of data such as graphing,

finding, and collecting data, particularly in graph analysis. Educators use pre-designed data sets to facilitate skill development in data handling.

2.6.2.3 Integration of Technology in Teaching Data. The use of technology in data gathering and analysis was a prevalent theme among the participants. The majority (five out of seven) reported incorporating technology into their data activities. Notably, three educators utilized the iNaturalist application, leveraging its features for species identification in student projects. It is essential to acknowledge that within this context, iNaturalist was primarily used to apply its built-in graphic and table representations to present findings rather than for direct engagement with raw data.

Furthermore, two participants employed local ArcGIS data as a resource, utilizing these pre-collected datasets to enrich and support student projects. This approach highlights the educational value of integrating geospatial technology to enhance students' analytical skills and project-based learning experiences. Conversely, two participants indicated they did not use digital platforms for their data activities, suggesting diverse methods and resources educators use to teach data literacy. This variation highlights the pedagogical approaches to equip students with data handling and interpretation skills and knowledge.

2.6.3 RQ 2- How the Approaches Align with Recognized Best Practices

This section presents responses to research questions 3a and 3b. The discussion will outline the alignment of teachers' approaches with the goals established in the *Framework for K-12 Science Education* (National Research Council, 2012) and their alignment with the best practices identified in the Synthesized Best Practices Model (SBPM). This analysis shows how these approaches align at two levels- the established standards-based goals for teaching data analysis and interpretation and the ideal best practices for instruction.

2.6.3.1. Comparative Analysis of Participant Responses Relative to Grade 12 Goals for Analyzing and Interpreting Data. This section presents a comparative analysis of participant responses against the grade 12 goals for analyzing and interpreting data, as specified in *A Framework for K-12 Science Education* (National Research Council, 2012). The study reveals that while the six data literacy goals—from systematic data analysis to technological tools—are generally reflected in the educators' teaching methods, suggesting alignment with the framework's objectives, not all educators explicitly addressed each goal. This variability might stem from differences in the grade levels and subjects they teach. Moreover, these goals are benchmarks for students to achieve by the end of grade 12, indicating they are expected to be met throughout K-12 science education.

These findings explore the practical application of critical data literacy practices defined in the *A Framework for K-12 Science Education* (National Research Council, 2012). The educators involved in this study utilized various strategies to engage students with data across multiple contexts. I detail how each educator incorporated specific practices to enhance students' data literacy skills.

In exploring various approaches for teaching data literacy, teachers demonstrated diverse techniques to engage students that connect to the goals from the framework. Kelly has students systematically analyze data by integrating ecological and climate change themes into her instruction by having students conduct thorough analyses of charts, graphs, and maps. She articulates the process as combining various data visualizations to highlight ecological or climate-related phenomena, thereby helping students make connections between disparate data sets. Similarly, Molly emphasizes the importance of collecting a broad range of data types,

including data from the internet, expert interviews, and field data, fostering a comprehensive approach to data gathering and analysis.

Despite the extensive use of digital and graphical tools, the educators' data analysis and interpretation approaches did not use physical models. Also, statistical methods were absent from teacher responses. There were no specific mentions of using statistical methods in the teaching of analyzing and interpreting data. This absence highlights a potential gap in employing models or statistical methods in analyzing and interpreting data, which could further enhance and help students become more proficient at working with data.

Val employs technology extensively in her teaching, particularly advocating the use of spreadsheets in urban studies. She focuses on teaching students to organize data meticulously, which aids in precise analysis and interpretation. This method prepares students to handle 'tidy data,' enhancing their ability to work efficiently with complex datasets. Jes further enriches this technological integration by teaching students to visualize data in multiple formats. She encourages them to view pre-visualized data to grasp different representations such as maps, bar graphs, and dot plots, which is crucial for evaluating the strength and relevance of the data.

Moreover, Molly engages her students in identifying significant patterns in data by linking classroom activities to real-world issues, such as the impact of climate change on local biodiversity. This approach deepens students' understanding of data analysis and contextualizes their learning within pressing global challenges.

These findings suggest that teachers effectively align their data literacy instruction with most goals outlined in the established framework for teaching, analyzing, and interpreting data. Not all goals were addressed in the teachers' responses; however, it is important to note that these goals are set for achievement by the 12th grade. Therefore, it is understandable that not every

one of the six goals would be addressed by each teacher. Additionally, while they did not explicitly state their use of this framework, their teaching methods are likely influenced by standards derived from it. Overall, teachers' approach corresponded well with most of the framework's expectations, indicating a coherent integration of these goals into classroom practices among this group.

2.6.3.2 Connection to the SBPM. This analysis used deductive coding to examine how teachers' approaches aligned with the SBPM, focusing on experiences with ambiguity, engagement in the full data cycle, relevance, scaffolding, and collaboration in data literacy education.

Experience with Ambiguity. The teachers discussed incorporating data activities lacking predefined outcomes to foster critical thinking and problem-solving skills. Molly uses the Geo-Inquiry Process, emphasizing its open-ended nature:

If you're looking at the data that's collected in tree rings about CO₂ storage, and you look back in time about how much is stored in those tree rings, like 100 years ago, even go all the way back to see the Industrial Revolution, and how much is stored. So, there's just a whole different kind of data to read. You really have to learn how to break it apart. Even I was like not 100% sure how to read it. So, we had to do a lot of research to figure that out and contact some experts who could help us.

Jes highlights the importance of interpreting ambiguous data in real-world contexts. “Sometimes, the data doesn't give us a clear answer, and we have to figure out what it's trying to tell us, which is a real challenge for the kids but very real-world.” Val integrates prototype design, where students determine how their data supports their projects:

Prototypes look like lots of different things. They have to portray their data somehow, the data that supports it, in some kind of data that they have come up with in the process of building their prototype.

Overall, these findings indicate an emphasis student-centered learning, where educators leverage the inherent uncertainty of real-world data and design challenges to foster a learning environment rich in critical thinking and practical application. This approach prepares students to deal with ambiguity and encourages them to apply their knowledge innovatively and effectively.

Engaging Students in the Full Data Cycle. Several teachers discussed and emphasized aspects of using the full data cycle with students. Kelly emphasized starting with data collection to teach comprehensive data handling. “We engage the students from the start—collecting and processing, and then they have to present their findings, which teaches them the full cycle of data handling.” Jes described how students develop and follow their own inquiry investigations:

I have my eighth graders do some kind of inquiry where they have to gather their own data. They have to come up with their own simple, testable question and need it to be based on observable data. They have to collect it and then they have to organize it in some way that makes sense of what they're trying to achieve. They have to create some kind of graph.

Molly used the GeoInquiry process to emphasize diverse data sources and visualization techniques:

The kids collect various forms of data—from the Internet, by interviewing experts, and by actual field collection. Over time, we end up with all this information, and then we have to figure out a way to visualize it. I like to give them various forms of data that's already out there that's already visualized so that they can see, for example, data in a map

format, a bar graph, a dot plot, or a scatter plot. This helps them decide how they want to visualize their data for their project.

These responses demonstrate that some educators are committed to providing experiences that allow students to engage in the full data cycle. This approach seems to enhance their analytical skills and foster a deeper understanding of data's practical implications and applications in real-world contexts.

The Role of Relevance in Teaching Data Literacy. Relevance is noted as important in enhancing student engagement, as stated by several participants. Jes discussed using community-specific data. “We use data that directly impacts our community, which makes the students really care about the results.” Val and Molly discussed how local data can increase engagement and inspire deeper investigations, with Val noting. “They really are interested in that local data at KGIS (Geographic Information System).” Molly goes into detail on how relevant local data increases student engagement and motivation:

The kids are just so fascinated, and the kids who don't like to read will still read an article because they're so interested in the topic or they don't like to do math. They'll still gather data and put it into a graph, even though it's just because they're really interested in something that has to do with their home or something close to them.

These insights demonstrate that integrating local and authentic data makes learning more relevant and enhances student engagement by linking educational content with real-world applications. Overall, the teachers' strategies emphasize the profound impact of relevance in data literacy education, fostering a more engaging and meaningful learning experience for students.

Scaffolding Data Literacy Activities. Many of the teachers discussed different types of scaffolding and how it is essential for aligning activities with students' proficiency levels. They

design incrementally challenging tasks to build students' capabilities. Kelly details how she scaffolds instruction. "Each step of our data projects is designed to build on the last, ensuring all students are progressing within their capabilities." Molly and Val focus on setting clear parameters for data collection and deconstructing complex visualizations to enhance understanding. Molly notes on setting clear parameters:

I'm the one who helps them decide how it's going to be collected. We definitely have to set up the table or set up the parameters of what they're collecting. Then, once that's set up, you do an example, let's do it together kind of thing. After that, we just kind of let them go with it.

Val addressed the need to deconstruct complex data visualizations into simpler data sets to enhance student's understanding of how data is transformed into graphical representations. This reverse-engineering approach aids in comprehending data flow and visualization:

Let's take this graph and let's get the data out of it and put it back into a spreadsheet and see where it came from. I think maybe I need to deconstruct some graphs and see if we can take it back to the data and see how that looks.

Fay takes a more structured approach to scaffolding data activities, indicating that she believes it is time to take it a step further by including authentic data and larger sets:

So, I will give them charts occasionally to read or interpret on class work, homework, or on a quiz. I have not yet had them make one from a lab, but it's time to do that. Typically, we don't have a whole heap of data points, but we need to do that in graph data because that's a key skill. They need to know.

These findings highlight that several educators implement structured scaffolding strategies to teach data literacy, effectively building upon students' existing skills within their

ideal learning zone. Four participants detailed their systematic scaffolding methods, involving sequentially ordered tasks that build on each other to ensure individual student progression. They emphasized key techniques such as setting clear parameters for data collection, providing tangible examples, and simplifying complex data visualizations to enhance understanding. These approaches can not only aid in developing data literacy skills but also boost students' independence in handling data.

Collaboration. Several participants indicated that collaboration is important for deepening understanding and engagement. Kelly states the importance of group discussions. “Collaboration is key in our projects; they need to talk through their data interpretations and analyses to reach a consensus.” Molly describes fieldwork that fosters collective data analysis:

I take my kids out to the field, and then all the kids are working on a different area of wherever we are. And then we could just take all of that data and collectively put it together in one space.

Additionally, Molly discusses how structured group discussions play a crucial role in enhancing students' communication skills. “The students work in groups to discuss their data findings, which helps them learn from each other and improve their communication skills.” These three participants indicate that collaboration is important in teaching data literacy. It enables students to work together effectively, think critically, and communicate their findings clearly.

These connections to the SBPM illustrate that some of the educators' methods align with best practices in data literacy education. They highlight the importance of handling ambiguity, engaging students throughout the data cycle, and enhancing learning through relevant,

scaffolded, and collaborative activities. These strategies prepare students to manage complexity and apply their skills innovatively in real-world situations.

2.7 Discussion and Conclusions

In examining how teachers define data literacy and their approaches to teaching data analysis and interpretation, a broad consensus exists on the definition—examining and making sense of data. This aligns with descriptions in academic frameworks and literature on data analysis in science education (National Research Council, 2012; Schwarz et al., 2017). However, the varied teaching approaches suggest a broader and more interpretative nature to the practice, consistent with the literature calling for a detailed framework for teaching data analysis and interpretation (Pangrazios & Sefton-Green, 2019; Wolff et al., 2019).

The study's findings highlight the similarities in teaching data literacy within 6-12 science education, reflecting the dynamic and contextual nature of educational practices (Cobb et al., 2003; Jones & Carter, 2007; Lumpe et al., 2000; Luft et al., 2022). Teachers tailor their approaches to meet broader educational goals, the contextual factors in which they are influenced, and the specific needs of their students. The context significantly influences instructional methods, with community-centered activities enhancing relevance and engagement (Schwarz et al., 2017). This experiential learning approach uses local resources to make abstract data concepts tangible and relatable (Smith et al., 2019).

Basic data skills and curated experiences are crucial. All of the teachers emphasized basic competencies in data handling, such as graphing, finding, and collecting data, with a focus on graph analysis and interpretation (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Srikant & Aggarwal, 2017). Scaffolding is widely used to support learning, but the removal of supports to foster autonomy in data work is often overlooked.

Exposing students to diverse data visualizations broadens their analytical skills and prepares them for complex scientific data (Bowen & Bartley, 2014). Several of the teachers indicated that this was an important strategy in teaching about data. Integrating technology is critical for improving learning outcomes, enabling students to engage effectively with various stages of the data cycle (National Research Council, 2012; Rosenberg et al., 2022). Teachers indicated this by describing the tools they use, like iNaturalist and ArcGIS, to support data analysis and enhance learning experiences. However, not all of the teachers used digital platforms and chose pen-and-paper activities instead.

2.7.1 Connection to Framework (2012)

Responses to research question 3 show that many teachers align their practices with some of the data analysis and interpretation objectives for grade 12 outlined in A Framework for K-12 Science Education (National Research Council, 2012). However, some goals, such as using physical models or statistical methods, are not addressed.

Teachers incorporate essential data literacy practices in various ways, including systematic data analysis and comprehensive data collection. They use technological tools for data organization and visualization, engaging students with real-world data to contextualize their learning within global challenges. This diversity of pedagogical strategies demonstrates the need for flexible frameworks to accommodate varying levels of technological integration and cater to diverse learning environments (Rosenberg et al., 2022).

2.7.2 Connection to the SBPM

The teachers' approaches in data literacy education focus on experiences with ambiguity, engagement in the full data cycle, relevance, scaffolding, and collaboration. Some incorporated activities with no predefined outcomes to foster critical thinking and problem-solving, using

processes like Geo-Inquiry and prototype design to interpret ambiguous data, highlighting its importance in real-world contexts (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015).

The teachers engaged students in the full data cycle, from collection to visualization, often through inquiry-based learning where students develop and follow their investigations (Bowen & Bartley, 2014). They emphasized relevance by linking data activities with real-world applications, using community-specific data to enhance engagement and motivation (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Maybee & Zilinski, 2015; Srikant & Aggarwal, 2017).

Scaffolding is identified as essential, with teachers designing sequential tasks that build on each other, aligning activities with students' proficiency levels (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Srikant & Aggarwal, 2017). This approach ensures optimal learning by making new knowledge accessible with appropriate support (Bowen & Bartley, 2014).

Collaboration is identified as crucial for deepening understanding and engagement. Teachers used collaborative projects and group discussions to enhance communication skills and promote shared learning experiences (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Maybee & Zilinski, 2015; Srikant & Aggarwal, 2017). This fosters teamwork and cultivates vital social skills, supporting research that values cooperative learning (Johnson & Johnson, 2005). Integrating technology is highlighted as critical for improving learning outcomes. Tools like iNaturalist and ArcGIS facilitate comprehensive engagement with the data cycle, helping manage logistical challenges and enhance interaction with complex data sets (National Research Council, 2012; Rosenberg et al., 2022). This suggests an amendment needs to be made to the Synthesized Best Practices Model (SBPM) to emphasize technology, aligning with its role in

teaching data analysis and interpretation. Currently, the SBPM does not emphasize technology. Yet, it is a prominent goal within the *Framework* (National Research Council, 2012), and teachers have highlighted its role in their approaches to teaching data analysis and interpretation. Accordingly, integrating technology into the SBPM will make it more relevant to teaching and analyzing data. The update is reflected in Appendix D.

These findings show that educators' methods align with best practices in data literacy education, emphasizing handling ambiguity, engaging students throughout the data cycle, and enhancing learning through relevant, scaffolded, and collaborative activities. These strategies prepare students to manage complexity and apply their skills innovatively in real-world situations.

It is vital to continue exploring innovative ways to integrate emerging technological tools to enhance data visualization and analysis capabilities. Updating curricula and providing ongoing training for teachers will be essential to maintaining alignment with industry developments and academic standards.

2.7.3 Conclusion and Limitations

This study faces several limitations that could influence the validity and applicability of its findings. Firstly, the relatively small sample size and potential selection bias may limit the generalizability of the results, as the participants may not represent a diverse range of educational settings or teaching practices. Participants may have chosen to participate in the study because they already possess a keen interest in analyzing and interpreting data.

Additionally, relying on self-reported data from educators could introduce biases, as participants may present their practices in an overly favorable light or not recall all details accurately. Another significant limitation is the absence of direct evaluation of student outcomes;

without this data, it is challenging to ascertain the impact of the described teaching strategies on students' data literacy and academic performance. Addressing these limitations in future research could provide a more comprehensive understanding of the efficacy and applicability of data literacy teaching practices.

In conclusion, this study illuminates how secondary science educators approach data literacy in their practices, emphasizing the practical understanding, analysis, and communication of data. The integration of local resources, the strategic use of scaffolding, and the fostering of collaborative environments underscore data literacy's practical and innovative applications. They believed these practices deepened their students' understanding and prepared them to apply their knowledge across various contexts. Future educational strategies should enhance these components to ensure that data literacy education remains responsive and relevant to academic and industry standards. Emphasis should be placed on relevance by incorporating local resources into teaching data literacy. Collaboration should continue to be a foundational element, and scaffolding should be designed to gradually remove supports, allowing students to develop independence. Each grade level should build a strong foundation for the next, ensuring continuity in skill development.

What was missing in these approaches was a higher level of experiencing ambiguity and enabling students to work more independently with data. Providing opportunities for students to engage with ambiguous data and develop comfort and competence in comprehensive data handling is crucial for their growth and readiness to tackle complex real-world problems.

Chapter 3. Exploring Data Literacy Challenges in Secondary Science Education

3.1 Introduction

In K-12 science education, Science and Engineering Practice 4 (SEP 4), which focuses on analyzing and interpreting data, has received relatively less research attention compared to other practices such as modeling (Lehrer & Schauble, 2004; Passmore & Stewart, 2002; Schwarz et al., 2009) and argumentation (Berland & McNeill, 2010; Grooms et al., 2015). This study explores the challenges teachers face in teaching, analyzing and interpreting data. The Next Generation Science Standards (NGSS) (Lead States, 2013) emphasize SEP 4's role in science education but lacks specific instructional methods, focusing mainly on basic skills and competencies (Bowen & Bartley, 2020).

In addition, current standards-based models for teaching, analyzing and interpreting data do not always adequately reflect the practical ways working scientists engage with data (Bowen & Bartley, 2020), leading to variations in implementation and a significant gap in the standardization of curriculum and textbook development (Morris et al., 2015; Pangrazios & Sefton-Green, 2019). Previous studies by Banilower et al. (2018) and Rosenberg et al. (2022) have provided insights into how much teachers engage with SEP 4 and the resources they utilize. However, these studies fall short of thoroughly exploring the specific actions, methodologies, and challenges teachers face in this domain. Data analysis and interpretation comprise about 11% of NGSS Performance Expectations, underscoring their importance across all STEM education levels and subjects (Kastens, 2014). Yet, the lack of clear instructional guidance remains challenging for educators, necessitating a deeper investigation into how data literacy is taught and implemented in classrooms. This study builds on existing literature by using focus groups to collect detailed information from teachers about the challenges they face when teaching data

analysis and interpretation. Through this research, I aim to provide insights into the contextual factors, such as time, resources, and training, that prevent teachers from effectively teaching, analyzing, and interpreting data. The findings can guide the development of effective supports for integrating data literacy into curriculum and professional development, ensuring that educational practices are aligned with the demands of 21st-century data.

3.2 Literature Review

This literature review examines the theoretical and empirical research on data analysis and interpretation in science education, focusing on Science and Engineering Practice 4 (SEP 4) as outlined in the Next Generation Science Standards (NGSS). It explores current conceptualizations of these practices within science education and evaluates data literacy as represented in scholarly literature. The review aims to provide a comprehensive overview of how analyzing and interpreting data is conceptualized, highlighting gaps in our understanding of big, authentic, and messy data and identifying areas needing further investigation to enhance data analysis capabilities in science education.

3.2.1 What is Data Literacy?

Redefining data literacy within a 21st-century science education is crucial. As data becomes increasingly integrated into everyday life, students need to develop the proficiency necessary to navigate this data-rich environment effectively (D'Ignazio & Bhargava, 2015). Modern working with data involves handling messy data, working with authentic datasets, understanding data security and storage, and recognizing data's potential and limitations (Bowen & Bartley, 2020; D'Ignazio, 2022; Kjellvik & Schultheis, 2019; Rosenberg et al., 2020; Schultheis & Kjellvik, 2020) and yet these are not a focus of Science and Engineering Practice 4. It is essential to include the integration of real-world, complex data scenarios in classrooms to

mirror the authentic challenges faced in scientific and professional domains, enhancing the educational relevance and application of data literacy (Maybee & Zilinski, 2015; D'Ignazio & Bhargava, 2015; Grimshaw, 2015).

The evolving nature of data use in society necessitates a modern, standardized framework for teaching data literacy in K-12 education, preparing students for data-driven decision-making (Pangrazios & Sefton-Green, 2019; Wolff et al., 2019). Although NGSS provides a basic framework for teaching, analyzing, and interpreting data (NGSS Lead States, 2013), it lacks detail on advanced data handling techniques such as data cleaning and the full data cycle (Bowen & Bartley, 2020). Integrating data should be highlighted because of its vital role in scientific inquiry and broad educational practices (National Research Council, 2012; Schwarz et al., 2017). This integration calls for an expansive approach beyond fundamental skills, encompassing critical evaluation, ethical considerations, and diverse applications (Bowen & Bartley, 2020; D'Ignazio, 2022).

3.2.3 An Examination of NGSS Practice 4- Analyzing and Interpreting Data

The NGSS sets performance expectations across grade levels for Science and Engineering Practice 4 (SEP 4) - Analyzing and Interpreting Data (NGSS Lead States, 2013). This framework progressively builds data analysis skills from elementary through high school, detailing what students should achieve at each grade band, and is based on *A Framework for K-12 Science Education* (National Research Council, 2012). Starting in grades K-2, students begin by collecting, recording, and sharing observations through various methods such as pictures, drawings, and writings, using these to describe patterns and relationships to address scientific questions. By grades 3-5, they employ quantitative methods for data collection and start

representing data in graphical displays like bar graphs and pie charts to analyze and interpret phenomena (Lead States, 2012).

