AUTHENTIC DATA IN HIGH SCHOOL MARINE SCIENCE: IMPACTS ON DATA LITERACY AND REAL-WORLD CONNECTIONS

by

Caroline Elizabeth Bottega

A professional paper submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Science Education

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2024

DEDICATION

To my family for showing me the beauties of the natural world and supporting me as I continue to follow my curiosities and passions, most often on high elevation trails with a very heavy backpack.

To all of the teachers and professors I have learned from, who have motivated me to grow as an educator, and allowed me to realize that the greatest impact you can have on a student is to inspire.

To my fellow teachers, as we navigate this challenging and unique profession, for their support as we continue to put our whole selves into our work for the sake of helping others to grow as learners and individuals.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Data literacy is becoming increasingly relevant in science education as educators work to prepare students for success beyond the classroom. Despite its importance, data literacy is often overlooked in science curricula, depriving students of valuable opportunities to engage with real-world data and develop a deeper appreciation for the relevance of science. The study discusses changes in high school students' data literacy skills, confidence in data literacy, and connection to science following lessons that incorporated real-world, data-driven case studies in marine science. To assess changes, pre- and post-treatment tests and surveys were used as data collection instruments and analyzed using quantitative and qualitative methods. While student test scores did not statistically change following the treatment period, student surveys reflected increases in confidence in data analysis and deeper connections between real-world research and classroom learning. With repeated implementation, data literacy skills can be developed and should be reinforced throughout a students' learning career. Further, with real-world data and case studies, students can develop a deeper understanding of the applicability of classroom learning to real-world problems.

CHAPTER ONE

INTRODUCTION & BACKGROUND

Context of the Study

Throughout my teaching and educational career, I have interacted with students and science content at nearly every level, from graduate-level field data collection to communicating place-based science content to fourth graders. No matter the grade level or the setting, I have noticed the power that a real-world connection to science can play on student interest. Equally as powerful is the importance of proper communication of scientific information, particularly through data. Data literacy is becoming increasingly relevant as a core, cross-disciplinary skill, yet is often lacking in students, as structured implementation is not yet widespread in science curriculum (Wolff et al., 2017). One way data literacy can be incorporated into classroom science is through the use of real-world data, which has been shown to increase student interest in science beyond the classroom (Schultheis & Kjelvik, 2015). My goal as an educator is to make science applicable and relatable beyond the classroom. Therefore, this action research (AR) project sought to determine the impact on student's data literacy skills and their capacity to connect scientific concepts to real-world applications through the use of authentic, real-world data in the classroom.

I currently teach general biology, marine science, AP environmental science, and climate science at The Hun School of Princeton, a private boarding and day school for grades 6-12. I have the opportunity to interact with students across all high school grade levels, as general biology is a 9th and 10th-grade course, while marine science, AP Environmental Science, and climate science are 11th and 12th-grade science courses. Hun serves primarily New Jersey and Pennsylvania residents but with a boarding community of over 170 students the population is composed of students from around the world. The school is located in a predominantly affluent area, just two miles away from Princeton University and within the town of Princeton where the median household income is $177,000 (U.S. Census Bureau, 2022). The student body is predominantly White, making up nearly 60% of the population. The minority population is mostly comprised of Asian students within the boarding community (Hun By The Numbers, n.d.). With an acceptance rate of 35%, most students are of a high academic caliber. The Hun School is one of many private boarding schools in the central New Jersey area. Therefore, the competition to attract the best and brightest students is rigorous.

For the purposes of this study, I focused my attention on my marine science classes. The classes are separated by topic into three trimester electives that are offered throughout the academic year. During the implementation of my capstone, students were enrolled in the course titled Humans and the Marine Environment, which covers topics such as climate change, ocean acidification, and overfishing. Classes typically have 16 students and are comprised of students from 10^{th} grade to post-graduate students. This lab-based science, trimester elective is unique in that students often take the course as alternative to higher-level science courses. There are no prerequisites, therefore, students in the course do not have strong math or science foundations. Notably, students struggle when interacting with scientific data and case studies, which are used frequently in the course in order to teach key concepts and topics.

Even in the later years of their high school journey, I have observed a prevalent deficiency in fundamental data literacy skills among many of my students. Data literacy is a foundational skill that can continue to be built upon throughout a student's science education and can range from basic graph construction to higher-level evidence-based reasoning, hypotheses development, and data interpretation (Risdale et al., 2015). In an increasingly data-abundant world, students must be exposed to and feel comfortable understanding various types of data and the messages communicated through data. Therefore, with this research, I sought to reinforce data literacy skills through the use of real-world data and research case studies to reinforce the connection between classroom content and real-world implementation of science.

Focus Question

My focus question was, How will the use and manipulation of authentic data impact students' scientific data literacy skills and their capacity to connect scientific concepts to real-world science applications?

My sub-questions include the following:

1. Does repeated manipulation and analysis of authentic data increase students' confidence in their data literacy skills?
2. Does the use of authentic data from real case studies increase students' understanding of the application of science beyond the classroom?

CHAPTER 2

CONCEPTUAL FRAMEWORK

The Growing Relevance of Data Literacy in Science Education

Data literacy is becoming increasingly relevant in science education as educators work to prepare students for success beyond the classroom. Historically, there has been a strong emphasis placed on the need for students to be highly skilled in language literacy (Kjelvik & Schultheis, 2019). However, many U.S. students are not adequately prepared for the STEM needs of today's society and are lagging behind other developed countries (Schultheis & Kjelvik, 2015). In an ever-changing world, the collection, use, and availability of data are becoming increasingly common, both in everyday life and particularly in science (Kjelvik & Schultheis, 2019). With such extensive and widely available databases, the use of data in the workplace necessitates proficiency in data literacy (Schultheis et al., 2022). Therefore, the growing relevance of data in society and science necessitates the need for students to develop robust data literacy skills as well.

Data literacy does not have a concrete or fixed definition among educators and researchers. Broadly, a data-literate student possesses the necessary quantitative and analytical skills that allow them to identify, collect, organize, evaluate, communicate, and transform data into actionable knowledge (Gibson & Mourad, 2018; Mandinach & Gummer, 2013; Ridsdale, 2015; Wolff et al., 2017). Each of these data literacy skills can be further broken down into specific abilities, such as developing hypotheses, data interpretation, and formulating methods for further action (Mandinach & Gummer, 2013). For students to properly develop and utilize data literacy skills, it is in the hands of the educator to implement a curriculum that specifically fosters mastery of such skills. Research indicates that the foundation of a data-literate student begins in the classroom; therefore, data literacy must be incorporated into the school curriculum (Wolff et al., 2017). However, the higher-order thinking skills, such as critical thinking or problem-solving, can be challenging to teach and may require professional development, utilization of new technologies, or support and collaboration from peers and administration. All of these factors require time and act as a barrier to implementation in the classroom (Risdale et al., 2015). Knowing the importance of data literacy as a cross-disciplinary skill, its relevance will only continue to grow. Therefore, it is essential that the challenges of implementation be addressed and data literacy skills be integrated into the classroom in engaging and relevant ways.

Challenges to Data Literacy Implementation in Science Classrooms

To keep content relevant and students adequately prepared for the future, educators are challenged with the task of adapting curriculum to incorporate data literacy skills in combination with content knowledge. However, many teachers do not feel well-equipped or properly prepared to incorporate data literacy into science curricula (Miller, 2021). Despite the richness and diversity of datasets available online today, a study published in the *British Journal of Educational Technology* found a gap between the data available to educators and the actual use of the data in the classroom (Rosenberg et al., 2022). Rosenberg et al. (2022) suggested that one reason for such a gap between available data and classroom use may be that teachers view such datasets as cumbersome, time-consuming, and a distraction away from content, rather than a means to supplement content knowledge. Further, it was concluded that educators were more likely to use familiar digital analysis tools, rather than learning to use new analysis methods as, again, new methods would take time to learn. As such, students are missing valuable opportunities to interact with real-world data and educators struggle to incorporate data literacy skills into science curriculum.

Strategies for Data Literacy Development in the Classroom

While studies and evidence of improvements in student data literacy exist, defined classroom strategies for proper implementation are less clear. Gibson and Mourad's 2018 paper, *Importance of Data Literacy in Life Science*, summarized discussions among biology educators at the University of Oklahoma, aiming to establish best practices for organized enhancement of data literacy in the classroom. While recognizing the absence of a singular skill development pathway, the discussions emphasized the importance of identifying specific data skills or concepts and linking them to learning outcomes. Moreover, Gibson and Mourad (2018) proposed readily available data-focused modules, like HHMI's BioInteractive's Data Points, as easy opportunities for easy integration into yearly curricula. The second key aspect of educator discussions underscored the necessity of departmental collaboration and administrative support. Risdale et al. (2015) emphasized the importance of collaboration and recommended online resources for seamless curriculum integration, offering an extensive list of data-focused ready-to-use materials. Although additional professional development for educators may be required, leveraging existing resources and guidance from conferences and professionals can aid in overcoming challenges associated with integrating data literacy into classrooms, allowing teachers to introduce students to essential data literacy practices and techniques.