As students advance to grades 6-8, the focus shifts to more formal quantitative analyses, distinguishing between correlation and causation and using basic statistical techniques to enhance precision and accuracy. By high school, they engage in detailed statistical analysis using advanced tools and models to make scientific claims and evaluate the impact of new data on existing models. This curriculum emphasizes proficiency in utilizing digital tools and various data representation methods, including spreadsheets, databases, and graphical displays, to convey data effectively (Lead States, 2012).

However, despite this structured approach, SEP 4 faces challenges that hinder comprehensive data literacy education. It primarily focuses on basic skills, failing to expose students to the complexities of big data, messy data, and authentic data sets, which could foster a more profound understanding of data's real-world applications (Bhargava et al., 2015). The current standards often overlook vital aspects of working with data, such as exploring the purpose of data, its potential manipulation, and the differences between quantitative and qualitative data. This oversight contributes to a deterministic view of data presentation and lacks a probabilistic perspective (Bowen & Bartley, 2014). Consequently, there is a pressing need for a more holistic data literacy framework that transcends basic skill competencies and fosters critical thinking and discernment in data analysis. This enhanced approach, as advocated by D'Ignazio (2022), should equip students with the skills to critically question and evaluate data, aligning with the evolving demands of a data-driven world. Doing so ensures that education in data analysis remains grounded in current standards and practices as outlined by the NGSS Lead States (2012), effectively preparing students for the complexities of modern data environments

3.2.4 Data Literacy and SEP 4

This section highlights two key scholarly works exploring data analysis and interpretation as delineated in Science and Engineering Practice 4 (SEP 4) of the Next Generation Science Standards (NGSS). The goal is to deepen the understanding of how SEP 4 is represented in the literature and to critique existing pedagogical methodologies.

Schwarz et al. (2017) emphasized the critical role of SEP 4 in Science and Engineering, particularly how it integrates with other practices and the importance of collaboration. This work discussed key elements essential for effective data engagement, such as using Venn Diagrams, GIS images, and mathematical thinking. However, while the book acknowledged the use of varied data sources and the importance of contextually relevant information, did not extensively cover critical data literacy skills or offer detailed strategies for comprehensive data analysis and interpretation.

On the other hand, Bowen and Bartley (2014) provided a focused examination of data literacy within SEP 4, offering specific strategies for educators to foster a deeper and more critical engagement with data. Their critique of SEP 4 centers on the conventional teaching approaches in science classrooms, which often involve sanitized or replicated data, thereby limiting students' exposure to the full data cycle experienced by field scientists. They argued that this practice does not adequately prepare students for engaging with authentic scientific data and tends to portray data in a deterministic rather than probabilistic manner. This discrepancy shows a significant gap between classroom learning and real-world scientific practices, impacting students' understanding of scientific investigations.

Together, these works advocated for a shift towards more probabilistic thinking in science education, aligning teaching practices with modern data analysis approaches. This

review contextualizes the instructional resources for SEP 4 and highlights the need for pedagogical advancements that embrace the complexities of real-world data engagement.

3.2.5 Four Frameworks for Teaching Data Literacy

This literature review synthesizes four frameworks that shape contemporary approaches to teaching data literacy, emphasizing relevance and empowerment for learners across various educational contexts. The frameworks represent how data literacy should be taught. The Open Inquiry approach, advocated by Bowen & Bartley (2014, 2020) and supported by Dimopoulos & Karamanidou (2013), promotes a shift from deterministic to probabilistic thinking in K-12 education by engaging students with the full data cycle—collection, analysis, and interpretation. This framework encourages investigation into ill-defined problems, fostering personalized learning and collaboration within a community of learners. The Four Core Pillars of Data Literacy, as outlined by D'Ignazio & Bhargava (2015) and supported by Raffaghelli et al. (2020), redefine data literacy to include agility and adaptability, focusing on empowering individuals to navigate their social ecosystems effectively. This framework emphasizes developing adaptable skills that transcend specific technologies and fosters a holistic understanding of data literacy as essential for societal inclusion and active citizenship. The Data-Informed Learning framework by Maybee & Zielinski (2015) and further discussed by Koltay (2016b) and Raffaghelli (2020), integrates data literacy within disciplinary subjects by building on learners' prior experiences, making it contextually relevant and a part of everyday learning. Lastly, the Workflow Through Data Science Tutorial Design by Srikant & Aggarwal (2017), supported by Miao et al. (2022), involves middle-grade students working directly with the data cycle with relatable data sets, emphasizing hands-on collection and practical application. Together, these frameworks advocate for a shift in educational paradigms where data literacy is not only about technical skills but also

about engaging learners in ways that are meaningful to their personal lives and community involvement, thus establishing data literacy as a critical component of modern science education and necessary for developing informed, capable citizens in a data-driven world.

In conclusion, while the literature provides a foundational understanding of data literacy within the context of 21st-century science education, notable gaps persist. These include a lack of detailed strategies for teaching advanced data handling techniques such as data cleaning and managing the full data cycle. Additionally, the literature falls short in addressing the integration of ethical considerations and diverse contextual applications into data literacy education. Current educational frameworks, such as the NGSS (Lead States, 2013), provide a basic structure but do not offer sufficient guidance on handling real-world, messy data or fostering critical evaluation of data sources. This oversight limits students' capacity to engage effectively with authentic datasets and to develop a nuanced understanding of the role of data in scientific inquiry and decision-making. Addressing these gaps is crucial for developing more comprehensive data literacy education that covers technical skills and emphasizes critical thinking, ethical reasoning, and adaptability across various educational and real-life contexts.

3.3 Conceptual Framework

3.3.1 The Role of Contextual Factors in Teacher Practice

In science education, understanding the impact of contextual factors on teaching practices is crucial for fostering effective instructional strategies. This section highlights the interaction of various elements within educational environments and their influence on teacher practice.

Factors such as adherence to state standards, the backgrounds of teachers and students, and the available instructional materials shape the educational landscape (Cobb et al., 2003; Jones &

Carter, 2007; Lumpe et al., 2000; Luft et al., 2022). These factors can either enhance or hinder a teacher's ability to deliver effective instruction.

Teacher self-efficacy is a significant influence (Bandura et al., 1999), where teachers' confidence in their abilities directly affects their instructional practices (Dong et al., 2018; Thibaut et al., 2018b; Wright & Waxman, 2022). A strong sense of self-efficacy, supported by robust content knowledge often acquired through hands-on training, is vital for positive outcomes. Furthermore, the importance of professional development is underscored, with effective programs markedly improving how teachers implement classroom instruction (Dong et al., 2018; Thibaut et al., 2018b; Wright & Waxman, 2022). Additionally, both administrative support and collegial interactions are pivotal, facilitating the adoption of innovative teaching methods and the exchange of best practices (Dong et al., 2018; Thibaut et al., 2018b; Wright & Waxman, 2022).

The impact of teaching experience varies, with extended tenure sometimes fostering resistance to new teaching methods. However, accumulated knowledge can also enrich a teacher's effectiveness (Thibaut et al., 2018b; Wright & Waxman, 2022).

Resource constraints, particularly time and access to materials, also emerge as significant barriers, restricting the ability to implement desired instructional methods and engage in professional development (Dong et al., 2018; Thibaut et al., 2018b; Wright & Waxman, 2022).

Addressing these contextual factors is essential for developing supportive educational environments that enhance teaching practices and improve educational outcomes in science education. This approach acknowledges the complexity of educational contexts and emphasizes the need for a multi-faceted support system.

3.4 The Study

3.4.1 Research Design

This study identifies challenges and obstacles teachers face in teaching, analyzing, and interpreting data. The goal is to determine what factors inhibit their ability to teach data literacy in a relevant and meaningful way. This study aims to contribute to a deeper understanding of teachers' challenges to aid in future research to identify supports for teaching data literacy. This study employs the same data sources and research design presented in Chapter 2. However, it shifts the focus to exploring challenges and obstacles in teaching analyzing, and interpreting data.

This study utilized broad qualitative research methodologies (Ezzy, 2002; Silverman, 1999) to investigate teachers' challenges in their practice regarding analyzing and interpreting data. Although previous research, such as surveys by Banilower et al. (2018) on teacher practice with SEP 4 and Rosenberg et al. (2022) on tool utilization, has shed light on various aspects of teaching practices, there remains a gap in the literature regarding a comprehensive examination of the specific challenges teachers face. To bridge this gap, the qualitative approach employed in this study was instrumental in facilitating a thorough exploration of these challenges, enabling a deeper understanding of teachers' obstacles.

I chose focus groups as the most suitable method for data collection in this study, and several key rationales underpin this decision (Arksey & Knight, 1999). Firstly, the focus group method facilitated participant comfort during discussions, creating a conducive environment for open dialogue (Arksey & Knight, 1999; Rabiee, 2004). To foster a sense of camaraderie within the cohort, the initial phase involved a discussion on the grade levels we instruct and brief insights into our respective content areas. I actively participated in this dialogue, sharing my

experiences to emphasize my role as a teacher and researcher, promoting mutual understanding among participants (Wilkinson, 1998). Focus groups were used to cultivate diverse discourse surrounding the challenges faced in teaching, analyzing and interpreting data (Arksey & Knight, 1999; Rabiee, 2004; Wilkinson, 1998). This methodology facilitated the incorporation of insights from educators with diverse backgrounds while also enabling the exploration of group dynamics. Consequently, participants could leverage each other's experiences, identify commonalities, and contribute to a more enriched dataset than individual interviews. Additionally, focus groups effectively gathered insights, opinions, and interpretations within a social context, specifically within the classroom (Busetto et al., 2020; Wilkinson, 1998).

This approach aligns with the study's goals of promoting participant comfort, fostering diverse perspectives, and identifying commonalities of experience and the social context of the participants. The focus group followed a semi-structured format (Busetto et al., 2020; Creswell & Báez, 2020; Wilkinson, 1998) with the following prompting questions-

- Can you describe how you use SEP 4 in your classroom?
- What obstacles do you face in teaching data in your classroom?
- Can you give me an example of a time when you didn't feel like you could effectively teach working with data?
- Can you tell me how comfortable you feel using data in your classroom?

This interview style was chosen because it closely relates to oral history types of interviews, as the goal was for the participants to speak about their own experiences and practice in their own words (Arksey & Knight, 1999). These questions guided the discussion without imposing constraints, fostering an open and expansive discourse. Although I couldn't be sure teachers would consistently adhere to the designated topic, I was prepared to use prompts and questions to

guide the discussion. I remained open to the possibility of the research taking an alternative approach to exploring what teachers deem significant within this topic (Busetto et al., 2020; Roulston, 2013; Wilkinson, 1998). I articulated the results of this research through a detailed thematic analysis of the narratives obtained from the focus groups. This analysis addresses the central research question:

What contextual factors impede effective data literacy teacher practice in 6-12 science education classrooms?

By analyzing the data gathered, this study aims to enrich the academic discourse by providing concrete, real-world examples of pedagogical challenges in teaching data analysis and interpretation.

3.4.2 Participants

My university's review ethics board approved this study. Seven science educators were recruited through a hybrid Convenience-Purposeful sampling strategy (Creswell & Báez, 2020). The participant demographics can be seen in Appendix E. Based on availability, participants were grouped into small focus groups of 2-3, with no group exceeding five members to foster more intimate discussions. Recruitment targeted 6-12 grade educators from East Tennessee school districts and online teaching forums.

The recruitment utilized a survey distributed across six school districts and various online Facebook groups, focusing on educators' familiarity with data literacy and Science and Engineering Practice 4. Interested participants indicated their willingness to join a 1-hour focus group by providing contact details at the end of the survey.

3.5 Data Collection and Analysis

The primary data source for this analysis was the transcripts from a series of semi-structured focus groups of participants teaching science in one or more grades 6-12. This source was also used in Chapter 2, albeit with a focus on different questions and responses. I sought to enable participants to share their experiences and articulate their specific challenges and obstacles while teaching data analysis and interpretation. The questions were broad and open-ended, allowing teachers to construct their own understanding of what this means and share their experiences. It also allowed teachers to build off each other's answers to get a full picture of how they teach and an opportunity for them to identify support and commonalities with the other participants. The author conducted the focus groups. It is important to note that the author had previous interactions with four of the participants through their job as an educator. These interactions were of a professional nature including professional developments and professional organization meetings. It is also important to note that the author has 15 years of science teaching in middle grades. This experience influenced the questions for the interview in addition to the inductive coding for the analysis. I conducted three focus groups on Zoom, then transcribed and cleaned the transcripts (1 hour each).

3.5.1 Data Analysis

I conducted a thematic analysis utilizing inductive coding to scrutinize the data. This approach aligned with my background knowledge and expertise as a science educator, which means my personal experiences could influence my data interpretation (Boyatzis, 1998; Harper & Thompson, 2011). For analysis, I reviewed the recordings of each focus group session, cleaned the transcripts, and cross-referenced the transcripts to ensure an accurate representation of participants' responses (Boyatzis, 1998). I reviewed the video along with the transcription and

engaged in the observation and documentation of nonverbal cues displayed by the participants. Initially, these cues were incorporated into the transcript and distinguished by highlighting them in orange to signify their nonverbal nature. Acknowledging nonverbal cues was necessary for discerning nonverbal indicators of agreement or dissatisfaction. For instance, an individual might have refrained from agreeing verbally but could convey concurrence by nodding.

I systematically analyzed the transcripts (Boyatzis, 1998; Harper & Thompson, 2011) to identify teachers' challenges. My analysis process is outlined in Appendix F. During multiple iterations of the transcript review, I highlighted pertinent elements (inductive coding) in the transcript document related to the strategies employed in teaching data literacy. Subsequently, I categorized similar items by assigning distinct colors. I highlighted the key passages in different colors to identify them quickly during the subsequent analysis.

The first round of coding generated 15 unique codes across the three focus group transcripts. During the second and third rounds of coding, the codes were refined into ten codes. These codes and examples can be found in Appendix H. Analysis revealed that the most prevalent codes included professional and preservice learning, resources, curriculum, student experience with data, technology, standards, district policies, time, teacher agency, and student engagement. From here, I identified the patterns among participant responses and used this to identify five major themes- insufficient training, curriculum, time, resource availability, and student prior knowledge and engagement, as defined in Appendix I.

3.6 Findings

In this section, the findings are delineated and illustrated through selected responses from educators, illustrating the key outcomes of the study.

3.6.1 Insufficient Training

Three participants highlighted a significant deficiency in data literacy training across educational levels, emphasizing the need for collaborative efforts to enhance teaching effectiveness in analyzing and interpreting data. Kelly, a high school biology teacher, noted her personal challenges in mastering data literacy, which were not addressed during her formal education:

Data literacy is something that I did not receive in school. Not even in just preservice training but, in my education. I didn't go to a public school. I still don't feel like data literacy was something that I was very proficient at. I think I had to work pretty hard and with colleagues to master that in a way that I could teach it to students.

Her account highlights the widespread absence of focused data literacy training, even at the college level. She remarked:

Even in college, I don't remember a focus on it. There was no 'Here's some data now work with it, you know, extrapolate.' Now, I'm trying to actually do it with freshman in high school, I don't remember getting that in college.

Jes expressed a desire for interdisciplinary collaboration to strengthen data literacy instruction but was skeptical about its feasibility:

It would be awesome to get with the math teachers and science teachers of that grade level and come up with a really good collaboration. I don't see that happening a lot where I'm at.

She also pointed out the limited focus on creating data within training sessions, indicating that most training involved pre-made data, which did not sufficiently prepare teachers for guiding students in data creation. Val indicated she wished she and her colleagues were able to observe

other teachers and how they teach certain skills to strengthen her skills that she didn't get from preservice training or professional development:

If we had more time to observe other teachers. I would like to be able to observe how they teach spreadsheet development. I wish that I had a mentor teacher who could do that on a very simple level.

Molly revealed that her only exposure to data training was in math classes, highlighting a fragmented approach to data literacy that often omits its application in science education. She pursued further professional development on her initiative, driven by her passion for science: "I think what I've developed over time has been through various trainings and programs that I participated in because I just love science." She stated that most of her learning on analyzing and interpreting data was independently:

Just my own learning and science is what background I got (*on analyzing and interpreting data*) when I was before I became a teacher. I think what I've developed over time has been through various trainings and programs that I participated in because I just love science. I've sought out trainings that have to do with that.

The findings from the study highlight a significant gap in data literacy training across educational levels, as reflected in the experiences of several participants. Many of these teachers, including those who have pursued specialized training, reported a lack of focused instruction on data literacy during their formal education, extending even into their college years. This deficiency forced these teachers to seek professional development opportunities independently or work collaboratively with peers to develop the necessary skills to teach data analysis effectively. These teachers frequently mentioned wanting more interdisciplinary collaboration and structured mentorship programs to enhance their data literacy education. However, the feasibility of such

initiatives remains uncertain in many of their schools. Overall, these findings note the absence of comprehensive data literacy training in their teacher education programs, highlighting the need for systemic changes to integrate data skills into the preservice training programs, thereby enhancing the overall quality of teaching and learning in this crucial area.

3.6.2 Curriculum

The discussion about the curriculum highlighted several challenges faced by teachers, particularly in terms of accessibility and effectiveness. This study encompasses participants from diverse educational contexts, including districts, content areas, and grade levels. Such variability may influence their access to and implementation of the curriculum, affecting the teaching and learning processes. This variation illustrates the disparities between schools within a district, region, state, or country and highlights how these differences can impact educational practices and outcomes. Teachers such as Kelly, a high school Biology teacher, Jes, a middle school STEM teacher, and Dora, a middle school science teacher—all working in public schools—were required to develop their curricula as they were not provided one. Kelly found an open source "storylines" curriculum to guide her classes, while Jes and Dora crafted their curriculum from scratch.

On the other hand, educators like Anna, who teaches Biology and Chemistry at a private school, and Fay, Molly, and Val, science and STEM teachers at public middle and high schools, respectively, were provided with curricula. However, these curricula did not meet their expectations or fully address their teaching needs. They noted that they were insufficient, especially in effectively teaching and integrating data analysis.

Fay expressed her dissatisfaction with the textbook provided, criticizing its structure and integration of content: "I've been doing stuff in a way that throws the whole textbook out of

order just because I wanted to keep stuff relevant. There is nothing specific to analyzing and interpreting data within the text.”

Molly echoed these sentiments, pointing out the random and unsupported inclusion of data analysis in the curriculum:

It’s just random information charts and graphs based on questions, but not in any particular subject area. In science, you just kind of stumble upon it. It’s not in our curriculum. It’s not like there’s a specific unit on data in science, which you know it should be.

Val discussed the challenges with a data-specific curriculum that was not aligned with her teaching needs: “I’m doing the pilot for Data Educator. Which is a platform that we’re piloting, and I’m not finding it very useful.” Val felt the curriculum was too detached from her topics, so she started from scratch to make data lessons relevant and useful. Kelly pointed out another gap in the curriculum related to instructional support for teaching data:

It’s missing how to design a lesson and ways to think about inquiry and collecting data.

What does that look like? I don’t know. We have to figure that out because that’s not typically a science feature of how the curriculum should be taught.

The discussion about the curriculum in educational settings revealed significant challenges related to its accessibility and effectiveness, particularly in data analysis. Several teachers had to develop their own curricula because they were not provided one. In contrast, others who received official curricula found them insufficient and misaligned with their teaching needs. Common criticisms included the lack of specific content for effectively teaching and integrating data analysis, the random and unsupported inclusion of data-related materials, and a general disconnect between the curriculum content and practical teaching requirements. These

issues highlight a pressing need for curricula that are adaptable to the evolving educational demands of data literacy.

3.6.3 Time

Time constraints were a recurring theme in teachers' discussions, particularly in integrating data analysis into their curriculum. They indicated that time constraints are often compounded by factors such as state testing mandates and limited collaboration time, which restrict the depth and quality of data analysis activities in the classroom.

Jes noted that due to time constraints imposed by state testing, she could only collaborate with other teachers on in-depth data projects at the end of the school year: “When we're done with testing, it's like here we have extra time. We'll throw it at the very end when the kids are super wild.”

Anna shared her struggles with time while teaching AP biology. She described her experiences with the curriculum, which included in-depth labs for collecting and analyzing data:

When I was teaching AP biology, they (curriculum) would have a lot of very in-depth labs that involved collecting lots of data and analyzing it deeply. A lot of my struggles came when time just wasn't allowing for it. It would be great if we could have just slowed down and sat on that for days.

This reflects how time limitations prevented her from delving deeper into data analysis, which she felt was crucial for her students' learning.

Molly also highlighted the time challenge, especially when covering all required content. She discussed the difficulties of integrating additional elements like data analysis into an already packed curriculum. Molly remarked, indicating the tendency to overlook non-mandated content due to time pressures:

I think a lot of teachers, anytime that you're told to weave something in, especially elementary because all teachers have so much to get done. It's like, 'Oh, I guess that's expendable because they didn't bother putting that into its own unit. You know what I mean, or it's not already woven in for me. So that means I don't have to do it really.

Kelly emphasized the benefits of cross-curricular work to strengthen students' skills but pointed out that planning time was a significant barrier: "It's a lot harder for us because we just don't have the time to interact with planning." This suggests that more time spent planning could enhance the effectiveness of cross-curricular integration.

Echoing Kelly's thoughts, Molly, a 6th-grade teacher who teaches all content areas, noted that integrating subjects was a practical approach to managing time constraints:

I have to do science in math. I have to do social studies during math, and I have to do reading during social studies. Otherwise, there's just not enough time to get it all to fit in, so you just have to start doing a lot of crossovers.

This strategy reflects her method of ensuring comprehensive coverage of all subjects by blending them together due to time constraints.

Time constraints consistently emerged as a challenge in discussions among teachers, particularly in integrating data analysis within their curriculums. These constraints, often exacerbated by state testing mandates and limited collaboration time, significantly restrict the depth and quality of data analysis activities in the classroom. Teachers noted that these time limitations necessitate a more surface-level engagement with data analysis, which is crucial for student learning. Many responded by integrating subjects to efficiently cover required content, highlighting a need for more flexible and supportive scheduling practices to enhance educational quality and enable effective teaching methodologies.

3.6.4 Resource Availability

This theme relates to the curriculum, highlighting that teachers often need to develop their activities because of deficiencies in the existing curriculum or the complete absence of one. Specifically, one teacher emphasized the need for equipment, a common requirement in science education. The science teachers often required specific equipment to conduct investigations and collect data. The lack of equipment can, in turn, create a lack of resources for working with data. Val discussed using the tools available to her to analyze and interpret data. She was frustrated that they have access to Google Spreadsheets, but it does not always do what they need to do to graph and visualize data: “I don't feel like I know what tools are out there to help me get them there.” Kelly discussed that there were some resources through her district, but it wasn't clear how to access them unless you knew they were there: “The district has some resources out there. It's not available unless you're looking for it. It's not easily available.”

Molly talked about finding experts to help her and her students understand how to read tree ring data. She indicated that it took several steps and reaching out to her own built network to find assistance in reading the tree rings: “I was not 100% sure how to read it. I had to do a lot of research to figure that out and contact some experts who could help us with that.” She continued to say that she does this often when she encounters types of scientific data that she is unclear on how to read:

I'm on my own, so I will just reach out, depending on the project. I'll reach out to people I know in or not in the district. People that I know within my network that I've created over the past few years that are that are environmental educators. Then they put me in touch with the specific people.

She expressed a sense of isolation, indicating that when she needs resources, she lacks the necessary support from her school or district to obtain them.

Anna indicated that having equipment for her various courses is a struggle. There are concepts that she would like to teach through inquiry and allow students to gather authentic data, but she often does not have the equipment necessary to do this and include it:

We have labs that are given to us in our curriculum, but it's not always very feasible. If it's something like chemistry, you must have a lot of lab equipment. We are a private school, but it doesn't mean we have many resources. It's a slow build of getting all the resources that we need and maintaining what we have.

The findings reveal a significant challenge in teaching data analysis and interpretation due to a lack of adequate and accessible resources. Teachers frequently struggle with insufficient instructional tools and scientific equipment and experts, hampering their ability to engage students in data-centric educational activities effectively. Within each focus group, teachers collectively decided to compile a shared document containing resources, highlighting a proactive, collaborative effort to enhance accessibility to materials that could assist in teaching data analysis and interpretation. Additionally, relying on personal networks for expert guidance underscores a broader issue of institutional support deficiencies. These challenges emphasize the need for enhanced support systems to improve resource accessibility and strengthen data literacy education in science classrooms

3.6.5 Student Prior Knowledge and Engagement

The findings on student engagement and motivation revealed educators' multifaceted challenges in fostering student interest and proficiency in data analysis and interpretation. A recurring theme across the discussions was the struggle with student motivation, as exemplified

by Dora: “Sometimes I get work back that's not what it could be. It's just the easy way to put something down, to have it be done.” The issue of engaging students was particularly pronounced in transitioning grade levels. For example, Dora noted a significant disinterest among incoming sixth graders, attributing it partially to their preference for video games over academic content: “We've struggled with that with our incoming sixth graders this year. It's almost like if it's not video games, why should I listen when I can sneak and play my favorite game?”

The difficulty in engaging students was related to their interests and their foundational skills in interpreting and working with data. Teachers noted a general lack of readiness to handle data-driven assignments, especially among younger students and those transitioning into high school. Molly expressed frustration over students' inability to understand basic graphical representations: “You could put a pie chart out there, and they wouldn't be able to tell you what it really means.”

Another critical issue discussed was the inconsistency in how data analysis is taught across different grades, which affects the sequential development of these skills. As Molly pointed out: “It's really just kind of random what you get, what the kids get. It's almost always coming through a math curriculum.”

Additionally, there's a noticeable lack of experience with practical data handling, such as setting up spreadsheets and using data visualization tools effectively. This gap is evident in teachers' efforts to guide students step-by-step through processes that should ideally be more intuitive by their grade level. Val shared: “Many of them are sophomores, and they probably made a spreadsheet last year in a few classes, but they really don't know how to start from data and set up their own spreadsheet.”

The findings on student engagement and motivation highlighted significant challenges educators face in stimulating student interest and developing proficiency in data analysis and interpretation. Educators reported a pervasive issue with motivating students, especially during grade-level transitions. For instance, a marked disinterest was observed among incoming sixth graders, often preferring video games to academic content. This lack of engagement was a matter of interest and reflected inadequate foundational skills necessary for data-driven tasks. Many students showed difficulties with basic data concepts, such as interpreting graphical representations.

Furthermore, inconsistent teaching of data analysis across different grades led to uneven development of these crucial skills. Teachers also noted a general lack of practical experience among students in handling data, such as setting up and utilizing spreadsheets effectively, which often left students unprepared for more advanced data manipulation tasks. These issues collectively suggest the need for more structured and consistent approaches to teaching data literacy across educational levels.

3.6.6 Summary of Findings

The findings, such as insufficient training, curriculum challenges, time constraints, resource availability, and student engagement, depict a multifaceted struggle to integrate data literacy into science education effectively. The study reveals significant gaps in training, with teachers frequently lacking proper instruction in data literacy during their formal education, leading them to seek additional professional development independently. Curriculum issues are also prominent, as some educators are forced to create their teaching materials due to the unavailability or inadequacy of provided resources, highlighting a disconnection between curriculum content and practical teaching needs. Time constraints further complicate these

issues, with strict scheduling and state mandates limiting the depth of data analysis activities possible within the classroom. Additionally, resource limitations are a critical barrier, with many teachers lacking the necessary tools and equipment to conduct thorough scientific investigations, forcing reliance on personal networks and finding their own solutions. Finally, a key challenge is student engagement and prior knowledge, where teachers struggle to motivate students and build on inconsistent foundational skills in data handling across educational levels. These findings highlight the urgent need for systemic enhancements in educational frameworks to foster a more robust integration of data literacy, emphasizing the development of resources, teacher training, and curriculum design to support effective data literacy education.

3.7 Discussion and Conclusions

This study builds upon Banilower's (2018) and Rosenberg's (2022) work by allowing teachers to share their experiences with analyzing and interpreting data, providing a deeper understanding of their challenges in integrating this into their practice. Despite the limited sample size, the insights into contextual factors such as time, training, curriculum, and resources—which have long been studied as influences on teacher practice (Cobb et al., 2003; Jones & Carter, 2007; Lumpe et al., 2000; Luft et al., 2022)—are significant. Specifically, teachers voiced how these factors impact their ability to analyze and interpret data, raising questions about whether these challenges affect all science and engineering practices equally or if data analysis is simply not prioritized. Challenges in teaching, analyzing and interpreting data is an area for further investigation.

The discussion around preservice and in-service training was particularly prominent among the teachers in the focus groups. Interestingly, those reporting inadequate training came from diverse educational backgrounds. For example, Kelly, a high school biology teacher with

an undergraduate degree in biology, expressed surprise at her lack of data literacy training, stating, "Data literacy is something that I did not receive in school." She discussed this gap across her educational journey, from high school to her teaching licensure program.

Similarly, Molly, whose licensure covers all K-8 subjects, noted that her understanding of data primarily came from her math courses in her education program and that even with a master's in economics, she felt unprepared for working with data. These experiences suggest a broader issue beyond just teaching within 6-12 science education; there could be deficiencies across various degree programs. It will not translate into their practice if teachers lack the foundational knowledge to teach a subject effectively (Dong et al., 2018). Additionally, teachers like Val, Kelly, and Jes expressed a desire for more collaboration with colleagues to enhance their skills in analyzing and interpreting data. This finding is corroborated by existing research, which emphasizes the importance of peer support and collaboration as vital components of professional growth for teachers (Thibaut et al., 2018b). These findings suggest a significant need for improved training and support in data literacy, both at the educational and district levels, to better equip teachers for implementing these essential skills in their classrooms.

Curriculum accessibility and quality emerged as significant issues in the findings. All teachers reported the need to either create their curriculum due to its unavailability or modify existing curricula because of their inadequacy. Fay and Molly expressed frustration with textbooks and other resources that do not adequately integrate data analysis and interpretation, necessitating developing or procuring alternative materials. This situation aligns with findings from research on curriculum specific to data analysis and interpretation. Science textbooks often lack abstract or data-centered representations, typically featuring numerous illustrations that do not align with scientific principles (Lee, 2010). This inadequacy increases teachers' workload and

adversely affects the quality of education they can provide. Teachers, especially those at the K-8 level, depend on curriculum materials for content knowledge, which significantly influences science instruction (Stern & Roseman, 2004; Morris et al., 2015). Therefore, the absence of a robust curriculum or the presence of an inadequate one presents a substantial obstacle. Given that 46% of teachers rely on these materials as their primary instructional resource (Morris et al., 2015), the quality of curricula can either facilitate or hinder effective teaching of data analysis and interpretation.

Time constraints further exacerbate these challenges. Lack of time is a major contextual factor that impacts teacher practice (Luft et al., 2022; Thibaut et al., 2017, 2018a; Wright & Waxman, 2022). Anna's inability to fully engage in analyzing and interpreting data due to time constraints highlights a broader issue where practical limitations often diminish curricular ambitions. Educators like Kelly and Molly have suggested that more efficient planning and integration across subjects could help mitigate some of these time-related challenges. Jes expressed frustration that more in-depth data activities had to be pushed to the end of the year after testing for students to participate. These time-related constraints confirm findings from previous studies and identify the need for our education system to streamline processes to ensure teachers have sufficient time for planning and collaboration. Moreover, understanding the specific impacts of time constraints on various teaching activities could guide policymakers and educational leaders in creating more supportive frameworks for teachers. These insights would be valuable in refining educational strategies and improving the overall efficacy of instruction, particularly in data-intensive subjects

Resource limitations represent a significant barrier within the education system. Kelly's struggles to access district resources and Anna's difficulties due to inadequate lab equipment in

science classes exemplify how insufficient resources impede effective teaching and learning. While these resource issues are often connected to curricular needs, they extend beyond curriculum requirements, as demonstrated by Anna's specific need for science equipment. The scarcity of resources restricts teachers' ability to deliver content effectively, impacting both what they can teach and how they teach it (Dong et al., 2018; Thibaut et al., 2017, 2018a; Wright & Waxman, 2022). Furthermore, Molly's reliance on her personal network to access necessary expertise highlights another challenge—educators are forced to find support the education system should ideally provide independently. This situation underscores the need for systemic improvements to ensure educators have the resources and support necessary for effective teaching.

Furthermore, persistent challenges include engaging students and ensuring they possess the necessary foundational skills for data analysis. Dora's observations about student motivation and Molly's comments on the lack of basic data interpretation skills among students point to a need for a more consistent and engaging approach to teaching data literacy from an early age.

These findings indicate a need for systemic improvements across training, curriculum development, and resource allocation to better support teachers in integrating data literacy skills into their classrooms. Enhanced support structures and reevaluating educational priorities are essential to address the identified gaps and ensure students receive data analysis and interpretation skills that lay the foundation for more effective and engaging data literacy education from an early age.

3.7.1 Limitations

The small sample size could significantly affect the generalizability of the results. With fewer participants, asserting that the findings represent a broad range of experiences and

viewpoints is difficult. This limitation is significant when considering the diversity of educational environments and teaching methodologies across different schools and districts.

Additionally, this study relies heavily on self-reported data from educators. While self-reports are a valuable source of information, they are susceptible to certain biases. Participants may consciously or unconsciously present their practices in a more favorable light or may not recall past events and practices with complete accuracy. This reliance on subjective accounts could lead to overestimating the effectiveness of certain practices or missing nuances that only more objective methods could uncover.

These limitations suggest that while the findings provide valuable insights, future research could incorporate a larger, more diverse sample to extend the findings or focus on teachers within one district.

Chapter 4. The Nature of Data Literacy in a Middle Grades Science Curriculum Unit

4.1 Introduction

Working with data is a fundamental part of science (Bowen & Bartley, 2014; Gebre, 2022; NGSS Lead States, 2013). Its importance is noted by being one of the eight science and engineering practices in the Next Generation Science Standards (NGSS) (Lead States, 2013). However, compared to other science and engineering practices, such as modeling (Lehrer & Schauble, 2004; Passmore & Stewart, 2002; Schwarz et al., 2009) and argumentation (Berland & McNeill, 2010; Grooms et al., 2015; Manz, 2015; Rowland et al., 2017), there is less guidance on the specifics of Science and Engineering Practice 4: Analyzing and Interpreting Data. What is generally available are materials for involving students in analyzing and interpreting data in conjunction with the other practices or guides linking them (NGSS Lead States, 2013; Duschl & Bybee, 2014; Schwarz et al., 2017). What is missing is the specifics of working with data in more contemporary ways.

Data collection is considered a large part of science and science education. However, there can be a comprehensive and varying difference in how students work with data versus scientists (Bowen & Bartley, 2020). *A Framework for K-12 Science Education* outlines the goals for the practice of analyzing and interpreting data (2012). It states that students should work with raw data, as scientists do, and organize and categorize it (National Research Council, 2012, p. 61–63). The performance expectations guide the specifics of the content. However, they fall short in several ways, including describing a) what raw data is, b) what forms it can take, d) what digital tools could be used to organize it, and d) what size the data sets should be. These specifics are often left to the teacher or the curriculum (Devetak & Vogrinc, 2013). This poses a challenge

as, in the absence of clear guidance, the instruction of this concept may vary significantly, and the depth at which it will be taught remains uncertain.

Curricula serve as a potential guiding force within a multifaceted system of influences upon instruction (Banilower et al., 2018). Textbooks are a prevalent part of a curriculum that provides opportunities to support students in science instruction (Altan & Cagiltay, 2022; Bryce, 2013; Devetak & Vogrinc, 2013; Lee, 2010; Morris et al., 2015). In addition, they provide support to teachers in the teaching of science education. This implies they have a substantial capacity to influence content and pedagogy (Bryce, 2013; Devetak & Vogrinc, 2013; Stern & Roseman, 2004). The teaching of science is not merely reciting facts but should focus on teaching students to understand the nature of science (Dimopoulos & Karamanidou, 2013; Lee, 2010; Morris et al., 2015; Stern & Roseman, 2004). The most recent iteration of science standards identifies the importance of inquiry-based learning, focusing on the nature of science used by scientists in the field (Dimopoulos & Karamanidou, 2013; Morris et al., 2015). Science textbooks have traditionally presented a sanitized and deterministic portrayal of science, diverging from the inherent nature of science within scientific disciplines (Bowen & Bartley, 2014; Dimopoulos & Karamanidou, 2013). Utilizing data in the real world differs significantly from sanitized depictions found in curricula. Scientists often confront raw, messy data from the outset of research, requiring thorough cleaning and interpretation to derive meaning from it (Schultheis & Kjolvik, 2020).

There are few studies specifically addressing data activities in science curricula and those that have looked more broadly at data activities in curriculum. Morris et al. (2015) conducted a comprehensive analysis of data activities across 20 textbooks, focusing on their integration into chapter content, the methods used by students to record data, the complexity of mathematical

analyses applied, the use of data to support phenomena, and their role in making predictions. However, their study did not align these findings with the Next Generation Science Standards (NGSS) Science and Engineering Practice 4 (Lead States, 2013), nor did it delve into comprehensive and specific aspects of data handling. In contrast, our study developed deductive codes rooted in both Practice 4 and insights from teacher experiences. We examined the types of data, research methods, variables, and how students engaged with the data cycle. Furthermore, we extended our analysis by comparing our findings with a best practices model for teaching data in science classrooms, offering a more refined approach to the pedagogical integration of data activities in curricula.

Our work is intended to address this gap by evaluating the alignment of a prominent middle-grade curriculum with the NGSS framework and contemporary approaches to working with and comprehending data in the 21st century. Our rationale is that curricula is a contextual factor that influences teacher practice and student learning. By analyzing a single grade level, we can take an in-depth look at how it is being presented within a curriculum and how that aligns with NGSS standards and best practices to identify what works and what needs work concerning data literacy in science education curricula.

4.2 Literature Review

4.2.1 Discussing Data Literacy in Science and Engineering Practice 4

The existing literature on Science and Engineering Practice 4 (analyzing and interpreting data) is limited. This section aims to spotlight two scholarly works that specifically delve into the analysis and interpretation of data within the framework of the Next Generation Science Standards (NGSS) that we used to situate better our understanding of teaching resources for analyzing and interpreting data. Through this review, we seek to illuminate the representations of

analyzing and interpreting data in the literature that address instructional methodologies and critique pedagogical approaches.

Schwarz et al.'s (2017) edited book identifies the pivotal role of Practice 4 in Science and Engineering activities. This resource emphasizes key elements for effective data engagement and defines data analysis, advocating for tools like Venn Diagrams, GIS images, and mathematical thinking. It highlights the interconnectedness of Practice 4 with other scientific practices, emphasizing collaboration. The book acknowledges the importance of analyzing observations and working with relevant, contextually pertinent information. While it supports collaboration and relevant data, it broadly defines data as facts, statistics, observations, or information from natural and built environments (Schwarz et al., 2017). Although it discusses diverse data sources, it falls short of explicitly addressing critical data literacy skills or providing specific analysis and interpretation strategies. While it does offer more specifics on working with data, it still comes across as broad and doesn't delve deeply into modern data iterations.

Bowen and Bartley's (2014) book on data literacy specifically addresses analyzing and interpreting data. This resource offers targeted strategies for educators to guide students toward a deeper and more critical engagement with data, aligning with scientists' analytical and interpretive approaches. In addition, it offers critiques of how data is taught in science classrooms and provides examples and strategies for helping students make sense of data in a way that is more aligned with current data literacy practices. Typically, students interact with sanitized or replicated data, lacking the comprehensive experience of the full data cycle employed by field scientists (Bowen & Bartley, 2014, 2020). While such data aids in understanding established relationships, it falls short of preparing students for authentic scientific data practices (Bowen & Bartley, 2020; Dimopoulos & Karamanidou, 2013).

Moreover, data representation as absolute or deterministic in the classroom diverges from the more probabilistic nature of scientific outcomes in practical applications, revealing a disparity between classroom reasoning and field applications. The figure in Appendix A illustrates the potential impact of this contrast on students' perceptions of scientific investigations and their findings. This work contributes to pedagogical foundations that encourage students to approach science more probabilistically by providing a model aligned with modern data practices.

In summary, analyzing and interpreting data primarily focuses on basic data analysis skills but lacks the depth and breadth required for comprehensive data literacy in the 21st century. Real-world data demands a more critical understanding of data's purpose, potential biases, and the ability to navigate and question data effectively. The current NGSS practices do not fully meet these requirements, emphasizing the need for a broader and more critical approach to data literacy in science education.

4.2.2 Current Models for Teaching Data Literacy

This section will explore four current frameworks for teaching data literacy. These four frameworks were chosen because they focus on data literacy across different grade levels and subjects. They were selected because there is limited work specifically related to data in science education. These frameworks provide valuable insights into the importance of data literacy in education and its relevance in various contexts, including science education. They are instrumental in helping us understand the significance of working with relevant data and its implications for empowering learners. They provide valuable insights into the multifaceted nature of data literacy and its relevance in education. They highlight the importance of working with relevant data and integrating data literacy into learning contexts, particularly in science

education. These frameworks offer flexible approaches that can be adapted to suit various educational levels and settings, ultimately empowering learners to navigate the data-driven world effectively.

Open inquiry (2014) provides a basic framework for allowing students to work with data in a more relevant way, with the full data cycle- collection, analysis, and interpretation, and supports a shift in thinking from deterministic to probabilistic thinking in K-12 science education (Bowen & Bartley, 2014, 2020; Dimopoulos & Karamanidou, 2013). Bowen and Bartley (2014) define open inquiry as investigating a problem that is ill-defined, allowing for student modification, allowing students to experience uncertainty and ambiguity, starting at the current knowledge state of the student, and is used to locate conceptual foundations (start where they are), allows materials and discursive practices and resources to be shared and developed within a community of learners, and allows for newcomers to draw on others' experiences and expertise (p. 6). Engaging in an investigation involving an ill-defined problem will enable students to modify the project as they progress. This approach encourages students to grapple with uncertainties and ambiguities when drawing conclusions, fostering a deeper understanding of the subject. By starting with the student's current knowledge in each project, the learning experience becomes more personalized and relevant.

Furthermore, such investigations create a collaborative learning environment, allowing for sharing materials, discursive practices, and resources within a community of learners. This collaborative setting offers newcomers the chance to draw on the expertise of others and utilize any available resource that could enhance their learning, promoting a dynamic and interactive learning process (Bowen & Bartley, 2014). Levels of inquiry can be described as a scale ranging from 0 to 3, where zero represents a level in which problems, procedures, and conclusions are

explicitly provided to students. In contrast, level 3 provides an environment in which students can explore issues, devise procedures, and formulate conclusions independently (Bowen & Bartley, 2014). The conceptualization of open inquiry provides a scaffolded instructional framework to facilitate students toward working a level 3 inquiry, fostering deeper engagement and autonomy with investigations. This inquiry, aligned with the practices of scientists in real-world contexts, contributes to enhancing data literacy and developing higher-order thinking skills in K-12 education (Fontichiaro & Johnston, 2020). It highlights the significance of mirroring scientific practices in the educational framework to foster a deeper understanding of data literacy. Science textbooks and curricula generally stay around the inquiry levels of 0-1, occasionally reaching 3, while scientists work typically in a 3 (Bowen & Bartley, 2014). This illustrates a need for a more comprehensive approach to working with data in K-12 science that reflects the nature of science in the field.