Connection Between Classroom Content and Real-World Application

Utilizing freely available scientific data fosters a vital link between classroom learning and real-world application. Research, including studies by Woolley et al. (2013) and Archer et al. (2013), highlights the importance of connecting classroom science content with professional applications to enhance student engagement and success. Despite student interest in STEM subjects, disengagement often occurs when the relevance to future careers is unclear. Faulconer et al. (2020) and Maass and Engeln (2019) support this view, emphasizing the necessity of real-world connections for students to perceive the applicability of their learning in society. The Australian Council for Educational Research (ACER, 2007) underscores the need for educators to address how the nature of science interacts with society, enabling students to see the active role science plays in various aspects of life. Woolley et al.'s (2013) longitudinal study corroborates ACER's findings, demonstrating increased student engagement when classroom learning aligns with real-world applications and careers. In light of this research, the centrality of authenticity and relevance in science curricula becomes evident, emphasizing the importance of educators incorporating science through a more authentic lens, focusing on real-world application and integration. This approach allows educators to develop strategies that naturally engage students in content that might otherwise be forgotten.

Using authentic data in science curricula can be employed as an approach to enhance students' connection to science and integrate the relevance of science beyond the classroom. Authentic data are classified as “quantitative or qualitative information collected from real-life phenomena” (Kjelvik & Schultheis, 2019, p. 2). Research suggests that authenticity in content and making real-world connections with content increases student performance and motivation to learn (Hulleman & Harackiewicz, 2009). Hulleman and Harackiewicz's (2009) study of 262 high school science students utilized intentional motivational interventions in lessons designed to enhance the relevance of science content to students' lives, ultimately showing increased interest and performance in science. Following the implementation of this revelatory study, further research has supported the findings. Erwin’s (2015) study of real-world learning through the use of real-world data found that students felt more committed to a task when they knew the task and background were authentic in nature. With the understanding of the role of real-world data in the classroom, educators are encouraged to incorporate authentic data to increase interest in science content.

Using Authentic Data to Increase Science Data Literacy Skills

The use of authentic data in the classroom not only promotes interest, motivation, and engagement in STEM, but can also be used as an opportunity to incorporate data literacy skills. Kjelvik and Schultheis (2019) are at the forefront of the integration of authentic data and data literacy skills. In their most recent paper, Kjelvik and Schultheis reiterate the intrinsic link between data literacy and authentic data by stating, “the use of these authentic data sets in the classroom has the unique potential to develop student data literacy and draw out connections between quantitative reasoning and data science” (p. 2). Additional studies support this statement and find that when interacting with authentic data students have a greater opportunity to practice critical thinking skills, data analysis techniques, and engage in evidence-based arguments, all skills that a student with strong data literacy skills should possess (Gould et al., 2014; Holmes et al., 2015; Kerlin et al., 2010). Researchers have concluded that data literacy skills improve when working with authentic data, suggesting a greater effort should be placed on using authentic data manipulation as a medium for practicing data literacy skills. Therefore, with the manipulation of authentic data, students will be able to build a toolkit of data literacy skills that can be applied across a variety of disciplines, while simultaneously building a deeper connection with science content.

Incorporating Authentic Data in the Classroom

In the classroom, students typically encounter authentic data through various sources such as online databases, textbook tables, and figures, or data collected by students themselves. Risdale et al. (2015) underscores the use of real-world data as an opportunity not only to inspire students but also to integrate a diverse set of data literacy skills. An exemplary educational tool that achieves this is Data Nuggets, a free online resource created by Elizabeth H. Schultheis and Melissa K. Kjelvik from Michigan State University. Data Nuggets employs real-world data and research stories to facilitate the practice of quantitative skills and engagement with the scientific process (Schultheis & Kjelvik, 2015). Each lesson within Data Nuggets includes relevant research background information, a research question and hypothesis, practice in graphing and interpreting data, and an opportunity to consider further research. The overarching aim is to involve students in applying science beyond the classroom, hone essential data literacy skills, and foster enthusiasm for real-world scientific applications.

As resources like Data Nuggets become more abundant, they serve as prime examples of how data literacy skills can be seamlessly integrated into classroom content, providing an authentic perspective that not only cultivates essential data literacy skills but also heightens students' interest in and connection to the subject matter. Schultheis and Kjelvik (2019) suggest implementation strategies, including scaffolding data complexity, and emphasize the ongoing need for resource development and educator training. They also highlight the necessity for further research to identify specific learning outcomes and advance best practices for enhancing data literacy in STEM. This approach not only taps into students' natural curiosity about the world but also deepens their connection to classroom material, all while honing crucial data manipulation skills essential for success in an increasingly data-rich future.

Resources such as Data Nuggets are becoming increasingly plentiful and serve as a prime example of how data literacy skills can be woven into classroom content through an authentic lens, which not only develops essential data literacy skills but increases interest in and connection to classroom content. By incorporating data skills and science content in this way, students are more likely to develop connections between data and their everyday lives. Further, students will engage their natural curiosity about the world and develop a deeper connection to the classroom material, all while practicing essential data manipulation skills that will prepare them to excel in an inevitably data-rich future.

Conclusion

As educators strive to equip students for success beyond the classroom, the relevance of data literacy is growing. In response to the expanding field of big data, recent science education standards prioritize data literacy skills, emphasizing the need for curricula to include data analysis, interpretation, student-led investigations, and evidence-based explanations (Rosenberg et al., 2022). To support these standards, various classroom activities and digital learning platforms, leveraging authentic data and integrating real research stories, are being developed (Kjelvik & Schultheis, 2019). Despite challenges in data literacy implementation, the availability and user-friendliness of resources are improving and educators are witnessing a positive shift, as they strive to build data literacy into curriculum and bring science concepts to life through real-world connections.

CHAPTER THREE

METHODOLOGY

Demographics

The goal of my action research was to determine the impact on students' scientific data literacy skills and their capacity to connect scientific concepts to real-world science applications through the use and manipulation of authentic data. In an increasingly data-driven world, data literacy is an important cross-disciplinary skill that students often lack. Particularly in science, students struggle to understand data and see the application of science concepts beyond the classroom. With my mixed method research design, I utilized data from authentic case studies to assess changes in students' data literacy skills and connection to real-world science practices. By teaching data literacy skills through the medium of authentic data, my research was able to accomplish two goals: (1) to assess changes in student data literacy skills, and (2) build connection between classroom topics and real-world science practices.

I conducted my research with two classes of high school marine science students at The Hun School of Princeton, a mid-sized private boarding school in Princeton, New Jersey (*N*=29). One section held 14 students, while the other held 15 students. The course, titled Humans and the Marine Environment, was a trimester-long elective course comprised primarily of junior, senior, and post-graduate students. Students will take the marine science course as an elective as an alternative to a higher-level science course, like physics or Advanced Placement as there are no prerequisites to the marine courses. Generally, students in the marine science course do not have strong math or science foundations. In particular, students struggle when interacting with scientific data and case studies. For my quasi-experimental design, I implemented the same treatment across both classes and assessed changes before and after the treatment period, with the pre-treatment results acting as quasi-control. The research methodology for this project received an exemption by Montana State University's Institutional Review Board and compliance for working with human subjects was maintained (Appendix A).

Treatment

Due to the structure of the trimester schedule in my school, I began teaching the course three weeks prior to the start of my capstone project. Prior to implementation, students were introduced to the various ways in which human activities have impacted marine ecosystems, with a focus on microplastics and overfishing. Lessons were taught through more traditional methods in which students learned about broad concepts without real-world context or data to supplement lessons. Since my capstone was developed with a quasi-experimental design and relied heavily on pre- and post-treatment data collection and analysis, as seen in my research methodology cycle (Figure 1).

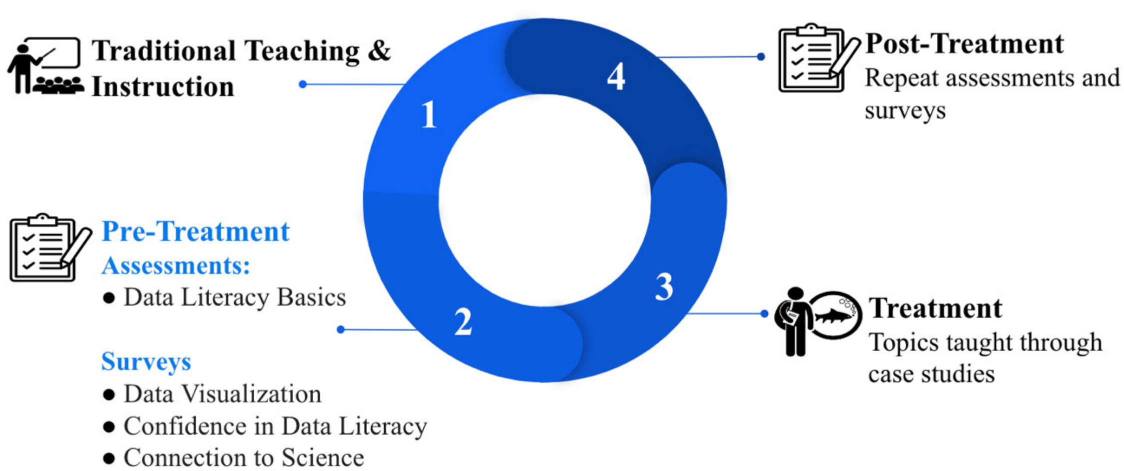


Figure 1. Research methodology cycle.