D'Ignazio and Bhargava's (2015) framework challenges the conventional view of data literacy as mainly basic skill competencies, the ability to read, interpret, and communicate data effectively. The first pillar emphasizes agility and adaptability, cultivating adaptive capacities and resilience. This approach prioritizes the development of skills rather than fixating on specific platforms and technologies, recognizing the inevitability of technological evolution. The second pillar is envisioned as building upon essential features across all subcategories of literacy, conceiving literacy as a continuum rather than isolated domains. The third pillar involves empowering individuals to navigate their current ecosystems and societies in ways that are personally significant to them. Data literacy has the potential to empower citizens and communities (D'Ignazio & Bhargava, 2015; Raffaghelli et al., 2020). The fourth pillar advocates for providing diverse pathways to accommodate individuals with varying data literacy needs and

capacities, acknowledging the complexity of the overarching system. The four pillars advocate for shifting from rigid technical competencies toward a holistic understanding of data literacy as essential for 21st-century life and not just for careers in STEM. Their framework emphasizes the significance of data literacy in fostering societal inclusion and citizenship and recognizing the relevance of data to learners' lives (D'Ignazio, 2022; Raffaghelli et al., 2020).

Maybee and Zielinski's (2015) framework stresses the importance of building upon learners' prior experiences when teaching data literacy. They emphasize that new ways of using data should be integrated into disciplinary core subjects and that learning should result in students developing new understandings of the subject being studied. This framework highlights the necessity of context in data-informed learning, reinforcing that data literacy should not be taught in isolation but incorporated into all learning contexts (Koltay, 2016b; Raffaghelli, 2020). This approach aligns with the concept of relevance, particularly in science education. Relevance has been stated as an essential aspect of teaching science, specifically in analyzing and interpreting data (Schwarz et al., 2017). Teaching data literacy through the lens of students' personal experiences and communities can engage students and deepen their understanding of the data cycle. While this framework initially aims at higher education, it can be adapted in K-12 contexts.

Srikant and Aggarwal's (2017) framework is particularly relevant to middle-grade science education, as it directly involves K-12 learners in the data collection process and emphasizes the need for relatable data. Their framework (Appendix C) promotes the full data cycle- data acquisition and collection, data processing and cleaning, analysis, data visualization and communication, relatable data sets (e.g., about themselves), and hands-on data collection and entry (Miao et al., 2022). Scaffolding the instruction ensures learners can comfortably engage

with data to solve real-world, contextual problems related to their experiences. This framework is a strong foundation for building data literacy in science education (Wolff et al., 2019). It demonstrates that learners can be motivated and engaged with relatable and meaningful data. While this framework originated from a specific study on data literacy, it can be adapted for use in various contexts where data is a valuable learning resource.

4.2.3 The Role and Impact of Science Textbooks

Textbooks are crucial as curricular tools in science education, serving as the primary instructional resource for 46% of educators and a secondary resource for another 40% (Morris et al., 2015). These materials significantly influence science instruction by providing vital information and pedagogical guidance, which is particularly important in K-8 settings where science teachers may not have specialized in science subjects during their training (Stern & Roseman, 2004).

The complexity of science content, which includes intricate vocabulary, language, and the need for visual integration, demonstrates the importance of textbooks (Devetak & Vogrinc, 2013). Effective textbooks incorporate textual and pictorial content, enhancing teaching and learning experiences (Stern & Roseman, 2004). However, the deterministic writing style of many textbooks may present science as an absolute field, which deviates from the principles of scientific inquiry and can lead to misconceptions among students and teachers (Bowen & Bartley, 2014; Lee, 2010).

Morris et al. (2015) analyzed 20 textbooks encompassing 731 activities, revealing that only about half of these activities engaged students with data. The analysis identified six categories for meaningful engagement with data but found that few activities provided

meaningful opportunities for this type of engagement, and instructional support was often lacking (Morris et al., 2015).

When evaluating a science textbook, it's essential to consider content alignment with scientific principles, pictorial components' substance and educational value, and the clarity of instructional support. These factors ensure that textbooks serve their intended educational purposes and support effective learning (Devetak & Vogrinc, 2013; Stern & Roseman, 2004). Moreover, textbooks should be critically examined as they significantly shape teachers' approaches in their classrooms (Devetak & Vogrinc, 2013).

4.3 Conceptual Framework

4.3.1 Synthesizing Best Practices for Teaching Data Literacy in 6-12 Science

We synthesized information from the four frameworks mentioned in the previous section into a singular framework titled Synthesized Best Practice Model (SBPM) for Teaching Data Literacy in 6-12 Science Education. We sought adaptable frameworks and identified four distinct ones: Open Inquiry (Bowen & Bartley, 2014), Data Informed Learning (Maybee & Zilinski, 2015), Foundational Pillars of Data Literacy (D'Ignazio & Bhargava, 2015), and Workflow through Data Science Tutorial (Srikant & Aggarwal, 2017). These frameworks, addressing various educational levels and content areas, were analyzed for their applicability to 6-12 science education to determine best practices for teaching data literacy. Through an in-depth examination, we synthesized a model incorporating commonalities among the four frameworks, allowing for comparative analysis between existing and ideal conceptualizations of data literacy education. The connections are identified in Appendix D. This synthesized model is tailored to meet the unique needs of teaching data literacy within the 6-12 science education. This model envisions a framework based on current research, serving as a foundation for comparative

analysis in this curriculum. It facilitates evaluating the alignment between existing practices and the idealized framework for teaching data literacy in science education.

4.3.2 *OpenSciEd Curriculum*

OpenSciEd, an open-source science curriculum tailored for grades 6-12, is a significant contextual factor with the potential to impact data instruction profoundly. This curriculum employs a storyline structure that integrates research-driven principles emphasizing student coherence, epistemic agency, and meaningful science (McNeill & Reiser, 2018; Reiser et al., 2021). Aligned with Next Generation Science Standards (NGSS), it emphasizes merging scientific practices and concepts within each lesson, thereby nurturing collaborative knowledge-building (McNeill & Reiser, 2018). Noteworthy for its student-centered, phenomena-driven learning experiences, OpenSciEd's unique attributes encompass accessibility, adaptability, and peer-reviewed certification (McElhaney et al., 2022). Its logic model emphasizes alignment with state and district policies and strives for enhancements in student engagement, application of scientific practices, and equitable outcomes (Edelson et al., 2021).

Incorporating diverse stakeholder collaboration, NGSS practices, and open-source accessibility, OpenSciEd is a prime example of high-quality, reform-oriented science education content. Despite uncertainties regarding its widespread utilization, partnerships with states, educators, and educational entities illustrate its expanding reach and potential impact on science education (OpenSciEd, 2023). Furthermore, collaborations with data literacy platforms like Tuva underscore efforts to enrich curriculum content with authentic data sets and tools (OpenSciEd, 2023).

The OpenSciEd curriculum exhibits characteristics akin to a textbook by incorporating textual and visual components to enrich the learning experience. In our literature review, we

found it pertinent to consider it as a textbook, particularly concerning the utilization of Science and Engineering Practice 4, where data is frequently conveyed through the visual elements in textbooks. In its digital format, distinct from traditional textbooks, the virtual version shares similar elements, offering a basis for assessing the data components within the curriculum. It is key to note that printed versions are available for this curriculum. However, our focus is on the accessible digital format. Our focus in this analysis is directed solely towards the materials engaged with by students, excluding the teacher support text.

4.4 Purpose and Research Questions

This study aims to conduct an in-depth examination of the nature of data within the OpenSciEd 7th-grade science curriculum. The study includes analyzing the types of data, data visualizations, data sizes, and data variables presented. Additionally, the study will explore the types of data analysis used and how students are expected to engage with the data. Furthermore, the curriculum's alignment with the Next Generation Science Standards (NGSS) and the synthesized best practices model will be evaluated to determine its effectiveness in teaching data literacy. This research will address two central questions about data practices in OpenSciEd:

RQ 1: What characterizes the nature of the data used by students in the OpenSciEd 7th-grade curriculum?

- RQ 1a: What types of data are included?
- RQ 1b: What types of visualizations are included?
- RQ 1c: What is the size of the data sets?
- RQ 1d: What are the sources of data?
- RQ 1e: What are the variable types?

RQ 2: How is analyzing and interpreting data incorporated within the OpenSciEd curriculum at the seventh-grade level?

- RQ 2a: What methods of data analysis are students participating in?
- RQ 2b: How does this data analysis align with modern working with data in the 21st century?

4.5 The Study

The primary objective of this study is to examine the curriculum as a contextual factor influencing data literacy instruction in science education classrooms. Specifically, the research will thoroughly analyze the data practices embedded in the 7th-grade curriculum from OpenSciEd. This section will explain the methodology, data source, and analysis

4.5.1 Research Design

We used qualitative content analysis to analyze the 7th-grade science curriculum from OpenSciEd. Qualitative content analysis is a versatile method to explore how concepts manifest within extensive document sets (Gläser-Zikuda et al., 2020; Grbich, 2012). This method will permit interpretation of textual or visual content by applying coding techniques to discern recurring themes (Grbich, 2012; Purwani et al., 2017). In this analysis, we used deductive coding to analyze the curriculum (Grbich, 2012) and answer RQ 1 and RQ 2a. This approach aligns with the study's objective of evaluating how the curriculum presents the nature of data. To address RQ 2b, we compared the analysis from RQ 1 and 2a to the synthesized model of best practices for teaching data literacy (SBPM) (Appendix D) to determine how the curriculum aligns with the best practices defined in the literature.

4.5.2 Data Source

The data source for this analysis is the 120 lessons in the 7th-grade curriculum from OpenSciEd. It was selected because it is accessible to every educator and aligns with NGSS (Edelson et al., 2021). This curriculum offers all text materials digitally at no cost to educators. In addition, it meets the expectations of the criteria through Ed Reports and the NGSS Design Badge. Ed Reports and NGSS provide educator-led reviews of K -12 math, ELA, and science curricula based on alignment with standards and usability (*Ed Reports*, 2023, *NGSS Design Badge | Next Generation Science Standards*, 2024.). We selected the 7th-grade level for analysis, recognizing a research gap in the literature concerning middle grades. This grade band is particularly significant as studies indicate a decline in STEM interest during this period (Genareo et al., 2016). We used the Elements document (which lists the SEPs and what lessons they are in) provided in each curriculum unit to determine which lessons to analyze.

4.5.3 Coding Frame

We devised the deductive coding (Creswell & Báez, 2020; Xu & Zammit, 2020) structure for this study's qualitative analysis by merging insights from science education, our own teaching experience in K-12 science education and higher education, and data-focused research. We scrutinized the analysis and interpretation of data, leveraging the viewpoints of seasoned educators and data researchers to broaden our understanding of the practice's principles. Moreover, the best practices aligned with each code to ensure consistent support. The table in Appendix H describes each code.

4.6 Data Analysis and Inter-Rater Agreement

To identify instances of data utilization across the six units, we utilized the OpenSciEd Elements for NGSS, a document identifying lessons within each unit that incorporate Science

and Engineering Practices (SEPs). The analysis focused exclusively on lessons listed in the Elements of NGSS document that explicitly involved student interaction with data within all the Science and Engineering Practices. We opted for a comprehensive examination of all instances where data was incorporated to ensure thorough coverage within this curriculum. Nonetheless, while the primary emphasis lies on analyzing and interpreting data in this study, it is noteworthy that NGSS is recognized for incorporating scientific practices. Omitted from the study were self-assessments and teacher rubrics, aligning with the emphasis on understanding how students engage with data.

Two coders established inter-rater reliability. The initial coder refers to the author, and the subsequent coder denotes a research assistant affiliated with the Rosenberg lab. The initial and subsequent coder coded the first two units separately and then analyzed them concurrently to determine interrater reliability. For the remainder of the study, the initial and subsequent coded two units separately and met to determine interrater reliability for the remaining data. To conduct a systematic analysis of the curriculum, our initial step involved a detailed review of the "Elements of Next Generation Science Standards (NGSS)" document provided within the curriculum resources. This document guided our selection of specific lessons for further examination. Each lesson was analyzed to identify instances where the predefined codes from our codebook were demonstrated. These codes were then documented alongside the corresponding practice, lesson title, and document in the codebook for clarity and reference.

The curriculum is hosted on an open-source platform in Google Drive, structured in a non-traditional format where each lesson component, such as slides, handouts, and teacher guides, is stored in separate files. This necessitated extensive cross-referencing between multiple files to verify the presence of specific codes within the lessons. Additionally, due to the

scaffolded nature of the curriculum, it was imperative to review several preceding and subsequent lessons to determine the continuity and presence of these codes accurately. Observations and notes regarding these cross-references and the scaffolded integration of the curriculum were systematically recorded in a designated notes section at the end of each coded entry in the codebook. This approach ensured a comprehensive analysis of how well the curriculum aligns with the NGSS elements and facilitated a deeper understanding of its pedagogical structure.

Upon completing the coding process within the codebook, we proceed with the data analysis phase. We utilize the chart function in Google Sheets and the Common Online Data Analysis Platform (CODAP) to analyze the coded data. These tools were selected because of our commitment to open-source resources, as both Google Sheets and CODAP are freely accessible and provide the necessary functionalities for our analysis. In our code book, we listed each lesson that included data from the elements of the NGSS document. We created headers for each code and marked the corresponding box if a lesson included one of our codes. If a lesson did not include a code, we left the box blank.

After coding the entire curriculum, we used Google Sheets to perform a frequency count of the occurrences of each code. The frequency counts were then transferred to a separate spreadsheet, where we utilized the chart function to create bar graphs for analysis. Examples of these bar graphs can be found in Appendix L. This approach aligns with our resource constraints and supports our aim to maintain transparency and accessibility in our research methodology.

To address research question 2b, we utilized a deductive approach by aligning the codes from our codebook with the SBPM. This alignment evaluated how closely the curriculum adheres to established best practices in the literature. Additionally, we conducted a comparative

analysis between the definitions of these codes and the definitions of best practices to identify any discrepancies or variations.

The correlations established through this alignment are detailed in Appendix J. Using our initial analysis derived from the deductive codes, we could interpret the curriculum through the lens of the best practices model. This approach enabled a structured examination of the curriculum's integration with recommended educational practices.

4.7 Findings

4.7.1 RQ1- The Nature of Data Used by Students in the 7th-grade Curriculum

Several intriguing findings emerged from the curriculum analysis. The occurrences of each code for RQ 1 can be found in Appendix K. Students were observed working with raw data sets 50 times. However, these data sets were relatively small and collected by the students themselves (Figure L.1, Appendix L). Furthermore, the student can organize their data into pre-made blank tables 44 times, indicating a lack of opportunity to explore data organization or clean the self-collected raw data (Figure L.2, Appendix L). Throughout the activities, scaffolding was provided to support students in various aspects of the process. For example, the tables were pre-created for students to fill in activities involving data collection.

Another significant finding was that students could interact with primary data sources 45 times, which were presented in tidy and curated sets within preexisting or mostly filled-out tables. While students can engage with raw data and technically participate in the full data cycle, it became apparent that these activities did not fully mirror the authentic process experienced by scientists, and the data sets were mostly small, involved curated tables, and did not require data cleaning. Although this scaffolding is beneficial, there should be occasions for students to decide how to clean and organize the data independently.

Moreover, it was noteworthy that digital tools were utilized only twice, specifically through online tools. Only one activity in the curriculum aligned with statistical and probabilistic data analysis methods. In this instance, the students did not perform the statistical calculations themselves; instead, a digital tool conducted the analysis on their behalf.

In summary, our analysis of the curriculum highlighted several key findings. While students can engage with raw and curated data sets, independent data cleaning and organization opportunities were absent. Scaffolding was evident throughout, yet the curriculum fell short of mirroring authentic scientific practices. Limited use of digital tools further emphasizes the need for enhancements in data literacy instruction.

4.7.2 RQ 2A-Types of Analysis

In our investigation into how students engage with data, we closely examined the types of analysis they employed within various activities (Figure L.3, Appendix L). This examination revealed that students utilized quantitative analysis methods, which occurred 25 times throughout the curriculum. For example, one activity involved measuring the change in mass of a crushed bath bomb in a closed bottle. Students can record both initial and final mass measurements, along with the release of gas, enabling a comprehensive understanding of the experimental outcomes. Qualitative data collection was observed 34 times. Students typically recorded their observations without engaging in further qualitative analysis. The qualitative data was collected, but qualitative analysis methods were not emphasized and only used to support the quantitative analysis. For instance, during an activity where students dropped a bath bomb into water, their focus was on visually observing the experiment. However, subsequent quantitative measurements were conducted to support and validate their qualitative observations.

4.7.3 RQ 2B- Frequency of Student Data Practices

In examining the integration of various research skills within the curriculum, a distinct frequency pattern emerges across the activities of collecting, organizing, analyzing, and interpreting data. For this analysis, we assessed how often students engaged in these tasks individually and in combination throughout the lessons.

Our analysis included 120 lessons that incorporated data practices. The "Analyzing" data activities were the most frequent, occurring 98 times, highlighting its dominance within the curriculum. This was followed by "Interpreting" data, observed 83 times, "Collecting" data 49 times, and "Organizing" data 32 times. These frequencies suggest a curriculum focusing primarily on analyzing and interpreting data, which is critical for developing students' higher-order thinking and data literacy skills.

The most common task combination within lessons was analyzing and interpreting data, appearing in 44 instances. This frequent combination underscores the curriculum's strong emphasis on equipping students to examine data and critically derive meaningful conclusions.

Additionally, 26 lessons involved a comprehensive approach that spanned collecting, organizing, analyzing, and interpreting data within the same lesson. This approach promotes a holistic understanding of the entire data cycle. Lessons focusing specifically on data analysis were noted 15 times, emphasizing a specialized focus on this aspect of data handling. Engaging students in collecting, analyzing, and interpreting data, excluding the organizing component, was observed in 10 instances, emphasizing the dynamic aspects of data handling. Data collection only tasks occurred in 7 instances, likely serving as introductory activities to data management. Collecting and organizing data, which may represent more foundational or

introductory lessons, was noted in 4 instances. The occurrences of each code for RQ 2 can be found in Appendix K.

This analysis highlights the curriculum's emphasis on analytical and interpretive skills and indicates areas for potential enhancement in the comprehensive integration of data handling experiences.

4.7.4 RQ 2B *How does this data analysis align with the SBPM?*

The comparative analysis of the findings with the Synthesized Best Practices Model (SBPM) shed light on several key experiences with data within the examined curriculum. Firstly, the ability to navigate ambiguity, crucial for effective data analysis, is somewhat limited. While students have some opportunities to create their own ways of organizing data, most experiences involve already created tables for organization and many curated data sets for analysis. The inclusion of student-collected raw data sets allows for some level of decision-making and ambiguity, albeit within a structured framework. Furthermore, the curriculum provides opportunities for students to engage in the full data cycle with 26 instances observed. It is important to note that these instances predominantly involve small data sets.

The concept of relevance within the curriculum is primarily tied to the data students collect rather than broader community-based or open-inquiry approaches. While students may engage in activities where they list what they know, community relevance is not explicitly integrated into the curriculum.

Regarding the experiential practice, students predominantly engage with "messy" data collected by themselves rather than raw data sourced externally. It does emphasize a hands-on approach to data analysis. However, this limits their experience working with either curated or small student-collected raw data sets. Moreover, the curriculum provides sufficient scaffolding

for students' data analysis by offering opportunities to work with data ranging from pre-existing curated data sets to activities involving created blank tables. The structured progression in scaffolding throughout the curriculum, including instances where students are tasked with creating their own tables, facilitates a deeper understanding of data analysis processes. Finally, the curriculum provides opportunities for collaboration through group work and class discussions, enhancing students' learning experiences and fostering a collaborative learning environment. Overall, while there are areas for improvement, the curriculum effectively supports students' engagement with basic skill competency for data literacy. However, it is lacking in higher-order data skills, specifically working with large raw data sets and visualizing and communicating data and findings. The curriculum does a thorough job of scaffolding. It also does a great job of collaboration as the entire curriculum is set up for student discourse.

4.8 Discussion and Conclusions

The findings from this study reveal several critical insights into the nature of data usage, analysis types, and data handling practices within a 7th-grade science curriculum. The curriculum emphasizes foundational skills in data literacy, which aligns with the literature indicating that basic skills are how data literacy is generally defined (Gebre, 2022; Wolff et al., 2019). However, this emphasis also highlights significant areas for improvement, particularly in providing students with more advanced ways of working with data, a skill essential in modern data handling (Bowen & Bartley, 2014, 2020; D'Ignazio & Bhargava, 2015).

Students frequently engaged with raw data sets and were tasked with organizing their data into pre-made tables, demonstrating an effort to provide hands-on experience. However, these data sets were small and self-collected, limiting exposure to larger, more complex datasets that better reflect authentic scientific inquiry—using pre-made tables restricted opportunities for

students to independently clean and organize data, which are critical skills in data literacy (Bowen & Bartley, 2014, 2020).

The curriculum incorporated both quantitative and qualitative analysis methods, with a stronger emphasis on quantitative analysis. While students collected qualitative data, it was primarily used to support quantitative findings, suggesting an imbalance that might not fully develop students' abilities to conduct and appreciate qualitative research. The limited use of digital tools and statistical methods further shows a gap in the curriculum, as these tools are essential for modern data analysis (Rosenberg et al., 2022). Using statistical methods and digital tools is vital to modern data work (Rosenberg et al., 2020, 2022).

The curriculum provided numerous opportunities for analyzing and interpreting data, as evident from the frequency of these activities. However, the lesser emphasis on collecting and organizing data indicates a potential misalignment with comprehensive data handling practices. This focus on basic skills, rather than a more comprehensive approach to working with data, is consistent with findings in the literature (Bowen & Bartley, 2014, 2020; D'Ignazio & Bhargava, 2015). Although analyzing and interpreting data were common activities, a more balanced integration of all aspects of data handling could provide a more holistic learning experience. The structured progression and scaffolding within the curriculum are beneficial, yet there is a need for more opportunities that encourage independent data handling and critical thinking. Scaffolding is essential, but support must eventually be removed to allow students to work more independently and complexly with data (Bowen & Bartley, 2014, 2020).

While the curriculum shows some alignment with modern data practices, particularly in scaffolding and collaborative work, it falls short in offering experiences that fully reflect the ambiguity and complexity of actual scientific data, an essential aspect of the nature of science

(Bowen & Bartley, 2014, 2020). The limited use of large raw datasets and digital tools for analysis suggests that students are not adequately prepared for higher-order data skills. The predominant use of pen and paper for data work, as noted in the literature (Rosenberg et al., 2022), does not meet the expectations of modern data practices where datasets are often too large for manual handling (Rosenberg et al., 2020, 2022; Schultheis & Kjellvik, 2020). It is crucial to expose students to working with data in ways that are comparable to handling modern large datasets (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Maybee & Zilinski, 2015; Srikant & Aggarwal, 2017). Furthermore, the curriculum's relevance is primarily tied to student-collected data rather than broader community-based or open-inquiry approaches. This limitation could restrict students' understanding of data literacy's more comprehensive applications and implications. Without more extensive opportunities to experience the full data cycle, students cannot understand data in all its aspects (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Srikant & Aggarwal, 2017).

In conclusion, while the 7th-grade science curriculum provides a solid foundation for developing basic data literacy skills. However, it needs substantial improvements to fully prepare students for the complexities of data analysis in the 21st century. Critical areas for improvement include expanding data sets and complexity by introducing more extensive and complex datasets that mirror real-world scientific data, thereby helping students develop more advanced data-handling skills. Encouraging independent data cleaning and organization will foster critical thinking and practical skills essential for data literacy. Balancing quantitative and qualitative analysis methods will offer a more comprehensive understanding of data interpretation. Integrating modern tools and techniques, such as digital and statistical methods, into the curriculum will align student experiences with current data analysis practices. Lastly, enhancing

relevance through community-based and open-inquiry approaches will make data literacy more engaging and pertinent for students. The curriculum can better equip students with the skills and knowledge necessary to navigate and excel in a data-driven world by addressing these areas.

4.8.1 Future Directions

A comprehensive review of the entire curriculum and its alignment with the K-12 Framework for Science Education and the synthesized best practices model (SBPM) would offer a holistic view of how the curriculum addresses the goals set out in the framework over various grade spans. This would help identify how each grade level builds the foundation for the next. Comparing each level against the synthesized best practices model would also be beneficial, revealing how well the curriculum aligns with current best practices for teaching data literacy. Additionally, examining other curricula using the same methods would allow comparisons to see how different curricula align with the goals and best practices for teaching data analysis and interpretation. Furthermore, researching a teacher using a specific curriculum to see how it translates into their practice and its benefits and challenges would provide practical insights. Working towards designing a curriculum based on the SBPM and studying how it is implemented would be valuable in making science education curricula more aligned with best practices in the literature.

4.8.2 Limitations

Conducting an analysis of an open-source curriculum has limitations. In this study, we analyze one grade level of the curriculum (7th grade). By digging deep into one grade, we had the opportunity to construct a comprehensive and in-depth understanding of how a curriculum presents data literacy activities in one grade level.

While NGSS is an authoritative national standard for science teacher preparation (Bybee, 2014), we know that being directed only by NGSS limits us to focus on states that adopted NGSS or utilized the framework for their standards. This is important because different standards are available, and standards in different states may vary. For example, not all states adopted NGSS. States that do not use or develop their own standards based on NGSS may have different data practices. Future research may address this limitation by integrating several standards into the analysis.

Finally, it is noteworthy that there is a gap between theory and practice. Curriculum analysis can only reflect the quality of the curriculum from the theoretical perspective, which can provide insights for teachers when selecting instructional resources (Pinfield et al., 2020). However, teachers may adjust the content in the classroom (Fan & Kaeley, 1998; Lepik et al., 2015). Different teachers have different teaching styles and philosophies, which also impact the effectiveness of instruction (Qi et al., 2018; Toyama & Yamazaki, 2020; Giles et al., 2006). In addition, there are other factors impacting education, such as demographic and content factors. In this study, we solely investigated the curriculum.

Chapter 5. Overall Findings, Conclusions, and Implications

5.1 Key Findings

This dissertation comprises three studies that examined the practice of analyzing and interpreting data within grades 6-12 science education. Chapter 2 focused on teachers' approaches to teaching, analyzing and interpreting data, Chapter 3 focused on their challenges, and Chapter 4 evaluates a 7th-grade curriculum's approach to data literacy. The first study (Chapter 2) delved into teachers' specific strategies to teach data analysis and interpretation while taking contextual factors such as grade level and content into account (Lumpe et al., 2000; Thibaut et al., 2017, 2018b), contrasting with existing literature that primarily relies on survey data that specifies whether aspects of the practice are used or not (Banilower et al., 2018) or what tools are used to complete data activities (Rosenberg et al., 2022). This study provided a detailed look at educators' approaches to teaching, analyzing and interpreting data, highlighting the importance of community-based activities and relevance in teaching data literacy. Teachers often linked data activities to local phenomena and datasets, encouraging students to collect authentic local data and align with the literature, emphasizing the importance of relevance in engaging students with data (Schwarz et al., 2017).

Participants also stressed the necessity of teaching basic skills and supporting learning through scaffolded experiences, which are widely recognized as crucial in educational literature (Bowen & Bartley, 2014, 2020; D'Ignazio & Bhargava, 2015; National Research Council, 2012). The responses indicated that teaching and enhancing fundamental skills through curated experiences were key components of their instructional strategies.

In this study, I also evaluated how teachers' responses aligned with the objectives set by *A Framework for K-12 Science Education* (National Research Council, 2012), focusing on the

established goals for science education by the end of 12th grade. The findings indicated that the teachers in the focus groups successfully met several goals, such as systematically analyzing data, identifying patterns, and utilizing computer technologies. However, there was a notable absence of physical models in their data-handling approaches. This partial meeting of goals is understandable, given that the six goals are expected to be met by 12th grade, and the participating teachers represent a wide range of grades from 6th to 12th.

Additionally, I compared their responses to a Synthesized Best Practices Model for Data Literacy (SBPM). To some extent, the participants' responses aligned with all the practices outlined in the SBPM. They allowed students to experience ambiguity by creating their investigations and data collection activities. They also facilitated engagement with the full data cycle, albeit with high levels of scaffolding and using smaller datasets. The most significant connections were observed with Relevance, Scaffolding, and Collaboration practices. The teachers frequently discussed strategies for making data literacy instruction more relevant and engaging by utilizing local data and community resources. They emphasized the importance of structured scaffolding in setting up data experiences and noted how students collaborated on data projects, further enhancing the learning process.

A significant finding from comparing teacher practices to the SPBM was the role of technology in analyzing and interpreting data. The participants noted the necessity of technology in various aspects of data literacy, from collection to organization and visualization, affirming its importance as outlined in the *Framework for K-12 Science Education* (National Research Council, 2012). However, the SBPM (Appendix D) did not prominently feature technology but was implicit within the practices. This has led to the inclusion of a section on technology use in

SBPM (Appendix D), a significant contribution supported by both the *Framework for K-12 Science Education* (National Research Council, 2012) and empirical data from this research.

Chapter 2 of this dissertation enhances the existing literature by detailing the methods a sample of teachers use to teach data analysis and interpretation. It examines how these methods align with the established *Framework for K-12 Science Education* (National Research Council, 2012) and SBPM. The findings identified that relevance and community-based learning are prominent and interconnected in the teachers' approaches to teaching data analysis and interpretation. This study addressed a gap in the literature by offering specific examples of how teachers implement these practices in their classrooms, thereby providing insights into the practical application of data literacy in science education.

Despite its small sample size, the study's diversity in participant demographics—including different geographical locations, school types, grade levels, and educational backgrounds—offers a valuable foundational perspective on teachers' varied approaches to teaching data analysis and interpretation (Dong et al., 2018; Thibaut et al., 2017, 2018a, 2018b; Wright & Waxman, 2022). This diversity suggests that the findings could provide a foundational view of instructional practices in data literacy across different educational contexts.

Chapter 3 focused on the challenges teachers encounter in these endeavors. Given that both chapters 2 and 3 derive from the same transcript data, it is natural to observe some thematic crossover between them. As teachers discussed their methods, they frequently discussed the challenges intertwined with their approaches, illuminating gaps in the existing literature regarding teachers' difficulties in teaching data analysis and interpretation.

The participants' responses revealed a noteworthy deficiency in training related to data literacy education. Despite varied educational backgrounds, including one participant with an

undergraduate degree in biology—a core STEM area—there was a consensus about the lack of targeted instruction on data during their schooling. The study indicated that both preservice training and ongoing professional development significantly influence how data analysis and interpretation are taught in grades 6-12 for these teachers. This is supported by the literature in that teachers' educational backgrounds impact how they integrate concepts into their practice (Dong et al., 2018; Thibaut et al., 2017, 2018a, 2018b; Wright & Waxman, 2022). The lack of formal education in this area led participants to seek out learning opportunities independently (Chapter 2), whether through collaboration with colleagues or personal research, showing they recognize the importance of data literacy.

Resource constraints and time limitations were highlighted as major challenges, aligning with literature identifying these as significant contextual factors impacting teacher practice (Luft et al., 2022; Thibaut et al., 2017, 2018a; Wright & Waxman, 2022). Teachers cited inadequate access to data teaching materials and lab equipment for data-related investigations. The lack of lab equipment is a significant issue in science education because specific equipment is often necessary for gathering data in investigations in courses such as Chemistry. Without this equipment, the ability to collect data for various investigations is limited. Additionally, working with data involves numerous tools, leaving teachers unsure of where to start. This concern is valid, given the vast array of methods for collecting and analyzing data today. Navigating these tools is challenging, especially for teachers without strong training and background in data handling.

The teachers faced two time-related challenges. Firstly, they lacked sufficient time to work comprehensively with data, which limited their ability to delve into modern data practices. Secondly, they did not have enough time to collaborate or plan data activities despite recognizing

the importance of such collaboration for their professional growth. This finding reflects broader findings in educational research, which emphasize that teachers often lack the necessary time to engage deeply in their instructional practices (Thibaut et al., 2017; Wright & Waxman, 2022).

Teachers reported significant challenges in engaging students with data-related activities, observing a general lack of interest and motivation in learning about data. This finding aligns with the observations in Chapter 2, where teachers incorporated community-based activities to enhance relevance and engagement for students to increase student engagement. Despite these efforts, the present analysis reveals that the issue of student disengagement was noted. Teachers consistently identified this lack of interest as a significant barrier, suggesting that while community-based activities may offer some benefits, additional strategies are needed to address the widespread disinterest in data literacy. This finding highlights the importance of developing more effective approaches to make science education more appealing and relevant to students.

Curriculum limitations further compounded these challenges. Varied access to curricula was reported, with some teachers having to create their own due to the absence of provided materials. In contrast, others who had curricula provided found the curricula inadequate for effectively teaching data analysis and interpretation. This issue is critical as existing literature notes that science textbooks often present data in generic formats that do not align with scientific principles (Devetak & Vogrinc, 2013; Lee, 2010), a significant concern given that textbooks are a primary resource for 46% of science teachers (Morris et al., 2015). The inadequacy of curricular materials poses a substantial hurdle for teachers in effectively teaching, analyzing, and interpreting data.

Chapter 3's findings on curriculum challenges align well with the focus of Chapter 4, which examines how a science curriculum, specifically for 7th grade using the OpenSciEd open-

source curriculum, addresses the teaching of analyzing and interpreting data. This chapter assesses how the curriculum presents analyzing and interpreting data, its alignment with NGSS practices, and its alignment with the Synthesized Best Practices Model (SBPM). The analysis reveals that while the 7th-grade curriculum provides numerous opportunities for students to engage with data, these predominantly involve small, student-collected data sets or tidy, curated data sets. This finding aligns with existing literature, indicating that teachers frequently engage students in data activities involving small data sets and extensively utilize curated data sets (Rosenberg et al., 2022). A significant gap identified is the limited opportunities for students to use digital tools and engage in statistical methods. This gap contrasts the goals outlined in a *Framework for K-12 Science Education* (National Research Council, 2012) and the Next Generation Science Standards (Lead States, 2013), emphasizing the importance of utilizing digital tools and statistical methods to analyze and interpret data. The curriculum often provides pre-created tables, reducing chances for students to manage and organize data independently. It is important to note that almost all data-related activities are primarily conducted using worksheets. Scaffolding is consistently present, but there is minimal progression toward removing these supports to allow students to navigate the full data cycle independently. Scaffolding is essential, but it needs to be facilitated in a way that strips away support for students to be more independent (Bowen & Bartley, 2014).

The comparative analysis with the SBPM shows that the curriculum allows students to experience some ambiguity albeit within a structured context that frequently relies on predefined tables and curated data sets. It promotes involvement in the full data cycle in 26 instances out of 120 lessons, though these typically involve small data sets and lack substantial integration of relevance beyond that they collect the data. The curriculum effectively supports experiential

learning with "messy" student-collected data and fosters collaboration through group work and discussions. However, it falls short of developing advanced data skills such as handling large, raw datasets, utilizing statistical methods, using digital tools, and effectively visualizing and communicating findings. While it excels in nurturing basic data literacy and collaborative skills, there is room for enhancement in facilitating more comprehensive data analysis experiences.

This dissertation provides valuable insights into the teaching of data analysis and interpretation in grades 6-12 science education, emphasizing three key findings. First, teachers often use local phenomena and community-based activities to make data literacy relevant and engaging for students. This approach effectively teaches basic data skills through scaffolded experiences, crucial for student understanding. Second, teachers face significant challenges, including inadequate formal training, lack of resources, and difficulty engaging students. Many teachers feel unprepared for their educational backgrounds and must seek additional learning opportunities independently. Resource constraints and time limitations further hinder their ability to teach data literacy effectively. Third, the 7th-grade OpenSciEd curriculum provides some opportunities for data engagement but falls short in integrating advanced data skills, digital tools, and statistical methods. Most data-related activities rely on worksheets and pre-created tables, limiting students' ability to handle data independently. This finding indicates a need for curricula that better support comprehensive data literacy. This dissertation highlights the importance of making data literacy education relevant and engaging, improving teacher training and resources, and refining curricula to include advanced data skills and digital tools. These steps are essential to prepare students for a data-driven future and ensure they can effectively analyze and interpret data.

5.2 Conclusions Across Three Studies

The synthesis of these three studies highlights some of the contextual factors that influence the teaching of analyzing and interpreting data, highlighting three central themes—relevance, training, and curriculum. First, relevance is consistently identified by the participants as crucial in data literacy education. Participants noted in Chapter 2 that incorporating community and local resources into the curriculum enhances relevance and boosts student engagement and motivation. This also aligns with the SBPM, indicating that teachers are using this as a best practice in getting students interested in working with data. This finding also aligns with literature emphasizing the importance of making science content relevant to students' lives and communities (Schwarz et al., 2017), mainly when teaching complex skills like data analysis. However, it is important to note that textbooks and curricula are generically written because publishers aim to maximize profits by appealing to a broad audience, making it difficult for educators to find a curriculum that fits with the need for relevance by using local resources (D. Sobel, 2005).

Chapter 3 highlights findings that, like other challenges teachers face, are uniquely specific to working with data. These challenges arise from the distinctive nature of science education, which often requires specialized tools and resources. For instance, Anna discussed the difficulty of conducting data analysis in Chemistry due to a lack of necessary equipment. Investigations—and thus hands-on working with data—are impossible without the proper equipment.

Teachers reported feeling unprepared for their formal education and professional development to effectively teach data analysis and interpretation. In response, many sought out additional resources and learning opportunities independently. However, they faced obstacles

such as a lack of accessible quality resources, limited knowledge of available tools, and insufficient time to collaborate with colleagues on best practices in data literacy education. Teachers need to feel confident in their practice (Dong et al., 2018; Thibaut et al., 2018a; Wright & Waxman, 2022) and have first-hand experience with content (Thibaut et al., 2017, 2018a, 2018b) to successfully integrate concepts into their practice. These findings suggest that teachers do not have the experience or background to feel prepared to include analyzing and interpreting data fully into their practice.

Additionally, time constraints present a significant challenge. Time issues are twofold: firstly, teachers lack sufficient time to plan or collaborate, which impacts their ability to enhance their data-handling skills. Without adequate preservice training or targeted professional development, teachers often turn to their peers for assistance, yet still find themselves lacking support. Secondly, there is not enough time within the curriculum to teach data skills in the depth required to reflect modern scientific practices. Teachers expressed a desire to delve deeper into data work, but standards and pacing constraints prevent them from doing so. Higher-level data work is particularly time-consuming, especially for students lacking foundational data skills. This indicates the need for early introduction and clear definition of necessary data skills at each grade level to enable students to progressively engage with data more comprehensively.

Curricular limitations play a significant role in these challenges. While curricula like OpenSciEd include data activities, they often lack comprehensive coverage of necessary skills such as using digital tools, statistical methods, and a thorough engagement with the full data cycle. This gap indicates that existing curricular materials fail to provide the necessary depth and breadth for effective data literacy instruction, aligning with the literature on curriculum and data literacy (Lee, 2010; Morris et al., 2015). In relation to the SBPM, it effectively incorporates

scaffolding and collaboration. However, it does not adequately address the practices of experiencing ambiguity, engaging with the full data cycle, and providing comprehensive experiences with data. Additionally, these curriculum inadequacies can influence teachers' practices in analyzing and interpreting data (Devetak & Vogrinc, 2013; Morris et al., 2015), highlighting the importance of improving curricula to compensate for teacher education gaps.

Given that analyzing and interpreting data is one of the eight key practices outlined by the NGSS, these findings emphasize the importance of focusing on this practice and its integration with other SEP practices in research. Its significance may have been overlooked or diminished in favor of different practices that have been more extensively researched, such as argumentation (Berland & McNeill, 2010; Grooms et al., 2015; Manz, 2015; Rowland et al., 2017) and modeling (Lehrer & Schauble, 2004; Passmore & Stewart, 2002; Schwarz et al., 2009). This research indicates there are key issues specifically related to data. One major concern is the lack of comprehensive training on how to work with data effectively. This gap is reflected in curricula that fail to provide extensive, unscaffolded data practice. Moreover, the time required to engage with data thoroughly is often insufficient in schools, where teachers must cover numerous standards. Working with data is a fundamental aspect of science and should be more prominently featured in training, resources, and curricula. The inclusion of the SBPM alignment in these studies offers a valuable guide for incorporating more rigorous data activities in 6-12 science education and identifying where teachers and curricula are already excelling in these practices. Including these practices in this research can serve as a guide for future research and provide a foundational framework for teaching data literacy in 6-12 science education.

5.3 Implications Across Three Studies

The implications of this study highlight several key areas where improvements in data literacy education can be made. First, leveraging community resources to teach data analysis and interpretation can significantly enhance student engagement and the practical applicability of these skills. Preparing students for a data-driven future necessitates that science education prioritizes data literacy as an essential competence. This requires improvements in preservice and professional development programs and the integration of more comprehensive data literacy into curricula. Teachers need access to high-quality resources and continuous professional development opportunities to stay current with the latest tools and techniques. Additionally, the tasks required of teachers need to be reassessed to prioritize collaboration and align standards with the necessary pacing for effective instruction throughout the school year.

Using community resources to teach data analysis and interpretation can be an excellent way to make activities relevant and engaging for students. By incorporating data analysis activities connected to real-world problems and local community issues, educational practices can enhance student engagement and the applicability of data literacy skills. To prepare students for a data-driven future, education systems must ensure that data literacy is desirable and essential across many sectors. By equipping students with robust data handling and critical thinking skills, educators can prepare them to meet the challenges of an increasingly complex world.

Enhancing preservice education and professional development for teachers is crucial to equip them with better strategies and comprehensive skills for teaching, analyzing, and interpreting data effectively. In addition to providing the ability to use digital tools and statistical methods, these programs should help teachers feel more comfortable with their abilities and, in

turn, more effectively teach these skills. Educational institutions should develop targeted training programs focused on data literacy to address significant gaps in preservice and in-service teacher education. These programs should emphasize the theoretical and practical aspects of teaching data analysis and interpretation. School districts and educational leaders should provide regular workshops, seminars, and collaboration opportunities to support teachers in this area.

Curricula should enhance the presentation of data analysis and interpretation and establish clear connections to more advanced data handling methods. Comprehensive data literacy integration should be a priority, ensuring students learn to handle data and understand its relevance to real-world scientific inquiry. There is a pressing need to incorporate more advanced data handling skills, such as large datasets and more sophisticated statistical methods, to prepare students for higher education and careers in data-driven fields. Curricula should include digital tools to enhance data analysis and interpretation skills, reflecting current technological advancements and anticipating future data processing and visualization trends.

Teachers need access to high-quality, modern, data-centric resources. Local, state, and federal policymakers should ensure that schools have the necessary technological and material resources to teach data literacy effectively. Schools should foster an environment that promotes collaboration not only among students but also among teachers. Sharing best practices, data literacy project ideas, and resources within a community of educators can enhance teaching strategies and outcomes.

Lastly, teachers often lack time to teach data literacy appropriately. This indicates a need to reassess educational requirements to prioritize peer collaboration and possibly refocus standards to align with the pacing necessary for the school year. This approach would allow

teachers to teach practices more comprehensively, emphasizing depth of critical skills over a broad but shallow coverage of content.

5.4 Future Directions

This dissertation effectively highlights areas for future research, offering clear directions for deeper exploration of data literacy in grades 6-12 science education. First, the sample sizes in Chapters 2 and 3 were small and limited; future research should address this by examining larger sample sizes. A larger sample size would provide more diverse responses, offering a broader perspective on teachers' approaches and challenges in teaching data analysis and interpretation. It will allow me to determine if these themes hold up over a larger sample. Ideally, this would involve more focus groups, each limited to 2-5 participants, to maintain intimacy and promote conversation.

Another approach would be to focus on a specific school or district, allowing for a more detailed examination of the contextual factors affecting that district. This could provide specific insights into how data analysis is viewed and approached in a centralized area and the challenges faced, yielding targeted local improvement strategies. Additionally, conducting focus groups with K-12 and higher education teachers would reveal similarities and differences in approaches and challenges across different grade levels.

Given the prominence of community-based resources and relevance in the approaches studied, future research should investigate how teachers incorporate place-based principles into teaching data analysis and interpretation. A study focusing directly on this could provide insights into how these principles can be widely used in teaching data literacy.