I needed to ensure that my pre-assessment data could be used as a quasi-control or baseline before treatment, against which to compare the post-assessment results (Edmonds & Kennedy, 2017). Therefore, prior to implementation, I intentionally kept data analysis activities limited to ensure that my pre-treatment data truly assessed students' baseline knowledge of data literacy skills coming into the course. Similarly, I did not utilize real-world case studies in my lessons prior to the capstone treatment to ensure that the prior lessons would not influence the pre-assessment results. By starting the course teaching in a more traditional manner, I aimed to keep conditions as consistent as possible prior to the start of my study.

I began treatment by distributing the series of pre-assessment instruments to both classes. Over the course of one class period, students were given the Data Literacy Basics (Appendix B), Confidence in Data Literacy (Appendix E), and Connection to Science (Appendix F) assessments. Students completed the surveys independently in class. The following class period, I distributed the open-ended survey portion of the pre-assessment. In a similar format, students were given the Data Visualization Tools (Appendix G) survey to complete independently in class. Each data collection method gave me deep insight into students' prior data literacy abilities and connection to science before teaching my lessons and served as a baseline for comparison at the end of the treatment period.

During the following weeks of the treatment period, my lessons focused on topics of climate change, ocean acidification, and coral bleaching. At the start of the series of classes, I introduced students to the Graphing Checklist (Appendix D) and reinforced skills outlined in the Data Literacy Skills Table (Appendix C). Data related specifically to climate change, ocean acidification, and coral bleaching were used to weave in real world application of classroom concepts and provide opportunities for students to interact with authentic datasets. With this strategy, students began to build a connection to science beyond the classroom and see the different ways in which data is used to communicate information in science. In the following classes, the lesson topics were taught through case studies and students learned about the Claim Evidence Reasoning (CER) structure of analysis. Students had the opportunity to practice CER analysis repeatedly and were provided with feedback based on the CER Rubric (Appendix H).

Towards the end of the treatment period, students were asked to demonstrate their new skills and understandings of data literacy and real-world science practices through a final presentation. The presentation required students to pick a real-world case study related to marine science and communicate the research in a clear and complete manner, supported by real-world data related to the scientific study and a clear CER statement. The specific goal of the presentation was to see how well students could implement what they learned during the treatment period and students were assessed using a rubric that specifically analyzed skills relevant to the research (Appendix J).

My capstone project concluded with the distribution of the post-assessment and open-ended survey questions. Students took the same assessments and surveys in the same environment prior to the treatment to maintain consistency between the pre- and post-treatment results (Edmonds & Kennedy, 2017). The post-treatment results supported the results from the case study presentation and confirmed any changes in student ability over the course of the treatment period.

Data Collection and Analysis Strategies

Since my capstone research was developed with a mixed-method design, both qualitative and quantitative data were collected and analyzed. With this design, qualitative and quantitative data worked collaboratively to interpret results in an approach known as data triangulation. While most of data collection was quantitative in nature, the qualitative data was collected and used to further explain the quantitative data (Mertler, 2017).

Data Collection Methods

Table 1. Data Triangulation Matrix.

		Focus Questions		
Data Collection Instruments		How will the use and manipulation of authentic data impact students' scientific data literacy skills and their capacity to connect scientific concepts to real-world science applications?	Does repeated manipulation and analysis of authentic data increase students' confidence in their data literacy skills?	Does the use of authentic data from real case studies increase students' understanding of the application of science beyond the classroom?
Pre-Assessment & Post-Assessment	Data Literacy Basics	X	X	
	Confidence in Data Literacy	X	X	X
	Connection to Science			X
Open-Ended Survey	Data Visualization	X	X	
Observation	Presentation Rubric		X	X

Pre- & Post-Assessment: Data Literacy Basics. The first of three assessments that I distributed to students at the start and end of the treatment period was the Data Literacy Basics assessment (Appendix B). The goal of this instrument was to establish a baseline of student data literacy skills prior to treatment and was used again at the end of the treatment to track changes in student performance. Pre- and post-assessment design is commonly used in educational research to assess changes over a treatment period (Spector, 1981). The Data Literacy Basics assessment consisted of fifteen multiple-choice questions, each with four options, and was broken into two sections: (1) Data Literacy Essentials, and (2) Data in the Real World. Each question was aligned with a targeted data literacy skill that was being assessed in the study (Appendix C). Quantitative data was collected from this instrument in the form of mean score based on strengths and weaknesses in the skills areas that were assessed.

Pre- & Post-Assessment: Confidence in Data Literacy. Similarly, the Confidence in Data Literacy assessment (Appendix E) was used as a pre- and post-treatment instrument and collected quantitative data on students' confidence manipulating data both in and out of the classroom. The instrument was comprised of ten questions and developed with a Likert design. Likert surveys are a commonly used research design in which participants assess their level of agreement on a scale ranging from *Strongly Agree* to *Strongly Disagree* (Mertler, 2017). Further, Likert surveys are commonly used to assess student attitude in educational research to analyze changes in attitude over time (DeCuir-Gunby & Schutz, 2017). The Confidence in Data Literacy assessment not only revealed how confident students were in their own data literacy skills in science but also their ability to translate those skills to other disciplines and real-world applications. This instrument supported the important research goal of assessing students' understandings of skill application beyond the classroom.

The goal of the Confidence in Data Literacy assessment was to assess student confidence in data literacy at the start of the treatment period, then redistribute the assessment at the end of the treatment period to assess for changes in students' confidence when manipulating, analyzing, and communicating data. For each question, the survey offered four options: *Strongly Disagree, Disagree, Agree,* and *Strongly Agree*. These responses were recoded as numerical scores (1-4) and analyzed using descriptive statistics. The results provided deeper insight into how well students felt they learned and absorbed the added information taught during the treatment period.

<u>Pre- & Post-Assessment: Connection to Science.</u> The last of the three pre- and post-assessment instruments used was the Connection to Science (Appendix F) assessment, which collected quantitative data on students' ability to understand how concepts taught in class can be translated to real world science. This instrument aligned with a key goal of my research, which was to determine student connection to science, and was designed in the same Likert style as the Confidence in Data Literacy assessment. The instrument was comprised of ten statements, each with four options in which participants assess their level of agreement on a scale ranging from *Strongly Agree* to *Strongly Disagree*.

The goal of this instrument was to learn more about how students viewed science both in and out of the classroom and if they were able to apply their classroom learning to solve real-world problems. As Byrne (2023) outlines in the SAGE Project Planner, the quasi-experimental design that I used for my study requires a benchmark with which to compare later results. Like the two previous pre- and post-assessment instruments, this Connection to Science assessment aligned with my quasi-experimental design. This instrument was distributed at the start and at the end of the treatment period, so that I might gain insight on how my students attitudes changed over the course of the implementation of the treatment period.

Open-Ended Survey: Data Visualization. The Data Visualization survey (Appendix G) collected qualitative data regarding student data literacy skill level and confidence. While the multiple-choice style of questioning used in the pre- and post-assessment instruments directly aligned with the goals of the study, those instruments alone were not enough to understand the reason behind students' attitudes and abilities. In qualitative research, written reflection and open response questioning provides a deeper understanding of how individuals process and make decisions (Merriam, 2002). The Data Visualization survey provided the opportunity for students to justify their choices in a free response section, which allowed me to contextualize their quantitative results in a deeper way.

The Data Visualization survey was also given before and after the study period to determine changes in student data literacy skills and familiarity with particular data visualization tools. Students were given six different graph types and asked to select an option that described their familiarity with the graph, if they had seen that style of graph before, and if they felt comfortable interpreting the information on the graph. After selecting an option, students were asked to elaborate upon their interpretation of the graph and identify any key elements to the best of their ability. The open-ended style of questioning allowed for students to provide reasoning for their choices and offered greater insight on their level of data literacy beyond simple multiple-choice questions. After analyzing the open responses thematically and aligning reoccurring words and phrases with the Data Literacy Skills (Appendix C) targeted in my study, I was able to understand students' data literacy strengths and weaknesses in a more nuanced manner and piece out knowledge gaps in a more refined manner.

Observation: Presentation Rubric. The final data-collecting instrument in my capstone was designed as a final research project in which students worked in groups to present authentic data from a real-world case study related to marine science. Communication of data is an essential benchmark of data literacy proficiency and a skill that I included in my set of Data Literacy Skills (Appendix C). Since communication is an essential component in a students' data literacy toolkit, the presentation project emphasizes students' abilities to communicate all that they learned throughout the treatment period, from data manipulation to real-world science applications (Siddig et al., 2016).

The presentation rubric (Appendix I) aligns with all my capstone goals: (1) proper manipulation and interpretation of data, (2) use of authentic research, and (3) connection between classroom topics and real-world application. Students received quantitative scores based on the provided rubric and the data was analyzed quantitatively through mean scores. Scores were further broken down into two main categories, which aligned with my research goals: (1) Data Literacy Skills and (2) Connection to Science.

Analysis Strategies

With the mixed-methods design of my capstone research, both qualitative and quantitative data were collected and analyzed. In this study, both data types held equal value and supported one another to address my research goals (Mertler, 2017). The data collection methods and analysis strategies worked collaboratively to determine if student data literacy skills and connection to science increased after the lessons presented throughout the research period.

Quantitative Data Analysis Techniques. Quantitative data was collected across most of my research instruments. All three of my pre-assessment and post-assessment instruments collected quantitative data. However, analysis of the quantitative data differed across the suite of instruments. The Data Literacy Basics assessment collected student scores, from which I conducted *t*-tests. The *t*-test analysis technique is often used in research to assess before and after measurements (Denny et al., 2017). More specifically, the scores on the Data Literacy Basics assessment were analyzed using a paired-tailed *t*-test, as Denny et al. (2017) explained that a paired *t*-test is best used to compare pre-assessment and post-assessment results from the same set of individuals and reveal whether significant changes occurred during treatment. The analysis of the Data Literacy Skills pre- and post-assessment data was helpful in my study to reveal how well students' data literacy improved after the implementation of the data literacy lessons and practice assessments.

The Confidence in Data Literacy Skills and Connection to Science pre- and post-assessment instruments collected quantitative data through Likert survey scores. Since the instrument collected data on a four-point scale, ranging from *Strongly Agree* to *Strongly Disagree*, the data was analyzed and organized by category to obtain descriptive statistics. Both pre- and post- treatment assessments data were analyzed using this technique. Chi-squared goodness of fit test was utilized to determine whether any measurable changes in data literacy skills or attitudes towards connection to science had occurred as a result of my treatment (Mears et al., 2020).

The Data Visualization survey collected both qualitative and quantitative data that assessed student's prior experience with specific types of data visualization. Quantitative data was collected in this instrument when students were asked to select a response that fit their familiarity with a graph in question. Like the Likert survey instruments, the responses were organized by category in order to obtain descriptive statistics and chi-square tests were utilized to assess for measurable changes between pre- and post-treatment collection.

Finally, quantitative data was collected from the Case Study Presentation Rubric following students' final presentations in the form of average scores. For this instrument, only average scores were collected and there was no pre-treatment data collected from which to compare the results. However, in keeping with the spirit of triangulation, multiple methods and instruments work collaboratively to support a final theory or conclusion (Flick, 2011). Therefore, the results from the final presentation instrument worked to support the findings of other data collection instruments, such as any measurable changes in data literacy skills or connections to science found through chi-square tests or *t*-tests.

Qualitative Data Analysis Techniques. Qualitative data was collected in two of my research instruments. The first was through the Data Visualization Tool survey, which collected both qualitative and quantitative data. The survey assessed student's prior experience with specific types of data visualization, asked them to select a response that fit their interpretation of that particular piece of data, then asked them to elaborate on their understanding with a free response statement. The free response section asked students to identify as many pieces of the tool as possible, so responses were analyzed for key words and phrases that aligned with correct interpretations of the graph. The occurrence of correct phrases and answers were separated by category and analyzed for reoccurring statements and common themes.

CHAPTER FOUR

DATA ANALYSIS

The purpose of this action research was to examine changes in high school students' data literacy skills, confidence in data literacy, and connection to science following lessons that incorporated real-world, data-driven case studies in marine science. To assess changes, pre- and post- treatment assessments and surveys were used as data collection instruments and analyzed using quantitative and qualitative methods. Additional data collection instruments included a final case study presentation, which were assessed through a presentation rubric. The results of the pre- and post-treatment assessments and surveys indicated whether lessons intentionally taught with a focus on data analysis and interpretation would lead to improvements in students' data literacy skills and connection to science. While student assessment scores did not statistically change following the treatment period, student surveys reflected increases in confidence in data analysis and deeper connections between real-world research and classroom learning.

Impact on Student Data Literacy Skills

Data Literacy Skills Assessment Results

To assess changes in student data literacy skills, students completed pre- and post-treatment assessments and surveys that examined various types of data literacy proficiency: (a) basic data literacy skills, (b) familiarity with common data visualization forms, and (c) data interpretation. Students first completed the Data Literacy Skills assessment (Appendix C). Upon analysis of the average scores, there was not a statistically significant increase ($p < .05$) between the pre- and post- treatment assessment (Figure 2).

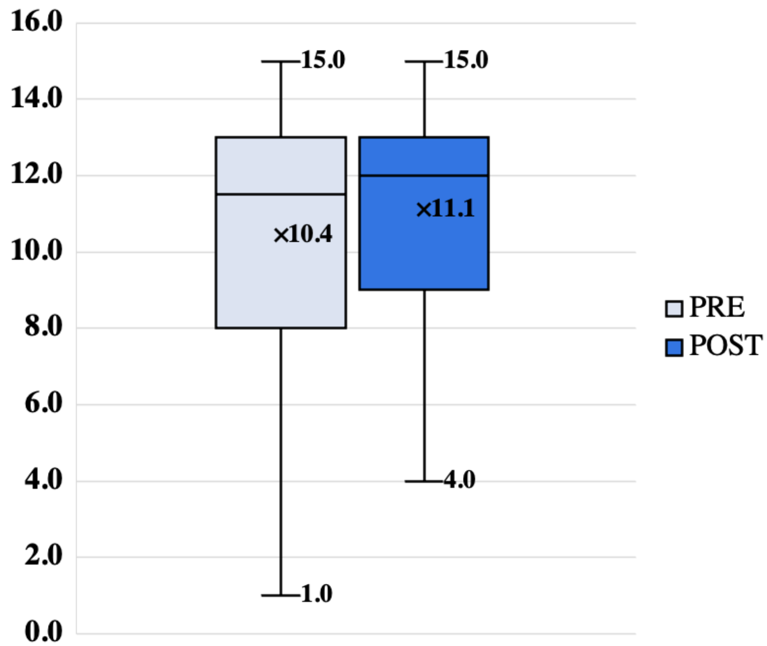


Figure 2. Score distributions of the data literacy skills pre- and post- assessment, (N=28).

As seen in Figure 2, the mean pre- treatment score for the data literacy assessment was 10.4, while the mean post- treatment score was 11.1, with a maximum possible score of 15 points. These data were subjected to a paired *t*-test, which show a statistically insignificant gain, $t(28)=2.05$, $p=0.55$. It is interesting to note that while the increase in the assessment average was not significant, the standard deviation for the pre- treatment assessment (SD=3.11) tightened in the post- treatment assessment (SD=2.62). A smaller standard deviation signifies a tighter distribution of scores, indicating that more students achieved scores closer to the average (University of Waterloo, 2023). Although the change in the mean was not statistically significant, the tightening standard deviations reveal progress among lower-scoring students.

Data Visualization Likert Survey Results

Students completed a pre- and post- treatment survey regarding familiarity with common data visualizations to reveal a more detailed understanding of students' data literacy proficiency (Appendix G). Students were asked to select an option that aligned with their familiarity with common data visualization. Of the six graph types assessed, students were more familiar and more confident in their ability to analyze four of the six graph types (scatter plot, candlestick chart, line graph with multiple variables, and line graph with one variable) after the course treatment. For example, students gained the most familiarity with the multiple variable line graph. Student responses increased notably from only 11% responding that they were familiar with and could interpret the graph before treatment to 50% following treatment. (Figure 3).