For the curriculum study in Chapter 4, a comprehensive review of the entire curriculum and its alignment with the *K-12 Framework for Science Education* (National Research Council,

2012) and the synthesized best practices model would offer a holistic view of how the curriculum addresses the goals set out in the framework over various grade spans. This would help identify how each grade level builds the foundation for the next. Comparing each level against the SBPM would also be beneficial, revealing how well the curriculum aligns with current best practices for teaching data literacy.

Additionally, examining other curricula using the same methods would allow comparisons to see how different curricula align with the goals and best practices for teaching data analysis and interpretation. Furthermore, researching a teacher using a specific curriculum to see how it translates into their practice and its benefits and challenges would provide practical insights.

In summary, future research should focus on increasing sample sizes, examining specific schools or districts, exploring place-based principles, and conducting comprehensive curriculum reviews. These steps will provide extensive insights and practical strategies for improving data literacy education.

References

- Aikenhead, G. S. (2013). Humanistic Perspectives in the Science Curriculum. *The Handbook of Research on Science Education*, 895–924. <https://doi.org/10.4324/9780203824696-36>
- Alharahsheh1, H. H., & Pius, A. (2020). A Review of key paradigms: positivism vs interpretivism. *Global Academic Journal of Humanities and Social Sciences*, 2(3), 39–43.
- Altan, T., & Cagiltay, K. (2022). An Eye Tracking Based Investigation of Multimedia Learning Design in Science Education Textbooks. *Educational Technology & Society*, 25(2), 48–61. <https://www.jstor.org/stable/48660123>
- Anderson, T., & Shattuck, J. (2012). Design-Based Research: A Decade of progress in Education Research? *Educational Researcher*, 41(1), 16–25. <https://doi.org/10.3102/0013189X11428813>
- Bandura, A., Freeman, W. H., & Lightsey, R. (1999). Self-Efficacy: the exercise of control. *Journal of Cognitive Psychotherapy*, 13(2), 158–166. <https://doi.org/10.1891/0889-8391.13.2.158>
- Banilower, E. R., Smith, P. S., Malzahn, K. A., Plumley, C. L., Gordon, E. M., & Hayes, M. L. (2018). *Report of the 2018 NSSME+*. Horizon Research Inc.
- Berland, L. K., & McNeill, K. L. (2010). A learning progression for scientific argumentation: Understanding student work and designing supportive instructional contexts. *Science Education*, 94(5), 765–793. <https://doi.org/10.1002/sce.20402>
- Berliner, D. C. (2002). Comment: Educational research: The hardest science of all. *Educational Researcher*, 31(8), 18–20. <https://doi.org/10.3102/0013189x031008018>
- Beyer, C. J., & Davis, E. A. (2009). Supporting preservice elementary teachers' critique and adaptation of science lesson plans using educative curriculum materials. *Journal of*

- Science Teacher Education*, 20(6), 517. <https://doi.org/10.1007/s10972-009-9148-5>
- Bhargava, R., Deahl, E., Letouze, E., Noonan, A., Sangokoya, D., & Shoup, N. (2015). Beyond Data Literacy: Reinventing Community Engagement and Empowerment in the Age of Data. *datapopalliance.org*.
<https://dspace.mit.edu/bitstream/handle/1721.1/123471/Beyond%20Data%20Literacy%202015.pdf>
- Bowen, M., & Bartley, A. (2014). The basics of data literacy: Helping Your Students (and You!) Make Sense of Data.
- Bowen, M., & Bartley, A. (2020). Helping students make sense of the “real world” data mess. *Science Activities: Classroom Projects and Curriculum Ideas*, 57(4), 143–153.
<https://doi.org/10.1080/00368121.2020.1843387>
- Boyatzis, R. E. (1998). *Transforming qualitative information: thematic analysis and code development*. <http://psycnet.apa.org/record/1998-08155-000>
- Busetto, L., Wick, W., & Gumbinger, C. (2020). How to use and assess qualitative research methods. *Neurological Research and Practice*, 2(1). <https://doi.org/10.1186/s42466-020-00059-z>
- Bryce, N. (2013). Textual Features and Language Demands of Primary Grade Science Textbooks: The call for more informational texts in Primary Grades. In *Springer eBooks* (pp. 101–120). https://doi.org/10.1007/978-94-007-4168-3_6
- Bybee, R. W. (2014). NGSS and the next generation of science teachers. *Journal of Science Teacher Education*, 25(2), 211–221. <https://doi.org/10.1007/s10972-014-9381-4>

- Cain, L. K., Williams, R. E., & Bradshaw, V. (2023). Establishing quality in qualitative research: trustworthiness, validity, and a lack of consensus. In *Elsevier eBooks* (pp. 336–350).
<https://doi.org/10.1016/b978-0-12-818630-5.11032-2>
- Campbell, T., Zhang, D., & Neilson, D. (2010). Model based Inquiry in the High School Physics Classroom: An exploratory study of implementation and outcomes. *Journal of Science Education and Technology*, 20(3), 258–269. <https://doi.org/10.1007/s10956-010-9251-6>
- Cobb, P., Confrey, J., diSessa, A. A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9–13.
<https://doi.org/10.3102/0013189x032001009>
- Crawford, B. A. (2007). Learning to teach science as inquiry in the rough and tumble of practice. *Journal of Research in Science Teaching*, 44(4), 613–642.
<https://doi.org/10.1002/tea.20157>
- Creswell, J. W., & Báez, J. C. (2020). *30 Essential Skills for the Qualitative Researcher*. SAGE Publications.
- Davis, E. A., Janssen, F., & Van Driel, J. (2016). Teachers and science curriculum materials: where we are and where we need to go. *Studies in Science Education*, 52(2), 127–160.
<https://doi.org/10.1080/03057267.2016.1161701>
- Devetak, I., & Vogrinc, J. (2013). The criteria for evaluating the quality of the science textbooks. In *Springer eBooks* (pp. 3–15). https://doi.org/10.1007/978-94-007-4168-3_1
- Devine, D., Fahie, D., & McGillicuddy, D. (2013). What is ‘good’ teaching? Teacher beliefs and practices about their teaching. *Irish Educational Studies*, 32(1), 83–108.
<https://doi.org/10.1080/03323315.2013.773228>
- D’Ignazio, C. (2022). Creative data literacy. *Information Design Journal*, 6–18.

<https://doi.org/10.1075/idj.23.1.03dig>

D'Ignazio, C., & Bhargava, R. (2015). Approaches to Building Big Data Literacy. In *Bloomberg Data for Good Exchange*.

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. National Geographic Books.

Dimopoulos, K., & Karamanidou, C. (2013). Towards a more epistemologically valid image of school science: revealing the textuality of school science textbooks. In *Springer eBooks* (pp. 61–77). https://doi.org/10.1007/978-94-007-4168-3_4

Doan, S. (2021). Changing the Subject: K–12 Teachers' Use of and Access to Science-Specific Instructional Materials, Feedback, and Professional Learning. RAND. Retrieved October 16, 2023, from https://www.rand.org/pubs/research_reports/RRA134-7.html

Duschl, R. A., & Bybee, R. W. (2014). Planning and carrying out investigations: an entry to learning and to teacher professional development around NGSS science and engineering practices. *International Journal of STEM Education*, 1(1). <https://doi.org/10.1186/s40594-014-0012-6>

Edelglass, W. (2000). Philosophy and place-based pedagogies. *Teaching Philosophy*, 69–80.

Edelson, D. C., Reiser, B. J., McNeill, K. L., Mohan, A., Novak, M., Mohan, L., Affolter, R., McGill, T. a. W., Bracey, Z. E. B., Noll, J. D., Kowalski, S. M., Novak, D., Lo, A. S., Landel, C. C., Krumm, A. E., Penuel, W. R., Van Horne, K., González-Howard, M., & Suárez, E. (2021). Developing Research-Based instructional Materials to Support Large-Scale Transformation of Science Teaching and Learning: The Approach of the OPENSIED Middle School Program. *Journal of Science Teacher Education*, 32(7), 780–804. <https://doi.org/10.1080/1046560x.2021.1877457>

Ed Reports. (2022). EdReports. Retrieved December 15, 2023, from

<https://www.edreports.org/reports/overview/openscienced-2023>

El-Deghaidy, H., & Mansour, N. (2015). Science Teachers' Perceptions of STEM Education:

Possibilities and Challenges. *International Journal of Learning and Teaching*.

<https://doi.org/10.18178/ijlt.1.1.51-54>

Fan, L., & Kaeley, G. S. (1998). Textbooks Use and Teaching Strategies: An Empirical Study.

Fletcher, S., & Luft, J. A. (2011). Early career secondary science teachers: A longitudinal study of beliefs in relation to field experiences. *Science Education*, 95(6), 1124–1146.

<https://doi.org/10.1002/sce.20450>

Fischer, C., Fishman, B. J., Dede, C., Eisenkraft, A., Frumin, K., Foster, B., Lawrenz, F., Levy, A. J., & McCoy, A. (2018). Investigating relationships between school context, teacher professional development, teaching practices, and student achievement in response to a nationwide science reform. *Teaching and Teacher Education*, 72, 107–121.

<https://doi.org/10.1016/j.tate.2018.02.011>

Fontichiaro, K., & Johnston, M. P. (2020). Rapid shifts in educators' perceptions of data literacy priorities. *The Journal of Media Literacy Education*, 12(3), 75–87.

<https://doi.org/10.23860/jmle-2020-12-3-7>

Garner, A. V., & Rosenberg, J. R. (2023). Utilizing iNaturalist to Support Place Based Learning and Data Analysis. *Science Scope*, 46(7), 54–60.

<https://doi.org/10.1080/08872376.2023.12315907>

Gebre, E. H. (2022). Conceptions and perspectives of data literacy in secondary education.

British Journal of Educational Technology, 53(5), 1080–1095.

<https://doi.org/10.1111/bjet.13246>

Genareo, V., Mitchell, J., Geisinger, B. & Kemis, M. (2016). University Science Partnerships:

- What Happens to STEM Interest and Confidence in Middle School and Beyond. K-12 STEM Education, 2(4), 117-127. The Institute for the Promotion of Teaching Science and Technology (IPST). Retrieved March 15, 2024 from <https://www.learntechlib.org/p/209565/>.
- Giles, J., Ryan, D. A., Belliveau, G., De Freitas, E., & Casey, R. (2006). Teaching style and learning in a quantitative classroom. *Active learning in higher education*, 7(3), 213-225.
- Gibson, J. P., & Mourad, T. (2018). The growing importance of data literacy in life science education. *American Journal of Botany*, 105(12), 1953–1956. <https://doi.org/10.1002/ajb2.1195>
- Gläser-Zikuda, M., Hagenauer, G., & Stephan, M. (2020). The potential of qualitative content analysis for empirical educational research. *DOAJ (DOAJ: Directory of Open Access Journals)*. <https://doi.org/10.17169/fqs-21.1.3443>
- Gorski, P. S. (2013). “What is Critical Realism? And Why Should You Care?” *Contemporary Sociology*, 42(5), 658–670. <https://doi.org/10.1177/0094306113499533>
- Governor, D., Bowen, M., & Brunsell, E. (2017). Big Data, small devices: Investigating the Natural World Using Real-time Data.
- Grbich, C. (2012). *Qualitative Data Analysis: An Introduction*. SAGE.
- Grimshaw, S. D. (2015). A framework for infusing authentic data experiences within statistics courses. *The American Statistician*, 69(4), 307–314. <https://doi.org/10.1080/00031305.2015.1081106>
- Grooms, J., Enderle, P., & Sampson, V. (2015). Coordinating Scientific Argumentation and the Next Generation Science Standards through Argument Driven Inquiry. *Science Educator*, 24(1), 45–50. <http://files.eric.ed.gov/fulltext/EJ1069981.pdf>

- Haag, S., & Megowan, C. (2015). Next Generation Science Standards: A National Mixed-Methods Study on Teacher Readiness. *School Science and Mathematics, 115*(8), 416–426. <https://doi.org/10.1111/ssm.12145>
- Haney, J. J., Czerniak, C. M., & Lumpe, A. T. (1996). Teacher beliefs and intentions regarding the implementation of science education reform strands. *Journal of Research in Science Teaching, 33*(9), 971–993. [https://doi.org/10.1002/\(sici\)1098-2736\(199611\)33:9](https://doi.org/10.1002/(sici)1098-2736(199611)33:9)
- Haney, J. J., & McArthur, J. (2002). Four case studies of prospective science teachers' beliefs concerning constructivist teaching practices. *Science Education, 86*(6), 783–802. <https://doi.org/10.1002/sce.10038>
- Harper, D., & Thompson, A. R. (2011). *Qualitative research methods in mental health and psychotherapy: A Guide for Students and Practitioners*. John Wiley & Sons.
- Hashweh, M. Z. (1996). Effects of science teachers' epistemological beliefs in teaching. *Journal of Research in Science Teaching, 33*(1), 47–63. [https://doi.org/10.1002/\(sici\)1098-2736\(199601\)33:1](https://doi.org/10.1002/(sici)1098-2736(199601)33:1)
- Hughes, H., & Bruce, C. S. (2012). Snapshots of informed learning: LIS and beyond. *Education for Information, 29*(3–4), 253–269. <https://doi.org/10.3233/efi-130940>
- Jones, M. G., & Carter, G. (2007). Science Teacher Attitudes and Beliefs. *Handbook of Research on Science Education, 1081–1118*. <https://doi.org/10.4324/9780203824696-43>
- Kang, E. J. S., McCarthy, M. J., & Donovan, C. (2019). Elementary Teachers' enactment of the NGSS Science and Engineering Practices. *Journal of Science Teacher Education, 30*(7), 788–814. <https://doi.org/10.1080/1046560x.2019.1630794>
- Kastens, K. (2014). Data Use in the Next Generation Science Standards [White Paper]. EDC Oceans of Data Institute. <https://oceansofdata.org/our-work/data-next-generation->

science-standards

- Kjelvik, M. K., & Schultheis, E. H. (2019). Getting Messy with Authentic Data: Exploring the Potential of Using Data from Scientific Research to Support Student Data Literacy. *CBE-Life Sciences Education*, 18(2), es2. <https://doi.org/10.1187/cbe.18-02-0023>
- Koltay, T. (2016a). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, 49(1), 3–14. <https://doi.org/10.1177/0961000615616450>
- Koltay, T. (2016b). Data governance, data literacy and the management of data quality. *IFLA Journal*, 42(4), 303–312. <https://doi.org/10.1177/0340035216672238>
- Leavy, P. (2013). Fiction as research practice: short stories, novellas, and novels. *Choice Reviews Online*, 51(04), 51–1917. <https://doi.org/10.5860/choice.51-1917>
- Lee, V. R. (2010). Adaptations and continuities in the use and design of visual representations in US middle school science textbooks. *International Journal of Science Education*, 32(8), 1099–1126. <https://doi.org/10.1080/09500690903253916>
- Lee, V. R., Wilkerson, M. H., & Lanouette, K. (2021). A Call for a Humanistic Stance Toward K–12 Data Science Education. *Educational Researcher*, 50(9), 664–672. <https://doi.org/10.3102/0013189x211048810>
- Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, 41(3), 635–679. <https://doi.org/10.3102/00028312041003635>
- Lepik, M., Grevholm, B., & Viholainen, A. (2015). Using textbooks in the mathematics classroom—the teachers’ view. *Nordic Studies in Mathematics Education*, 20(3-4), 129–156.
- Lowell, B. R., & McNeill, K. L. (2022). Changes in teachers’ beliefs: A longitudinal study of

science teachers engaging in storyline curriculum-based professional development.

Journal of Research in Science Teaching, 60(7), 1457–1487.

<https://doi.org/10.1002/tea.21839>

Luft, J. A., Navy, S. L., Wong, S. S., & Hill, K. (2022). The first 5 years of teaching science: The beliefs, knowledge, practices, and opportunities to learn of secondary science teachers.

Journal of Research in Science Teaching, 59(9), 1692–1725.

<https://doi.org/10.1002/tea.21771>

Lumpe, A. T., Haney, J. J., & Czerniak, C. M. (2000). Assessing Teachers' Beliefs about Their Science Teaching Context. *Journal of Research in Science Teaching*, 37(3), 275–292.

[https://doi.org/10.1002/\(sici\)1098-2736\(200003\)37:3](https://doi.org/10.1002/(sici)1098-2736(200003)37:3)

Manz, E. (2015). Representing student argumentation as functionally emergent from scientific activity. *Review of Educational Research*, 85(4), 553–590.

<https://doi.org/10.3102/0034654314558490>

Maybee, C., & Zilinski, L. (2015). Data informed learning: A next phase data literacy framework for higher education. *Proceedings of the Association for Information Science and*

Technology, 52(1), 1–4. <https://doi.org/10.1002/pra2.2015.1450520100108>

Mayes, R. W., Forrester, J. H., Christus, J. D. S., Peterson, F., Bonilla, R., & Yestness, N.

(2014). Quantitative Reasoning in Environmental Science: A learning progression.

International Journal of Science Education, 36(4), 635–658.

<https://doi.org/10.1080/09500693.2013.819534>

McElhaney, K. W., Baker, A. J., Chillmon, C., Kasad, Z., Liberman, B., & Roschelle, J. (2022).

An initial logic model to guide OpenSCIEd Research: Updated version.

<https://doi.org/10.51388/20.500.12265/152>

- McNeill, K. L., & Reiser, B. J. (2018). Open Source for opening minds: New OPENSIED Materials support science Standards. *Learning Professional*, 39(6), 44–48.
<https://eric.ed.gov/?id=EJ1200076>
- Miao, Y., Jong, M. S., & Dai, Y. (2022). Pedagogical Design of K-12 Artificial Intelligence Education: A Systematic review. *Sustainability*, 14(23), 15620.
<https://doi.org/10.3390/su142315620>
- Morris, B. J., Masnick, A. M., Baker, K., & Junglen, A. (2015). An analysis of data activities and instructional supports in middle school science textbooks. *International Journal of Science Education*, 37(16), 2708–2720. <https://doi.org/10.1080/09500693.2015.1101655>
- MS-LS2-1 — *The wonder of science*. (2022, October 4). The Wonder of Science.
<https://thewonderofscience.com/msls21>
- MS-PS1-2 — *The wonder of science*. (2021, September 30). The Wonder of Science.
<https://thewonderofscience.com/msps12>
- NGSS Lead States. 2013. *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.
- Next Generation Science Standards*. (2023). NGSS for All Students. Retrieved March 31, 2023, from <https://www.nextgenscience.org/>
- Nygren, T., & Guath, M. (2021). Students Evaluating and Corroborating Digital News. *Scandinavian Journal of Educational Research*, 66(4), 549–565.
<https://doi.org/10.1080/00313831.2021.1897876>
- OpenSciEd. (2023, December 20). *Homepage - OpenSCIED*. <https://www.openscienced.org/>
- Pajares, M. F. (1992). Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct. *Review of Educational Research*, 62(3), 307–332.

<https://doi.org/10.3102/00346543062003307>

Pangrazio, L., & Sefton-Green, J. (2019). The social utility of 'data literacy.' *Learning, Media and Technology*, 45(2), 208–220. <https://doi.org/10.1080/17439884.2020.1707223>

Passmore, C., & Stewart, J. (2002). A modeling approach to teaching evolutionary biology in high schools. *Journal of Research in Science Teaching*, 39(3), 185–204.

<https://doi.org/10.1002/tea.10020>

Pinfield, S., Wakeling, S., Bawden, D., & Robinson, L. (2020). *Open access in theory and practice: The theory-practice relationship and openness* (p. 256). Taylor & Francis.

PISA 2024 Strategic Vision and Direction for Science. (2020). In *oecd.org*. OECD. Retrieved \ June 17, 2023, from <https://www.oecd.org/pisa/publications/PISA-2024-Science-Strategic-Vision-Proposal.pdf>

Plummer, P. (2008). Focus group methodology Part 1: Considerations for design. *International Journal of Therapy and Rehabilitation*, 15(2), 69–73.

<https://doi.org/10.12968/ijtr.2008.15.2.28189>

Prado, J. S., & Marzal, M. Á. (2013). Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. *Libri*, 63(2). <https://doi.org/10.1515/libri-2013-0010>

Purwani, A., Rochsantiningsih, D., & Kristina, D. (2017). A Content Analysis of Bright 1: A 2013 Curriculum-Based Textbook for VII Grade Students. *Advances in Social Science, Education and Humanities Research*, 158, 88–100. <https://doi.org/10.2991/iccte-17.2017.98>

Qi, C., Zhang, X., & Huang, D. (2018). Textbook use by teachers in junior high school in relation to their role. *Research on mathematics textbooks and teachers' resources:*

Advances and issues, 29-51.

Radinsky, J., & Tabak, I. (2022). Data practices during COVID: Everyday sensemaking in a high-stakes information ecology. *British Journal of Educational Technology*.

<https://doi.org/10.1111/bjet.13252>

Raffaghelli, J. E. (2020). Is Data Literacy a Catalyst of Social Justice? A Response from Nine Data Literacy Initiatives in Higher Education. *Education Sciences*, 10(9), 233.

<https://doi.org/10.3390/educsci10090233>

Raffaghelli, J. E., Manca, S., Stewart, B., Prinsloo, P., & Sangrà, A. (2020). Supporting the development of critical data literacies in higher education: building blocks for fair data cultures in society. *International Journal of Educational Technology in Higher Education*, 17(1). <https://doi.org/10.1186/s41239-020-00235-w>

Reiser, B. J., Novak, M., McGill, T. a. W., & Penuel, W. R. (2021). Storyline Units: An Instructional Model to Support Coherence from the Students' Perspective. *Journal of Science Teacher Education*, 32(7), 805–829.