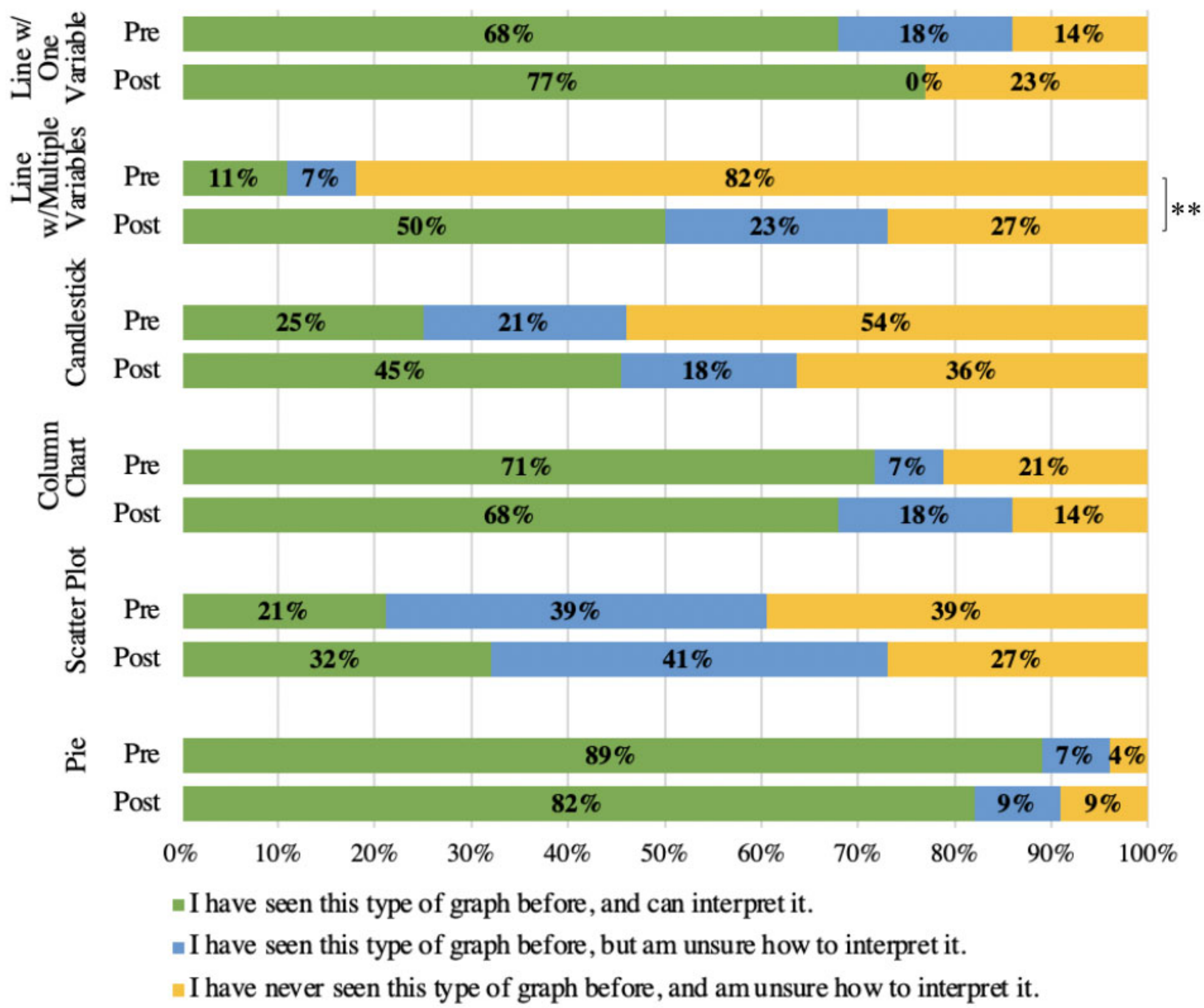


Figure 3. Results of the data visualization tools pre- and post- treatment Likert survey, (N=28). *Note*. **$p < .01$.

A Chi-Squared Goodness of Fit test was conducted to determine if the changes between the pre- and post- treatment responses were significant. This analysis technique is often used to compare responses between an expected population, the pre- treatment group, and an observed population, the post- treatment group (Mears et al., 2020). Upon analysis of the pre- and post- treatment responses, the only statically significant increase in familiarity was visible in the line graph with multiple variables $X^2(2, N=28) = 9.33$, $p = < .01$. Considering the high responses of familiarity with pie, column, and line charts, it is likely that students possessed some foundational knowledge of data visualizations prior to the course.

Data Visualization Open-Ended Survey Results

To gain insight into students' thought processes when analyzing different graph types, students were asked to elaborate on their level of understanding of the visualization through open-ended responses. Responses were analyzed through thematic analysis and separated into four main data interpretation proficiency categories. In the pre- treatment survey, many responses fell in line with a theme of misinterpretation (6%) and uncertainty about how to interpret the graph (46%). Several responses among the pre-treatment survey began with the phrase, "I don't know how to interpret this graph," which mirrors the lack of proficiency among students before the treatment. When analyzing the post- treatment responses, there were some notable shifts in student responses. More student responses fell into the *Moderate Proficiency* category, increasing from 29% to 48% of responses and *High Proficiency* responses saw a slight increase as well, increasing from 19% to 24% (Figure 4).

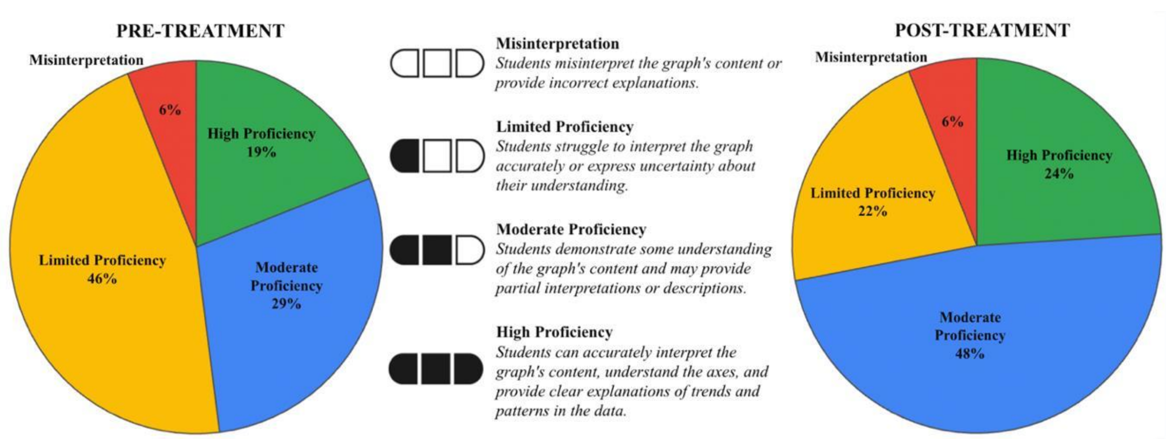


Figure 4. Thematic analysis results of the data visualization tools open-ended survey, (*N*=28).

Although subtle, the open-ended responses mirror the increases found in the quantitative analysis of the data visualization survey and support the modest increases in data literacy skill proficiency.

Statistically significant increases in specific areas of data literacy were found before and after the treatment period. While the average score on the data literacy skills assessment was not statistically significant between the start and end of the treatment period, data literacy skills were further broken down and analyzed in the data visualization survey. The visualization survey revealed that students did possess some prior knowledge about data visualization and analysis techniques and revealed some gaps in their knowledge. Upon analysis of the qualitative and quantitative analysis of the visualization survey, students reported statistically significant increases in familiarity and confidence interpreting multiple line graphs. Further, through thematic analysis of open-ended responses, data interpretation proficiency increased slightly throughout the treatment period.

Impact on Student Confidence in Data Literacy

Confidence in Data Literacy Likert Survey Results

Although it is helpful to understand how students' data literacy skills contribute to short-term comprehension of material, assessing students' attitudes towards data analysis is crucial for understanding their future interactions with data. Student responses were collected to assess the impact of repeated data manipulation on student confidence in various data literacy skills areas, such as data analysis, interpretation, and communication. In the Confidence in Data Skills Likert survey (Appendix E), students were asked to select an option that reflected their level of agreement to statements regarding their confidence working with data in various capacities.

Student responses regarding confidence increased in four of the five question areas (Figure 5). The most significant areas of confidence development were in students' abilities to create data visualizations and communicate results clearly to their peers.

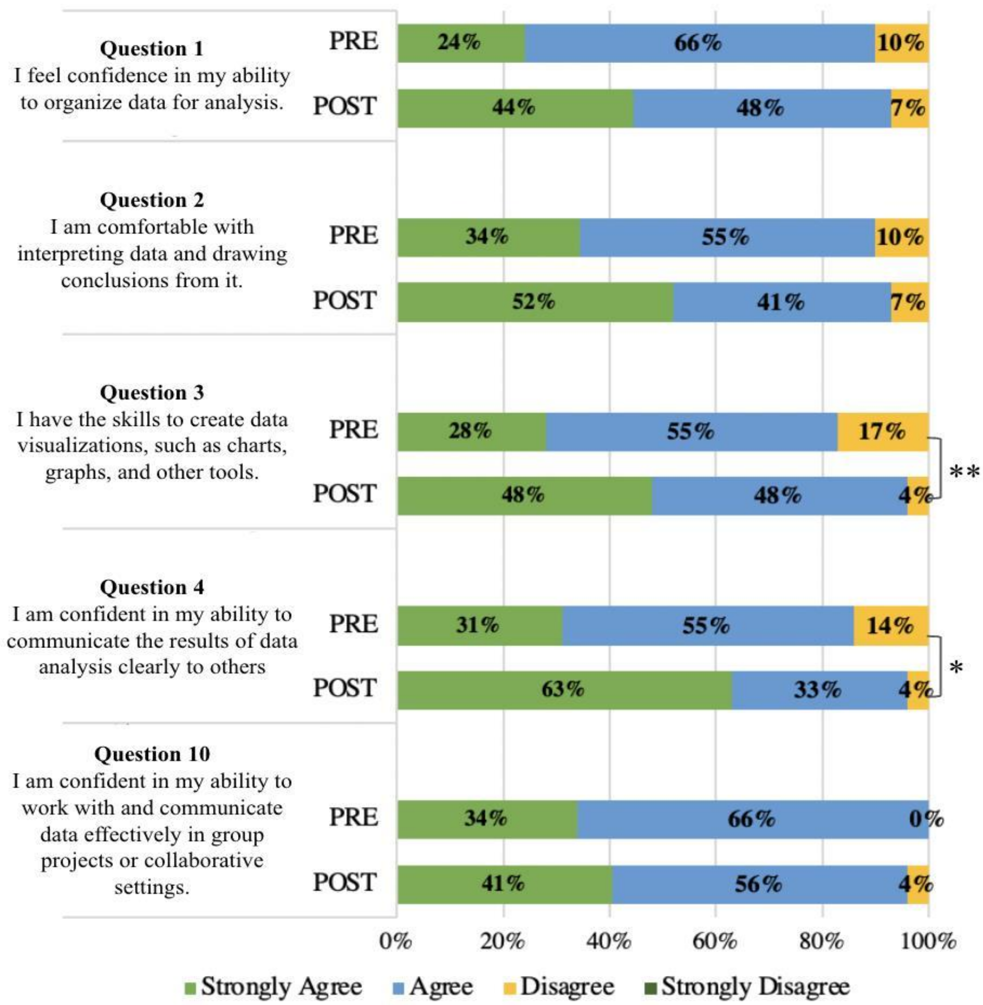
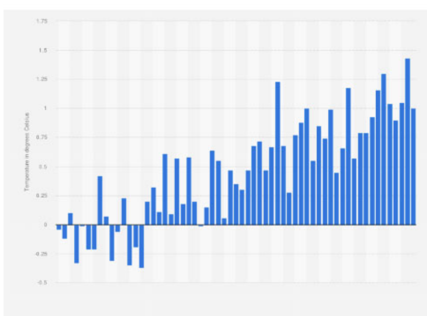


Figure 5. Results of the confidence in data literacy pre- and post- treatment Likert survey questions 1-4 & 10, (N=29). *Note*. *$p < .05$. **$p < .01$.

A Chi-Squared Goodness of Fit test was conducted for each question to determine if the changes between the pre- and post- treatment responses were significant. Responses to questions 3 and 4 saw statistically significant increases in student confidence, X^2 (2, N=29) = 7.25, $p < .05$. The questions targeted confidence in data manipulation as well as data communication. While student self-reported responses showed significant increases in areas of data literacy, the validity of these claims was supported by the student final project presentations.

Case Study Data Presentation Communication Results

Student success in the area of data communication was further supported by students' final presentation projects and write-ups, which were scored using the Case Study Data Presentation Rubric (Appendix I). Each student group successfully developed a Claim-Evidence-Reasoning statement that aligned with the CER Rubric (Appendix H). During the presentation, students clearly communicated the graphs and figures related to their chosen marine case study (Figure 6).



This is a bar graph that represents the water mean temperature variation in Puerto Rico from 1961-2021. As we can see on the graph, over the course of 60 years, the water temperature increases significantly. The reason why the water keeps getting warmer is because the atmosphere is trapping hot air on the earth and being absorbed into the ocean.

Figure 6. Sample of student work from the case study data presentation project write up.

Considering the quality and clarity of the presentations alongside the Confidence in Data Literacy survey, students exhibited a greater confidence and assurance in their ability to analyze and communicate real-world scientific findings throughout the treatment period.

Impact on Student Connection to Real-World Science

Connection to Science Likert Survey Results

Lastly, the impacts on students' connection to science was assessed following exposure to real-world marine science research case studies and manipulation of authentic datasets. The Confidence in Data Skills survey (Appendix F) focused on gauging students' perceptions of how relevant their classroom lessons were to real-world scenarios and whether they believed the skills learned in class were applicable in and out of the scientific field. Student responses regarding connection to science increased for nine out of the ten questions. (Figure 7). The most significant areas of development in building science connections were students' abilities to see the relevance and application of science in their daily lives, even if they did not have intentions of studying science in the future.

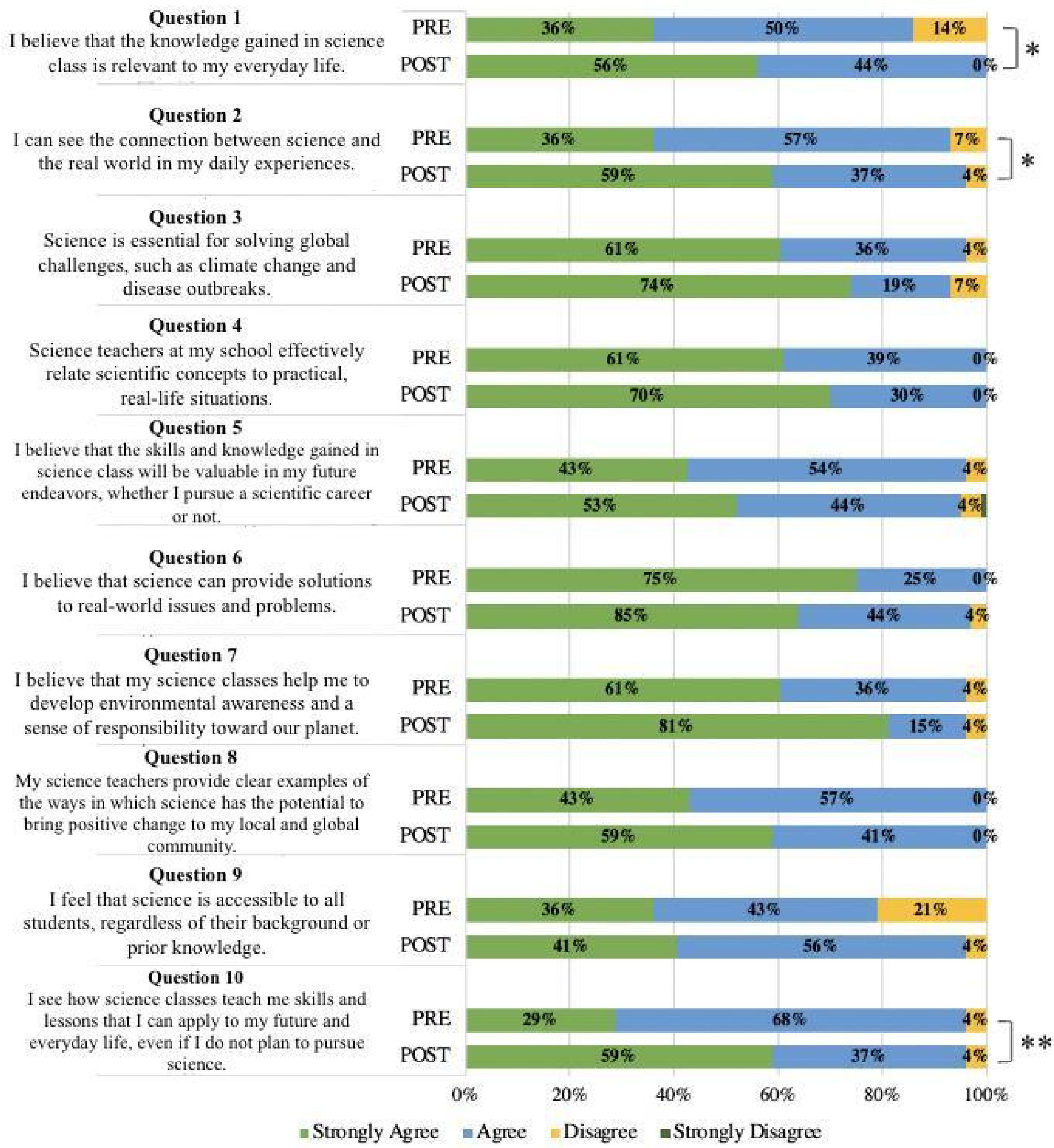


Figure 7. Results of the connection to science pre- and post- treatment Likert survey questions, (*N*=29). *Note*. **p* < .05. ***p* < .01.

A Chi-Squared Goodness of Fit test was conducted for each question to determine if the changes between the pre- and post- treatment responses were significant. Responses to questions 1 and 2 saw statistically significant increases (p = .03) throughout the treatment period. Question 10 reflected statistically significant increases in students' understanding of the application of science content to everyday life, even if they did not have a vested interest in science (p = .002). The outcome of these responses reflects an increase in students' abilities to see connections between classroom learning and real-world problem solving. The validity of these self-reported claims about deeper connection to and understanding of science application were further supported by the Case Study Presentations.

Case Study Data Presentation Connection to Science Results

Student awareness and ability to build connections from inside the classroom to outside the classroom became particularly apparent in the final presentation projects and write-ups. Within the Case Study Presentation Rubric (Appendix J), students were asked to identify connections between classroom topics and their chosen marine research case study. Groups were assessed on the how well their chosen case study aligned with concepts covered in class and the application of the study to real-world problems. All student groups (N=12) scored among the two highest categories of the rubric for this measure (Figure 8).