<https://doi.org/10.1080/1046560x.2021.1884784>

Rosenberg, J. M., & Koehler, M. J. (2015). Context and Technological Pedagogical Content Knowledge (TPACK): a Systematic review. *Journal of Research on Technology in Education*, 47(3), 186–210. <https://doi.org/10.1080/15391523.2015.1052663>

Rosenberg, J. M., Schultheis, E. H., Kjervik, M., Reedy, A., & Sultana, O. (2022). Big Data, big changes? The technologies and sources of data used in science classrooms. *British Journal of Educational Technology*. <https://doi.org/10.1111/bjet.13245>

Rosenberg, J. M., Edwards, A., & Chen, B. (2020). Getting Messy With Data. *The Science Teacher*, 087(05). https://doi.org/10.2505/4/tst20_087_05_30

- Rowland, A., Craig-Hare, J., Ault, M., Ellis, J. D., & Bulgren, J. A. (2017). Social media: How the next generation can practice argumentation. *Educational Media International*, 54(2), 99–111. <https://doi.org/10.1080/09523987.2017.1362818>
- Rubie-Davies, C. M., Flint, A., & McDonald, L. (2011). Teacher beliefs, teacher characteristics, and school contextual factors: What are the relationships? *British Journal of Educational Psychology*, 82(2), 270–288. <https://doi.org/10.1111/j.2044-8279.2011.02025.x>
- Savasci-Acikalin, F. (2009). Teacher beliefs and practice in science education. *Asia-Pacific Forum on Science Learning and Teaching*, 10(1).
https://www.ied.edu.hk/apfslt/download/v10_issue1_files/funda.pdf
- Schramm- Possinger, M., & Harris, L. (2021). Investigating Teacher’s Practices and Beliefs of Data Literacy to Enhance Pre-service Teacher Education. *SRATE Journal*, 30(1).
- Schultheis, E. H., & Kjervik, M. K. (2020). Using Messy, Authentic Data to Promote Data Literacy & Reveal the Nature of Science. *American Biology Teacher*, 82(7), 439–446.
<https://doi.org/10.1525/abt.2020.82.7.439>
- Schwarz, C. V., Passmore, C., & Reiser, B. J. (2017). Helping students make sense of the world using next generation science and engineering practices. NSTA Press.
- Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Acher, A., Fortus, D., Shwartz, Y., Hug, B., & Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, 46(6), 632–654. <https://doi.org/10.1002/tea.20311>
- Shreiner, T. L., & Dykes, B. M. (2020). Visualizing the teaching of data visualizations in social studies: A study of teachers’ data literacy practices, beliefs, and knowledge. *Theory and Research in Social Education*, 49(2), 262–306.

<https://doi.org/10.1080/00933104.2020.1850382>

Sobel, D. (2005). *Place-Based Education: Connecting classrooms and communities*.

<http://ci.nii.ac.jp/ncid/BA85280383>

Squire, K., MaKinster, J. G., Barnett, M., Luehmann, A. L., & Barab, S. A. (2003). Designed curriculum and local culture: Acknowledging the primacy of classroom culture. *Science Education*, 87(4), 468–489. <https://doi.org/10.1002/sce.10084>

Srikant, S., & Aggarwal, V. (2017). Introducing Data Science to School Kids. *Technical Symposium on Computer Science Education*. <https://doi.org/10.1145/3017680.3017717>

Stern, L., & Roseman, J. E. (2004). Can middle-school science textbooks help students learn important ideas? Findings from project 2061's curriculum evaluation study: Life science. *Journal of Research in Science Teaching*, 41(6), 538–568.

<https://doi.org/10.1002/tea.20019>

Tabak, I. (2004). Reconstructing context: negotiating the tension between exogenous and endogenous educational design. *Educational Psychologist*, 39(4), 225–233.

https://doi.org/10.1207/s15326985ep3904_4

Toyama, M., & Yamazaki, Y. (2020). Are there effects of a match between learning style and teaching style in an EFL classroom? *Innovation in Language Learning and Teaching*, 14(3), 243–258.

Vahey, P., Rafanan, K., Patton, C., Swan, K., Van 't Hooft, M., Kratcoski, A., & Stanford, T. (2012). A cross-disciplinary approach to teaching data literacy and proportionality. *Educational Studies in Mathematics*, 81(2), 179–205. <https://doi.org/10.1007/s10649-012-9392-z>

Wilkinson, S. (1998). Focus group methodology: a review. *International Journal of Social*

- Research Methodology*, 1(3), 181–203. <https://doi.org/10.1080/13645579.1998.10846874>
- Wolff, A., Wermelinger, M., & Petre, M. (2019). Exploring design principles for data literacy activities to support children's inquiries from complex data. *International Journal of Human-Computer Studies*, 129, 41–54. <https://doi.org/10.1016/j.ijhcs.2019.03.006>
- Xu, W. J., & Zammit, K. (2020). Applying Thematic Analysis to Education: A hybrid approach to interpreting data in practitioner research. *International Journal of Qualitative Methods*, 19, 160940692091881. <https://doi.org/10.1177/1609406920918810>

Appendix

Appendix A: How Curricula Affect Student Perspectives of Science Claims

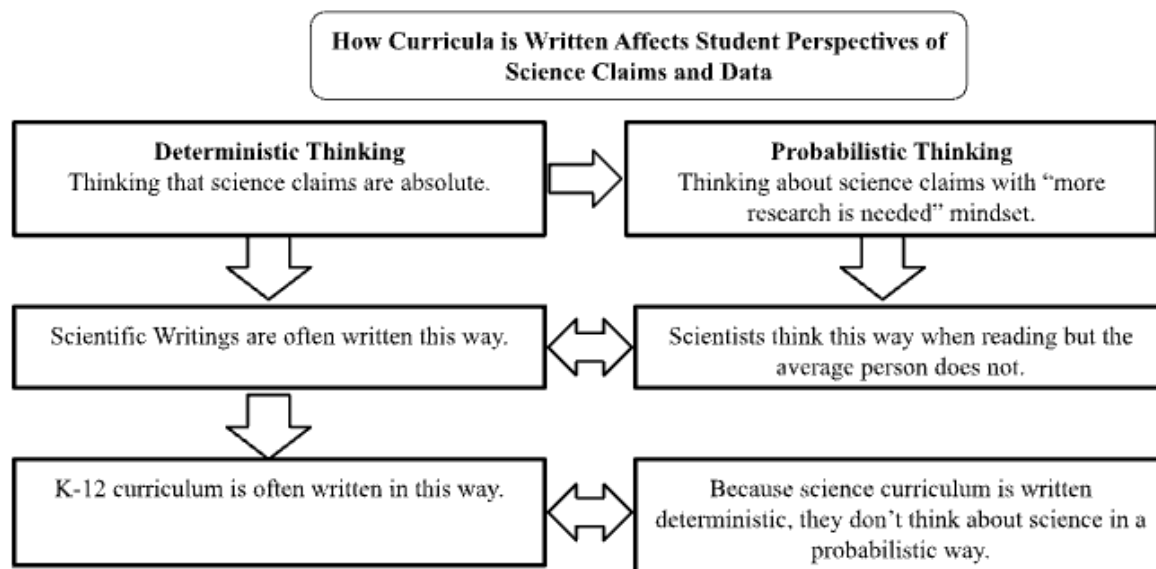


Figure A.1 Flow Chart Demonstrating Deterministic vs Probabilistic Thinking in Science Education (Bowen & Bartley, 2020)

Appendix B: Conceptual Framework Model

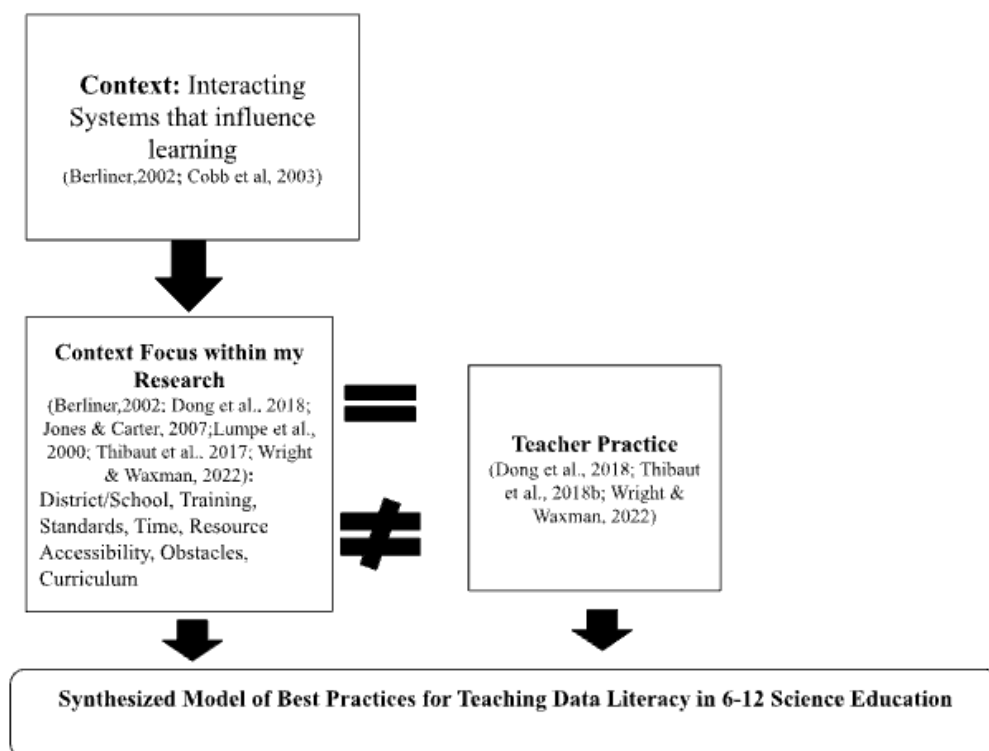


Figure B.1 *Conceptual Framework Model*

Appendix C: Frameworks for Teaching Data Literacy

Table C.1 *Levels of Science Inquiry (Bowen & Bartley, 2014, p. 7)*

Level of Inquiry	Problems	Procedures	Conclusions
0	given	given	given
1	given	given	open
2	given	open	open
3	open	open	open

Table C.2 *Data-Informed Learning Principles (Maybee & Zilinski, 2015)*

Principles	Course Example
New ways of using data must be built on students' prior experiences.	In an accounting course, the instructor may have students reflect on their experiences of balancing a checkbook and then relate that to a journal or general ledger.
Learning to use data should coincide with learning about a disciplinary subject.	A nuclear engineering course might apply the concepts of authority, quality, and accuracy to data repositories by looking up evaluated nuclear reaction data in two different databases and comparing the results.
Learning should result in students becoming aware of new ways of using data and developing new understandings of the studied subject.	A computer programming course may have students swap documented computer code with another team and rerun the script to see if they can replicate the process, having them learn about using and managing data within the context of programming.

Table C.3 *Workflow Through Data Science Tutorial Design (Srikant & Aggarwal, 2017)*

Design Elements	Defined
Full Data Cycle	Students should have hands-on exposure to the Full data cycle- collection, entry, visualizations, feature engineering, model building, model testing, and data permissions.
Relatable Data Sets	Students should be able to relate to and be interested in the data set.
Full Data Cycle	Students should be exposed to the collection and entry process. This provides greater ownership and enhancement of the activity. Avoid using prebuilt data sets.
Binary Variables	Keep data sets simple with no more than 3-4 variables.
Balanced Data Set	Make sure each independent variable is equally represented. Aim for 2 unique data points for the n independent variables.

Appendix D: Synthesized Best Practices Model

Table D.1 *Synthesized Model of Best Practices for Teaching Data Literacy in 6-12 Science Education*

Best Practices	Explained
Experience Ambiguity	Students encounter situations characterized by uncertainty and ambiguity, wherein they engage in activities that lack predefined pathways leading to specific outcomes (Bowen & Bartley, 2014; D’Ignazio & Bhargava, 2015).
Full Data Cycle	Students are engaged with aspects of the full data cycle- data acquisition and collection, data processing and cleaning, analyzing, and data visualization and communication (Avoiding curated “tidy” data) (Bowen & Bartley, 2014; D’Ignazio & Bhargava, 2015; Srikant & Aggarwal, 2017)
Relevance	Data activities are relevant to students. (Utilizing authentic data sets) (Bowen & Bartley, 2014; D’Ignazio & Bhargava, 2015; Maybee & Zilinski, 2015; Srikant & Aggarwal, 2017).
Experience	The data activities provide students opportunities to engage with and cultivate their experiences with data (Utilizing Messy data) (Bowen & Bartley, 2014; D’Ignazio & Bhargava, 2015; Maybee & Zilinski, 2015; Srikant & Aggarwal, 2017).
Scaffolding	Teachers design data-related activities that are scaffolded, taking into consideration the current proficiency levels of students and ensuring that these activities are situated within the learners' zone of proximal development (Bowen & Bartley, 2014; D’Ignazio & Bhargava, 2015; Srikant & Aggarwal, 2017).
Collaboration	Data activities provide opportunities for peer collaboration and discourse (Bowen & Bartley, 2014; D’Ignazio & Bhargava, 2015; Maybee & Zilinski, 2015; Srikant & Aggarwal, 2017).

Table D.2 *Updated SBPM for Teaching Data Literacy in 6-12 Science Education*

Best Practices	Explained
Experience Ambiguity	Students encounter situations characterized by uncertainty and ambiguity, wherein they engage in activities that lack predefined pathways leading to specific outcomes (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015).
Full Data Cycle	Students are engaged with aspects of the full data cycle- data acquisition and collection, data processing and cleaning, analyzing, and data visualization and communication (Avoiding curated “tidy” data) (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Srikant & Aggarwal, 2017)
Relevance	Data activities are relevant to students. (Utilizing authentic data sets) (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Maybee & Zilinski, 2015; Srikant & Aggarwal, 2017).
Experience	The data activities provide students with opportunities to engage with and cultivate their experiences with data (Utilizing Messy data) (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Maybee & Zilinski, 2015; Srikant & Aggarwal, 2017).
Scaffolding	Teachers design data-related activities that are scaffolded, taking into consideration the current proficiency levels of students and ensuring that these activities are situated within the learners' zone of proximal development (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Srikant & Aggarwal, 2017).
Technology	Technology should be utilized, where feasible (National Research Council, 2012), to provide students with experience using tools that facilitate working with data, including large datasets (Rosenberg et al., 2022). This approach is essential to mirror the methods employed by scientists and engineers in handling data, making the educational experience more relevant and practical.
Collaboration	Data activities provide opportunities for peer collaboration and discourse (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Maybee & Zilinski, 2015; Srikant & Aggarwal, 2017).

Appendix E: Participant Demographics for Chapters 2 and 3

Table E.1 Teacher Participants, Demographics, Focus Group Attendance

Participant Pseudonym	Gender	Race	Primary Roles	School	Focus Group Attended
Molly	Female	White	Middle grade science	Public	November 2023
Kelly	Female	White	High school science	Public	November 2023
Val	Female	White	High school STEM	Magnet Public	November 2023
Jes	Female	White	Middle grade STEM	Public	April 2024
Fay	Female	African American	Middle grade science	Public	April 2024
Dora	Female	White	Middle grade science	Public	April 2024
Anna	Female	White	High school science	Private	May 2024

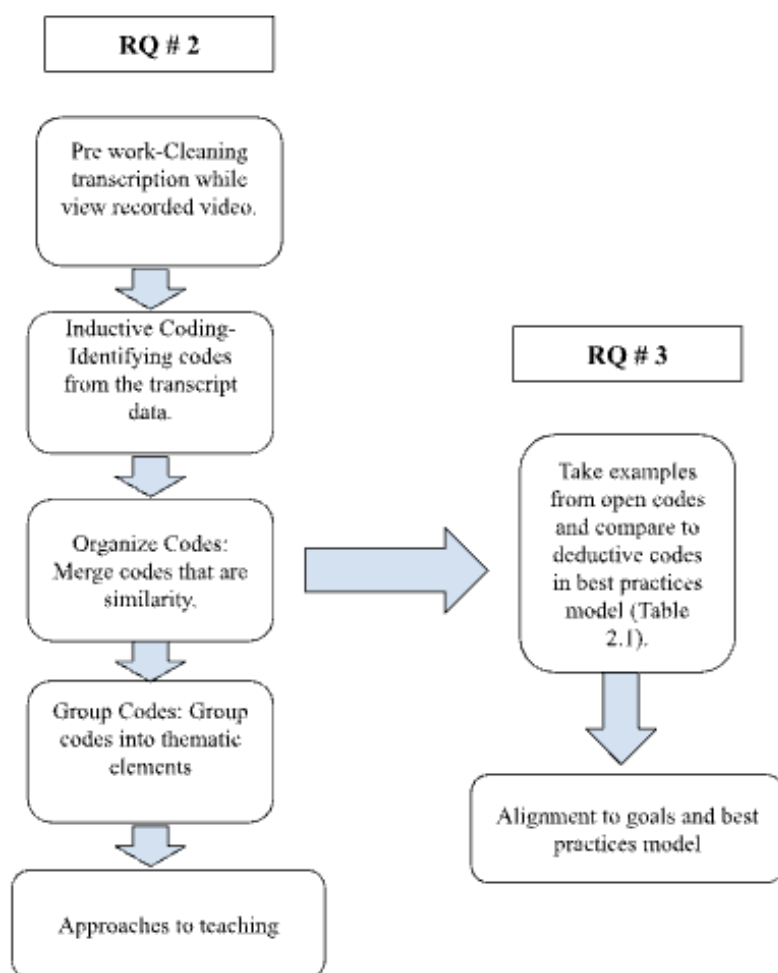
Appendix F: Modified Thematic Analysis Process

Figure F.1 *Modified Thematic Analysis Process for RQ 2 and 3 (Boyatzis, 1998)*

Appendix G: Codebooks for Chapter 2

Table G.1 *Inductive Codebook Chapter 2*

Code	Description	Examples
Basic Skills	Data Entry, basic calculations, being able to read and interpret basic visualizations and data, being able to identify basic errors in data.	“We're doing a lot of charts and graph analysis, map analysis and then kind of bringing all those together.” “But mostly we're doing a lot of charts and graph analysis and map analysis, and then kind of bringing all those together to make sense of them.”
Activities in Teaching Data	Data activities that teachers do.	“And I think something that I've done just started doing more and more in the past couple of years is the storylines for NGSS science.” “We will do labs where there's lots of analysis questions, where they are trying to connect what they saw to what we've been learning. So, for instance, just because it's fresh on my mind in chemistry. About a month ago we were talking about different types of reactions and then they were actually seeing chemical reactions.”
Scaffolded Experiences	Ways teachers scaffold data activities.	“I think at this point in the semester, we're gonna go a little more cookie cutter, and I give them the tables especially.” “We either collect it (data) ourselves or I'm finding a resource that has it pre-made.”
Relevance	The activities teachers use to get students engaged in data activities.	“The kids are definitely more engaged, obviously, when it is local.” “One thing that I do like to do is integrating local data that's already there.”
Data Literacy	How teachers define data literacy.	“Data literacy, I guess, would be them feeling comfortable looking at data table graph and making sense of it and understanding what they're looking at.” “Being able to read data in some kind of graph or chart and then use that data to make some kind of decision”
Technology	Technology that teachers use in data activities.	“We take deep dives into KGIS and looking at the topography and looking at where lines are drawn for different school zones.” “They count the species (<i>using iNaturalist</i>).”

Table G.2 *Deductive Codebook Chapter 2 Alignment with Framework (2012) Goals*

Code	Description	Examples
Systematically Analyze Data	Analyze data Systematically, either to look for salient patterns or to test whether data is consistent with initial hypothesis.	"But mostly we're doing a lot of charts and graph analysis map analysis, and then kind of bringing all those together to make sense of them in some sort of like ecological, or. you know, kind of climate change sort of way."
Data in Conflict	Recognize when data are in conflict with expectations and consider what revisions in the initial model are needed.	"Prototypes look like lots of different things, and that's one of the things that we got to. Some are a plan. Some are a program, and some are a thing. We have some people (<i>students</i>) making things, you know, they have to make sure that they have data that supports their research. They have to find data that supports the need for that (<i>prototype</i>)."
Technology and Data Practices	Use spreadsheets, databases, tables, charts, graphs, statistics, mathematics, and information and computer technology to collate, summarize, and display data and to explore relationships between variables, especially those representing input and output.	"Teaching them about tidy data and how to separate out data in their spreadsheet. We're doing a lot of charts and graph analysis, map analysis and then kind of bringing all those together."
Strength of Conclusions	Evaluate the strength of a conclusion that can be inferred from any data set, using appropriate grade-level mathematical and statistical techniques.	"I'd like to give them various forms of data that's already out there that's already visualized so that they can see here's the data in a map format, here's the data in a bar graph, here's it in a number like a dot plot or a scatter plot."
Recognize Patterns	Recognize patterns in data that suggest relationships worth investigating further. Distinguish between causal and correlational relationships.	"We're doing a lot of charts and graph analysis, map analysis and then kind of bringing all those together."
Collect from Physical Models	Collect data from physical models and analyze the performance of a design under a range of conditions.	

Table G.3 *Deductive Codebook Chapter 2 Alignment with SBPM*

Code	Description	Example
Experience Ambiguity	Students encounter situations characterized by uncertainty and ambiguity, wherein they engage in activities that lack predefined pathways leading to specific outcomes (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015).	"Sometimes the data doesn't give us a clear answer, and we have to figure out what it's trying to tell us, which is a real challenge for the kids but very real-world." "When we look at data from our experiments, it's not always straightforward; it's about interpreting what these numbers could mean in the broader context of our study."
Full Data Cycle	Students are engaged with aspects of the full data cycle- data acquisition and collection, data processing and cleaning, analyzing, and data visualization and communication (Avoiding curated "tidy" data) (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Srikant & Aggarwal, 2017).	"I have my eighth graders do some kind of inquiry where they have to gather their own data. They have to come up with their own simple, testable question and needs to be based on observable data. They have to collect it and then they have to organize it. They have to organize it in some way that that makes sense what they're trying to get. They have to create some kind of graph."
Relevance	Data activities are relevant to students. (Utilizing authentic data sets) (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Maybee & Zilinski, 2015; Srikant & Aggarwal, 2017).	"But if it has something to do with something happening locally, locally like they're way more interested." "We've done urban planning experiments. We've taken city walks. They are now building something for our city or our school."
Experience	The data activities provide students opportunities to engage with and cultivate their experiences with data (Utilizing Messy data) (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Maybee & Zilinski, 2015; Srikant & Aggarwal, 2017).	"We use a variety of data sources, not just the clean and easy ones, to show students what they'll actually face outside the classroom." "We focus on a lot of my classes. This, what is this: legit? Can you vet this data? And is it really necessary for what you're trying to back up?"