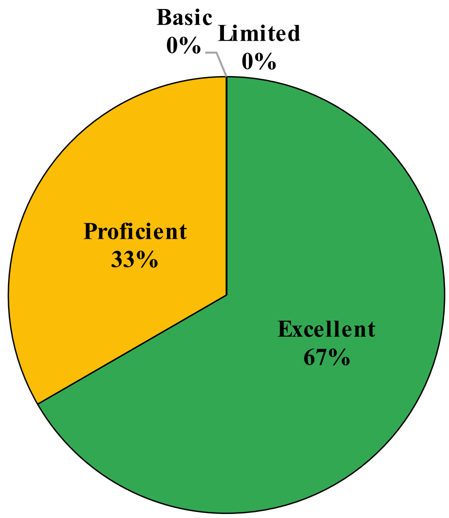


Figure 8. Breakdown of student group scores assessing the connection between classroom topics to real-world research using the case study presentation rubric, (*N*=12).

Upon analysis of the student presentation scores as well as the student surveys, it becomes evident that the use of authentic case studies when teaching classroom topics lead to a deeper sense of connection to the material beyond the lesson.

Summary of the Results

While scores on the Data Literacy Basics assessment did not statistically change following the treatment period, areas of the student surveys reflected statistically significant increases in data literacy abilities, confidence in data analysis, and deeper connections between real-world research and classroom learning. With repeated implementation, data literacy skills can be developed and should be reinforced throughout a students' learning career. Further, with real-world data and case studies, students can develop a deeper understanding of the applicability of classroom learning to real-world problems.

CHAPTER FIVE

CLAIM, EVIDENCE, AND REASONING

Claims From the Study

The goal was to determine the impact on students' scientific data literacy skills and their capacity to connect scientific concepts to real-world science applications through the use and manipulation of authentic data. Within my main research question, three claims were developed: (1) repeated data manipulation will increase students' scientific data literacy skills, (2) repeated data manipulation will increase students' confidence when interacting with scientific data currently and in the future, and (3) teaching scientific concepts through real-world case studies will increase students' understanding of the application of science and scientific skills to real-world problems beyond the classroom.

Data Literacy Skills

Upon analysis of the Data Literacy Skills assessment score, there was not a statistically significant increase ($p < .05$) between the pre- and post- treatment assessment. Therefore, no conclusions from the data literacy skills assessment can be drawn to support the claim that repeated data manipulation throughout the treatment period led to significant increases in students' scientific data literacy skills. However, the skills assessment was only one measure of student ability and only targeted basic data literacy techniques and data interpretation abilities.

Evidence in support of the claim that repeated interaction with data will increase students' data literacy skills is more compelling in the pre- and post-treatment data visualization survey. While statistically significant growth only occurred in one graph type, it is possible that students already held foundational data literacy knowledge and serves as an explanation for the statistically insignificant growth found in the other graph types. For example, most students expressed high rates of familiarity and confidence working with pie charts, column charts, and single line graphs even before the treatment period began (Figure 3). With a strong pre-existing foundation, there was not as much opportunity for growth compared to the more unfamiliar data visualizations. However, the treatment period effectively developed new data literacy skills, as there was a statically significant increase in familiarity with the graph types that students were most unfamiliar with pre- and post- treatment. Similarly, the assessment scores were not statistically significant pre- and post- treatment and the statistical significance on student surveys were scattered. However, through some survey responses, students demonstrated a statistically significant increase in understanding and familiarity with specific data literacy techniques and visualizations following intentional and repeated data manipulation during the treatment period.

Confidence in Data Literacy

Student confidence in data literacy statistically increased in certain areas over the course of the treatment period, supporting the claim that repeated data manipulation will increase students' confidence when interacting with scientific data currently and in the future. In four of the five confidence-related survey questions, students reported an increase in confidence regarding data skill proficiency, data analysis, interpretation, and communication of data. Several questions saw statistically significant increases in confidence working with data as well as communicating data. The outcomes of the student case study data presentations provided further evidence in support of the claim that repeated data manipulation will increase students' confidence when interacting with scientific data.

During the presentations, students were evaluated through a standard rubric that targeted accuracy of data and confidence when presenting data to their peers (Appendix I). All students scored proficient or excellent in all categories related to confidence in data communication and understanding. The confidence in which students presented data to their peers was evident throughout the presentation period. Throughout the presentations, students communicated their data clearly, used evidence to support their claims and developed a cohesive story connecting classroom learning to real-world problems. As evidenced through both the presentations and survey results, working with scientific data repeatedly can lead to greater student confidence and ability to work with and communicate data to others.

Connection to Real-World Science

The results from the student survey and presentation scores support the claim that teaching scientific concepts through real-world case studies will increase students' understanding of the application of science and scientific skills to real-world problems beyond the classroom. Through the Connection to Science survey, students reported feeling better equipped to manipulate data, both in the classroom and in their everyday lives following lessons that highlighted real-world research and analysis of authentic datasets. Even students with little pre-existing interest in science found that they had a greater understanding of the application of science content to everyday life. Students reported that they were able to see connection between classroom learning and real-world problem solving more clearly after the course. This sentiment was particularly evident in the student presentations, in which all students successfully developed clear connections between classroom topics and real-world problem solving. Collectively, the high student presentation scores support the findings from the surveys that the use of authentic case studies when teaching classroom topics lead to a deeper sense of connection to the material beyond the lesson.

Value of the Study and Consideration for Future Research

Data literacy skills are becoming increasingly relevant in an ever-changing world, both in and out of the scientific field (Kjelvik & Schultheis, 2019; Schultheis et al., 2022). However, data literacy is often not incorporated into classroom curriculums due to the challenges of adapting existing curriculum or educators feeling ill-equipped to incorporate data literacy into science curricula (Miller, 2021). This study highlighted the value of meaningfully incorporating data literacy skill-building and techniques into classroom lessons with the hopes of encouraging other educators to do the same.

Capacity for data literacy can be visible in many different ways, such as development of quantitative methods or analysis techniques that ultimately allow students to identify, collect, organize, evaluate, communicate, and transform data into actionable knowledge (Gibson & Mourad, 2018; Mandinach & Gummer, 2013; Ridsdale, 2015; Wolff et al., 2017). The development of data literacy skills is not a one-time learning experience, but can be and must be built upon through repeated measured. Therefore, the results only amplify the importance of incorporating data analysis and data communication practices into high school classrooms. In doing so, gains can be built upon throughout a students' high school career, both in and out of the field of science and in and out of the traditional classroom setting.

A key theme of this study was the impact that real-world stories can have on a students' connection to a topic. Science concepts and research, particularly in marine science, can feel very impersonal or distant to a high school student in a classroom miles away from the ocean.

However, by intentionally crafting lessons that focus on real people, real work, and real locations, the treatment was crafted to address that gap in understanding and students' abilities to relate to their classroom lessons. Students responded to this purposeful lesson design and felt an increased connection to the classroom material and its applicability beyond the classroom.

Although slight, throughout the study period, students showed early signs of data literacy skill development and increased engagement in their lessons when they were taught through the use of real-world case studies and data. Development of data literacy skills are continuous, so I encourage other educators to attempt to incorporate basic data literacy skills, such as data analysis or data communication, into their curriculum. Further, educators should continue to make lessons relatable to students, whether that is through story development to put concepts into context or by finding examples of scientific practices in their local community or region. Ultimately, students felt more connected to the content when it was relatable to them, regardless of their interests or desires to pursue science further in their schooling.

Impact of Action Research on the Author

Throughout my time as an MSSE (Master of Science in Science Education) student, I have learned the value of trusting the process. The development, implementation, and analysis of my action research project was truly an exercise in trusting the process because of the inevitable, unexpected variables of carrying out a project with high school students in an active classroom setting. As educators, it can often be difficult to see the impact of our teaching beyond averages and numbers in a gradebook. This capstone project brought to light the impact that educators can make on the development of our students, both as students and as people. I am encouraged to continue innovating and pushing myself as an educator to identify gaps in my students' learning and my own teaching and develop methods to address those gaps. Lastly, I am even more emboldened to make my lessons data-driven and relatable to students so that they can simultaneously develop critical data literacy skills, build an appreciation for the applicability of science beyond the classroom, and instill a curiosity and love for life-long learning.

REFERENCES CITED

Archer, L., DeWitt, J., Osborne, J. F., Dillon, J. S., Wong, B., & Willis, B. (2013). ASPIRES report: Young people's science and career aspirations, age 10 –14. *King's College London.*

Armstrong, T. (2019). Wrightsville beach, NC water temperature, climatology. https://www.weather.gov/ilm/WaterTemp-WrightsvilleBeach

Berg, C. A., & Smith, P. (1994). Assessing students' abilities to construct and interpret line graphs: Disparities between multiple-choice and free-response instruments. *Science Education, 78*(6), 527-554. https://doi.org/10.1002/sce.3730780602

Brown, M. (2012). Creating marine data scatter plots in ODV. https://marinedataliteracy.org/odv/scatter/scatter.htm

Chichekian, T., & Shore, B. (2017). Challenges to conducting a longitudinal study with classroom observations of teachers' first year of professional practice. *SAGE Research Methods Cases Part 2*. SAGE Publications, Ltd. https://doi.org/10.4135/9781526419378

DeCuir-Gunby, J., & Schutz, P. (2017). *Developing a mixed methods proposal: A practical guide for beginning researchers*. SAGE Publications, Inc. https://doi.org/10.4135/9781483399980

Denny, M., Denieffe, S., & Pajnkihar, M., (2017). Using a non-equivalent control group design in educational research. *SAGE Research Methods Cases Part 2*. SAGE Publications, Ltd. https://doi.org/10.4135/9781526419156

Edmonds, W., & Kennedy, T. (2017). *An applied guide to research designs*. SAGE Publications, Inc. https://doi.org/10.4135/9781071802779

Erwin, R. W. (2015). Data literacy: Real-world learning through problem-solving with data sets. *American Secondary Education, (43*)2, 18–26.