Table G.3 *Continued*

Code	Description	Example
Scaffolding	Teachers design data-related activities that are scaffolded, taking into consideration the current proficiency levels of students and ensuring that these activities are situated within the learners' zone of proximal development (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Srikant & Aggarwal, 2017).	"I start with simpler data sets and gradually introduce more complex ones as their skills improve." "I think maybe I need to deconstruct some graphs and see if we can take it back to the data and see how that looks."
Collaboration	Data activities provide opportunities for peer collaboration and discourse (Bowen & Bartley, 2014; D'Ignazio & Bhargava, 2015; Maybee & Zilinski, 2015; Srikant & Aggarwal, 2017).	"I take my kids out to the field. And then I have all the kids are working on a different area of wherever we are. And then we could just take all of that data and collectively put it together in one space." "Collaboration is key in our projects; they need to talk through their data interpretations and analyses to reach a consensus."

Appendix H: Codebook Chapter 3

Table H.1 *Inductive Codebook Chapter 3*

Code	Description	Example
Professional and Preservice Learning	Any learning that is specific to working with data.	“Even in college, you know, I don't remember a focus on it you know. Here's some data now work with it.” “I think the only training that I would receive with data would have been in a math class. It wouldn't even really be in any science methods class.”
Resources	The availability of resources in teaching working with data.	“I don't feel like I know what tools are out there to help me get them there.” “You have to have a lot of lab equipment and we are a private school, but it doesn't mean that we have large amounts of resources.”
Curriculum	What is available or not available to teachers in their provided curriculum?	“It's (<i>data</i>) just kind of random information charts and graphs based on questions, but not in with any particular subject area.” “It's (<i>district resources</i>) not available unless you're looking for it.”
Student Prior Knowledge	The previous learning or mastery of skills that students possess to do grade level appropriate work with data.	“I think that was my first thought (<i>student prior experience</i>) was like, when you put a graph up. Not that they can make sense of it. Just do you understand? Do you understand what this graph is? Do you understand the title? Do you understand what the y axis you understand what the X axis is? They don't come in and ninth grade with that either.” “You could put a put a pie chart out there, and they wouldn't be able to tell you what it really means or like how to break down percentages. They're not coming in with any of that.”

Table H.1 Continued

Code	Description	Example
Technology/Tools	Availability of tools, digital or not, to do data activities.	“We have a little school set at class set of iPads that we borrow from our school, and we'll take those with us to the park and we'll take photos with them, and then the photos will save that geolocation. Then when we come back into Wi-fi, we can upload it to INaturalist. But I feel like I still get more data collected from kids when I do the paper pencil version.” “I've pulled back in the last couple of years of how much I'm using (<i>Google Suite</i>) and I'm actually giving them a hard copy.”
Standards	That standards teachers must follow to teach data in science education.	“I, personally don't have a whole lot of knowledge of how it (<i>data</i>) is supposed to be taught within NGSS.” “I think the science engineering practices are a little elusive.”
Time	When time impacts the ability to teach about data or collaboration and planning.	“I have to do science in math. I have to do social studies during math. And I have to do reading during social studies, you know, like I have to. Otherwise, there's just not enough time to get it all to fit in, so that you just have to start doing a lot of crossover.” “If we don't get through all this, then you know, these kids have paid 90 something dollars for an exam that they don't get to use as college credits. So, between having the resource but not having the time and just having students that want to find the quickest answer.”

Table H.1 Continued

Code	Description	Example
School/District Policies	The policies, procedures and supports that impact teaching about data.	“What department would it (<i>teaching about data</i>) fall under? When I'm thinking about secondary schools like t, is this a science department problem? Is this a math department problem like, who's in charge of making sure that that information is being taught. It's a gray area.”
Teacher Agency	Teacher feels comfortable and confident in the teaching of data practices.	“I think that's probably why I do it (<i>teach about data</i>). I like I it. I make it a priority, but not every teacher's going to be like that.” “I'd just like to see, step by step, how to do it. I would like something that says, Oh, so now you have data. Now, what? What? How do? How do you walk your students through those steps to make it useful. I need that kind of instruction. I wish that I had a mentor teacher who could do that on a very simple level.”
Student Engagement	The ways in which students are or are not engaged in data activities.	“I get I get work back that's not what it could be. You know it'll be like two words instead of a paragraph or something, you know, not like what you think they could do if they really tried but just the easy way to put something down, to have to be done.” “There are just times where it's difficult to come up with creative ways that engage them and make them get into to reading a text or do something like that.”

Appendix I Themes for Chapters 2 and 3

Table I.1 *Themes from RQ #2*

Theme	Defined
Community-Based Activities and Relevance	Teachers use community-based activities to make the lessons more relevant for students.
Basic Skills and Curated Experiences	Basic skills with data, such as reading and interpreting graphs and providing students with curated data sets and experiences to scaffold this learning.
Integration of Technology in Teaching Data	Technology is used to support data activities, such as iNaturalist, to support data activities.

Table I.2 *Themes Defined*

Theme	Defined
Insufficient Training	Lack of preservice or in-service professional development training in analyzing and interpreting data.
Curriculum	Teachers either did not have a curriculum, or it was inadequate for their needs.
Time	Lack of time to collaborate with colleagues or plan.
Resource Availability	Lack of resources to teach analyzing and interpreting data, whether it is equipment, lessons, data, or knowing where to find them.
Student Prior Knowledge and Engagement	Students lack the foundational knowledge of analyzing and interpreting data and their lack of motivation to participate in data activities.

Appendix J: Codes and Definitions for Chapter 4 Content Analysis

Table J.1 *RQ 1 Codes and Definitions for Content Analysis*

RQ 1: What characterizes the nature of the data utilized by students in the OpenSciEd 7th-grade curriculum?			
Code	Definition	Research Question	Best Practices Alignment
Data Visualization: <i>Graphs, Tables, Maps, Infographics, Other</i>	Graphs: This includes, but is not limited to, line, pie, bars, histograms, and scatter plots. Tables: This is any table that includes organized data. Maps: This includes any map type. Infographic: This is any visualization that includes the transfer of data information to the reader.	RQ 1B	Full Data Cycle, Relevance, Experience, Scaffolding
Visualization Type: <i>Already-Created, Student Creates, Created Blank, Other</i>	Already Created: The visualization is complete and provided by the developers. Student Created: This is where students create the visualization from existing or collected data. Created Blank: The visualization is empty and ready for students to add data.	RQ 1B	Full Data Cycle, Scaffolding, Experience, Collaboration
Type of Data: <i>Raw, Primary Source, Made Up, Simulation-Based, Sensor-Based, Other</i>	Raw: Data in its original form, generally in a spreadsheet, and has not been cleaned. Primary Source: This is data cited from a primary source; this also includes student-gathered data such as a science article. Made up: Data that is not from research but is created to simulate actual data collected in order for students to work with it. Simulation-Based: Data from a computer-based model or simulation. Sensor Based: Data from probe ware such as Vernier.	RQ 1A	Full Data Cycle, Relevance, Experience, Collaboration
Student collected: <i>Yes, No, Other</i>	Yes: Students are asked to collect data No: The data are not collected by students	RQ 1D	Full Data Cycle, Relevance, Collaboration
Size of data set: <i>Small, Medium, Large, Other</i>	Small: The table contains fewer than 20 rows, 2 columns; or less than 40 cells in total. Medium: The table contains 20-100 rows and 2-4 columns, or 40-400 cells in total. Large: The table contains more than 100 rows and 5 columns, or more than 401 cells in total. Other: Other: The size of the data set is ambiguous, such as an infographic.	RQ 1C	Scaffolding, Experience

Table J.2 RQ 2 Codes and Definitions for Content Analysis

RQ 2- How is data analysis incorporated within the OpenSciEd curriculum at the seventh-grade level?			
Code	Definition	Research Question	Best Practices Alignment
How is it used by students? Do they need to: Collect, Organize, Analyze, Interpret, Other	Collect: Students are responsible for an investigation where they collect the data. Organize: Students take raw data and "clean" it up by putting it in tables, graphs, maps, etc. Analyze: Students look for patterns in data. Interpret: Students connect patterns to phenomena.	RQ 2A	Full Data Cycle, Relevance
What is the variable type? Nominal, Ordinal, Interval-Ratio	Nominal: Collected from two unrelated independent variables. (Compare cats and dogs or eye color, gender, etc.) Ordinal: collected from independent variables that are related but not in a way that it is measured (Small, medium, large. Medium is not 2 X larger than small, other examples: undergraduate, graduate) Interval- Ratio Data: independent variables that are direct measures (Distance from 15-20 is the same as 25-30)	RQ 2A	Full Data Cycle, Scaffolding, Experience
Research method: Quantitative, Qualitative, Both, Other	Qualitative (Nominal, Ordinal): students collect qualitative data such as color, observations, etc. Quantitative (Interval-Ratio): Students collect numerical data. Both: Students do both Quantitative and Qualitative in the same activity. Other: the type of research method is unclear.	RQ 2A	Full Data Cycle, Experience

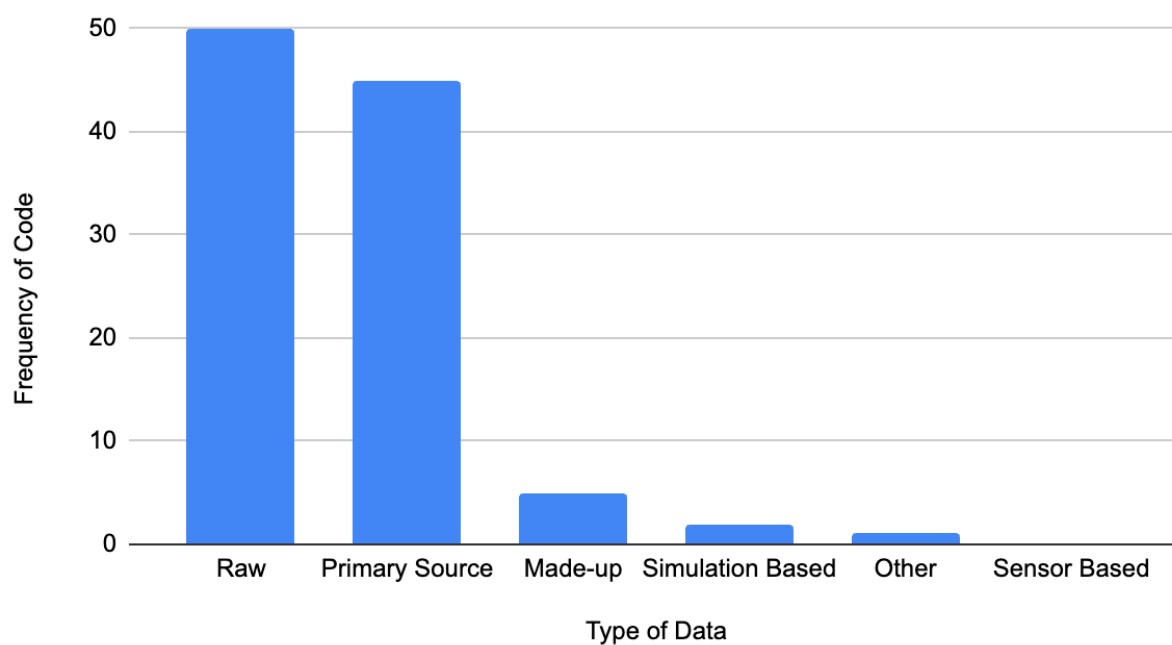
Appendix K: Occurrence of Each Code Chapter 4

Table K.1 Occurrence of Each Code for RQ 1

RQ 1: What characterizes the nature of the data utilized by students in the OpenSciEd 7th-grade curriculum?			
Code	Definition	Research Question	Occurrences of Code
Data Visualization: <i>Graphs, Tables, Maps, Infographics, Other</i>	Graphs: This includes, but is not limited to, line, pie, bars, histograms, and scatter plots. Tables: This is any table that includes organized data. Maps: This includes any map type. Infographic: This is any visualization that includes the transfer of data information to the reader.	RQ 1B	Tables- 74 Graphs-30 Infographics-6 Other-6 Maps- 5
Visualization Type: <i>Already-Created, Student Creates, Created Blank, Other</i>	Already Created: The visualization is complete and provided by the developers. Student Created: This is where students create the visualization from existing or collected data. Created Blank: The visualization is empty and ready for students to add data.	RQ 1B	Already Created-64 Created Blank- 44 Student Creates- 10 Other- 2
Type of Data: <i>Raw, Primary Source, Made Up, Simulation-Based, Sensor-Based, Other</i>	Raw: Data in its original form, generally in a spreadsheet, and has not been cleaned. Primary Source: This is data cited from a primary source; this also includes student-gathered data such as a science article. Made up: Data that is not from research but is created to simulate actual data collected in order for students to work with it. Simulation-Based: Data from a computer-based model or simulation. Sensor Based: Data from probe ware such as Vernier.	RQ 1A	Raw-50 Primary Source- 45 Made-up-5 Simulation-Based-2 Sensor-Based-0 Other-1
Student collected: <i>Yes, No, Other</i>	Yes: Students are asked to collect data No: The data are not collected by students	RQ 1D	Yes-48 No-69
Size of data set: <i>Small, Medium, Large, Other</i>	Small: The table contains fewer than 20 rows, 2 columns; or less than 40 cells in total. Medium: The table contains 20-100 rows and 2-4 columns, or 40-400 cells in total. Large: The table contains more than 100 rows and 5 columns, or more than 401 cells in total. Other: The size of the data set is ambiguous, such as an infographic.	RQ 1C	Small- 54 Medium- 34 Large-1

Table K.2 *Occurrences of each code for RQ 2*

RQ 2- How is data analysis incorporated within the OpenSciEd curriculum at the seventh-grade level?			
Code	Definition	Research Question	Occurrences of Code
How is it used by students? Do they need to: Collect, Organize, Analyze, Interpret, Other	Collect: Students are responsible for an investigation where they collect the data. Organize: Students take raw data and "clean" it up by putting it in tables, graphs, maps, etc. Analyze: Students look for patterns in data. Interpret: Students connect patterns to phenomena.	RQ 2A	(Combinations of tasks) Collect, Organize, Analyze, Interpret- 26 Analyze, Interpret-44 Analyze-16 Collect, Analyze, Interpret-10 Collect - 7 Collect, Organize-4 Collect, Organize, Analyze-2 Interpret-3
What is the variable type? Nominal, Ordinal, Interval-Ratio	Nominal: Collected from two unrelated independent variables. (Compare cats and dogs or eye color, gender, etc.) Ordinal: collected from independent variables that are related but not in a way that it is measured (Small, medium, large. Medium is not 2 X larger than small, other examples: undergraduate, graduate) Interval-Ratio Data: independent variables that are direct measures (Distance from 15-20 is the same as 25-30)	RQ 2A	Nominal-36 Ordinal-23 Interval-Ratio-8
Research method: Quantitative, Qualitative, Both, Other	Qualitative (Nominal, Ordinal): students collect qualitative data such as color, observations, etc. Quantitative (Interval-Ratio): Students collect numerical data. Both: Students do both Quantitative and Qualitative in the same activity. Other: the type of research method is unclear.	RQ 2A	Quantitative- 25 Qualitative-34 Both-7

Appendix L: Figures of Findings Chapter 4**Type of Data Included in Curriculum****Figure L.1** *Type of Data Included in Curriculum*

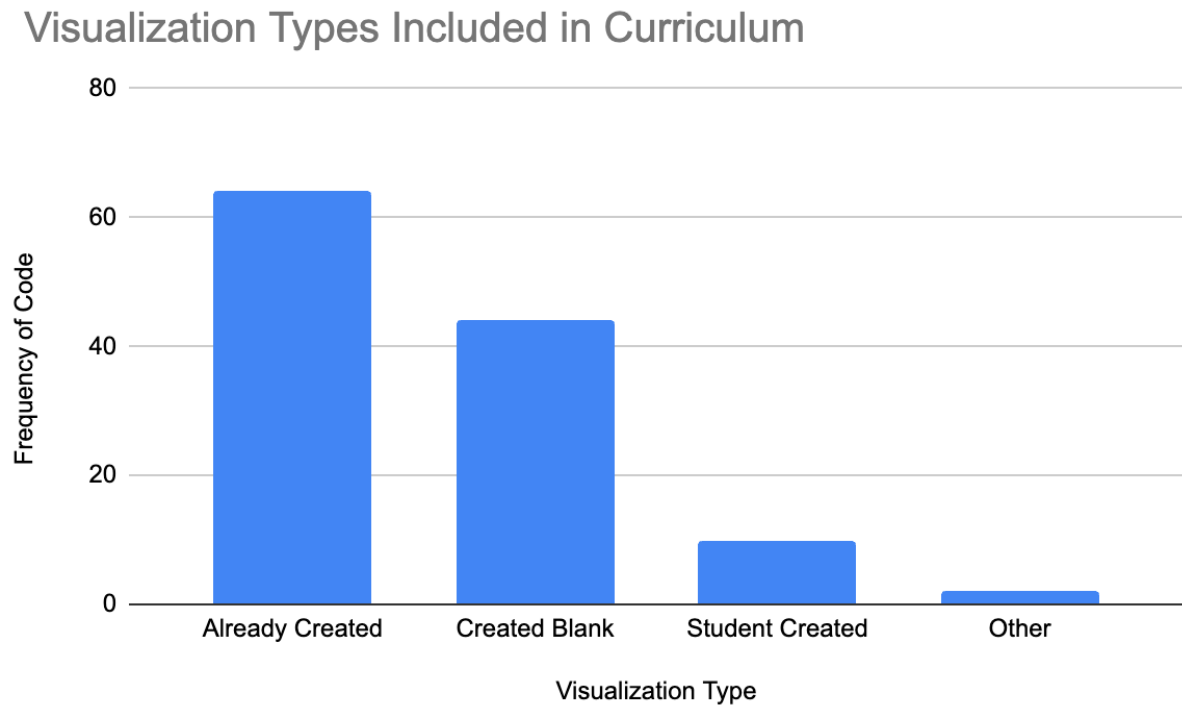


Figure L.2 *Type of Visualizations Included in Curriculum*

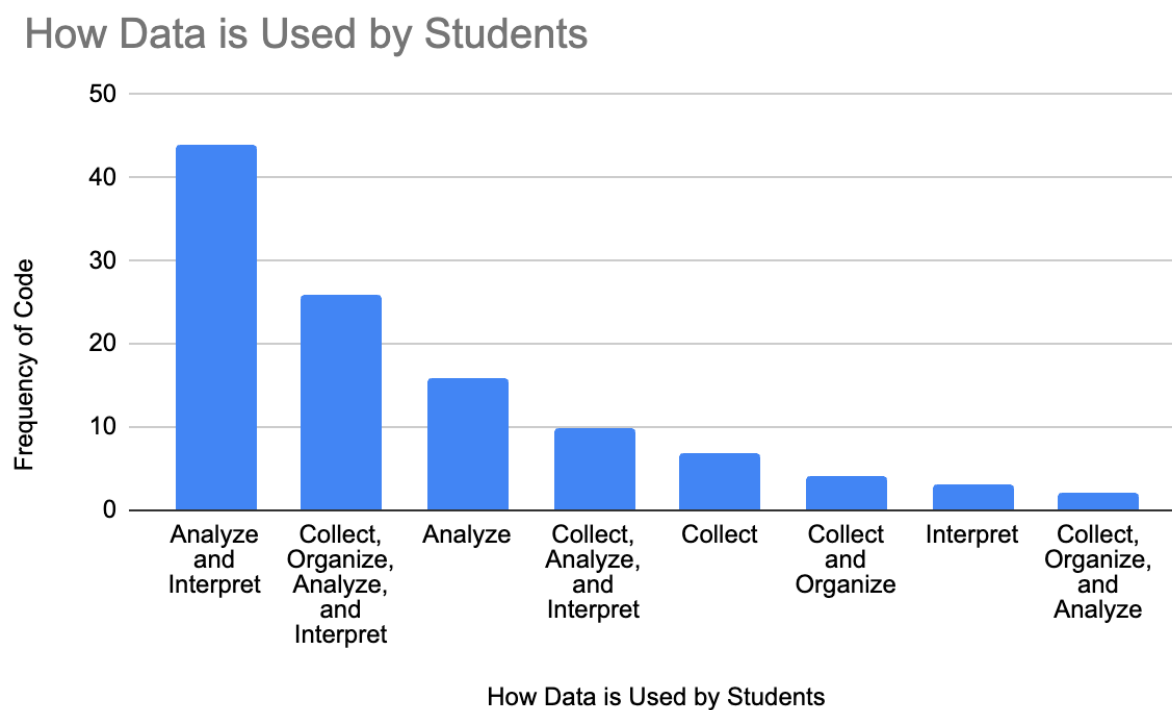


Figure L.3. *How Data is Used by Students*

Vita

Amanda was born and raised in Walland, Tennessee, nestled in the picturesque foothills of the Great Smoky Mountains. This unique environment instilled in her a profound appreciation for nature and the local culture, shaping her educational and research pursuits. Amanda attended Tennessee Technological University, where she earned a Bachelor of Science in Human Learning K-8. Her passion for education and dedication to improving teaching practices led her to further her studies at Tusculum University, where she obtained a Master of Arts in Curriculum and Instruction. Amanda's research interests are deeply influenced by her upbringing and experiences. She is particularly focused on data literacy in K-12 science education, aiming to enhance students' abilities to understand and use scientific data effectively. Additionally, she is passionate about place-based education, exploring how integrating local environmental and cultural contexts into science curricula can enrich the learning experience. Amanda also delves into the concept of Appalachian identity, investigating how the unique cultural heritage of the Appalachian region influences educational practices and student engagement. Through her work, Amanda strives to bridge the gap between local culture and educational practices, fostering a more inclusive and relevant learning environment for students in the Appalachian region and beyond.