Fabry, V.J., Langdon, C., Balch, W.M., Dickson, A.G., Feely, R.A., Hales, B., Hutchins, D.A., Kleypas, J.A., & Sabine, C.L. (2008). Ocean acidification's effects on marine ecosystems and biogeochemistry: Ocean carbon and biogeochemistry scoping workshop on ocean acidification research. *Eos, 89*(15), 143–143. https://doi.org/10.1029/2008EO150004

Faulconer, E. K., Wood, B., & Griffith, J. C. (2020). Infusing humanities in STEM education: Student opinions of disciplinary connections in an introductory chemistry course. *Journal of Science Education and Technology, 29.* https://doi.org/10.1007/s10956-020-09819-7

Flick, U. (2011). Qualitative research resigns. *Designing qualitative research*. SAGE Publications, Ltd. https://doi.org/10.4135/9781849208826

Gibson, J. P., & Mourad, T. (2018). The growing importance of data literacy in life science education. *American Journal of Botany, 105*(12). https://doi.org/10.1002/ajb2.1195

Gould, R., Sunbury, S., & Dussault, M. (2014). In praise of messy data: Lessons from the search for alien worlds. *The Science Teacher (National Science Teacher Association), 81*(8), 31–36. https://doi.org/10.2505/4/tst14_081_08_31

Holmes, N. G., Wieman, C. E., & Bonn, D. A. (2015). Teaching critical thinking. *Proceedings of the National Academy of Sciences - PNAS*, *112*(36), 11199–11204. https://doi.org/10.1073/pnas.1505329112

Hulleman, C. S, & Harackiewicz, J. M. (2009). Promoting interest and performance in high school science classes. *Science, 326*(12). https://doi.org/10.1126/science.1177067

Hun By The Numbers. Boarding School NJ | The Hun School. (n.d.). https://www.hunschool.org/about-hun/hun-by-the-numbers

Kerlin, S.C., McDonald, S. P., Kelly, G. J. Complexity of secondary scientific data sources and students' argumentative discourse. *International Journal of Science Education*, *32*(9), 1207–1225. https://doi.org/10.1080/09500690902995632

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). https://doi.org/10.1187/cbe.18-02-0023

Koltay, T. (2015). Data literacy: In search of a name and identity. *Journal of Documentation, 71*(2), 401-415. http://dx.doi.org/10.1108/JD-02-2014-0026

Maass, K., & Engeln, K. (2019). Professional development on connections to the world of work in mathematics and science education. *ZDM Mathematics Education, 51,* 967–978. https://doi.org/10.1007/s11858-019-01047-7

Mandinach, E. B., & Gummer, E. S. (2013). A systemic view of implementing data literacy in educator preparation. *Educational Researcher, 42*(1), 30–37. https://doi.org/10.3102/0013189X12459803

McKendrick, J. (2015). *Data driven and digitally savvy: The rise of the new marketing organization.* Forbes Insights. https://images.forbes.com/forbesinsights/StudyPDFs/Turn-The_Rise_of_the_New_Marketing_Organization-REPORT.pdf

Mears, K., Montelpare, W. J., Read, E., McComber, T., Mahar, Alyson, M., Ritchie, K. (2020). Introducing the goodness of fit chi-square. *University of Princeton Edward Island.* https://pressbooks.library.upei.ca/montelpare/chapter/statistical-analysis-when-data-are-counts/

Merriam, S. B. (2002). *Qualitative research in practice: Examples for discussion and analysis.* San Francisco, CA: Jossey-Bass Publishers.

Mertler, C. (2017). *Action research*. (Vols. 1-0). SAGE Publications, Inc. https://doi.org/10.4135/9781483396484

Miller, K., Yoon, S., Shim, J., & Cottone, A. (2021). Integrating data literacy into secondary school science: An exploratory study of a pilot professional development. *International Society of the Learning Sciences,* 781–784. https://doi.dx.org/10.22318/icls2021.781

Ministry for the Environment & Stats NZ. (2019). New Zealand’s environmental reporting series: Our marine environment. https://environment.govt.nz/publications/our-marine-environment-2019/issue-3-our-activities-at-sea-are-affecting-the-marine-environment/

Morrison, J., & McDuffie, A. R. (2010). Connecting science and mathematics: Using inquiry investigations to learn about data collection, analysis, and display. *School Science and Mathematics, 109*(1), 31-44.

Muskopf, S. (2017, October 21). Claim, evidence, reasoning (CER) rubric. *Biology Corner.* https://www.biologycorner.com/2017/10/21/claim-evidence-reasoning-cer/?cv=1

NOAA Ocean Data Exploration. (2022). Investigating sea level. https://dataintheclassroom.noaa.gov/sea-level/investigating-sea-level-teacher-resources

Peterson, R. A. (2000). *Open- and closed-end questions*. SAGE Publications, Inc. https://doi.org/10.4135/9781483349022

Plano Clark, V. & Ivankova, N. (2016). *Mixed methods research: A guide to the field*. SAGE Publications, Inc. https://doi.org/10.4135/9781483398341

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D.J., Kelley, D.R., Matwin, S., & Wuetherick, B. (2015). Strategies and best practices for data literacy education: Knowledge synthesis report. https://doi.org/10.13140/RG.2.1.1922.5044

Rosenberg, J. M., Schultheis, E. H., Kjelvik, M. K., Reedy, A., & Sultana, O. (2022). Big data, big changes? The technologies and sources of data used in science classrooms. *British Journal of Education Technology, 53*(5), 1179–1201. https://doi.org/10.1111/bjet.13245

Schultheis, E. H., & Kjelvik, M. K. (2015). Data nuggets: Bringing real data into the classroom to unearth students' quantitative & inquiry skills. *The American Biology Teacher, 77*(1), 19–29. https://doi.org/10.1525/abt.2015.77.1.4

Schultheis, E. H., & Kjelvik, M. K., Snowden, J., Mead, L., & Stuhlsatz, M. A. M. (2022). Effects of data nuggets on student interest in STEM careers, self-efficacy in data tasks, and ability to construct scientific explanations. *International Journal of Science and Mathematics Education*, 1–24. https://doi.org/10.1007/s10763-022-10295-1

Siddiq, F., Scherer, R., & Tondeur, J. (2016). Teachers' emphasis on developing students' digital information and communication skills (TEDDICS): a new construct in 21st century education, computers & education. *Computers and Education, (92-93)*, (1-14). https://doi.org/10.1016/j.compedu.2015.10.006

Spector, P. E., (1981). *One-group design.* SAGE Publications, Inc. https://doi.org/10.4135/9781412985673

Suter, W. N. (2012). *Introduction to educational research: A critical thinking approach*. SAGE Publications, Inc. https://doi.org/10.4135/9781483384443

Tytler, R., & Australian Council For Educational Research. (2007). *Re-imagining science education: Engaging students in science for Australia's future*. Acer Press

University of Waterloo. (2023). About standard deviation: Teaching assessment processes. http://uwaterloo.ca/teaching-assessment-processes/about-standard-deviation

U.S. Census Bureau (2022). *QuickFacts Princeton, New Jersey.* https://www.census.gov/quickfacts/princetonnewjersey

Wise, A. F. (2019). Educating data scientists and data literate citizens for a new generation of data. *Journal of the Learning Sciences, 29*, 165–181. https://doi.org/10.1080/10508406.2019.1705678

Wolff, A., Gooch, D., Cavero Montaner, J.J, Rashid, U., & Kortuem, G., (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics, 12*(3), 9–26. https://doi.org/10.15353/joci.v12i3.3275

Woolley, M. E., Rose, R. A., Orthner, D. K., Akos, P. T., & Jones-Sanpei, H. (2013). Advancing academic achievement through career relevance in the middle grades: A longitudinal evaluation of CareerStart. *American Educational Research Journal*, *50*(6), 1309–1335. https://doi.org/10.3102/0002831213488818

World Economic Forum. (2022). Oceans are getting warmer – but what were their temperatures 100 years ago? https://www.weforum.org/agenda/2022/06/oceans-warming-climate-change/

APPENDICES

APPENDIX A

MONTANA STATE UNIVERSITY’S INSTITUTION REVIEW BOARD EXEMPTION

Hello Bottega, Caroline,

Your protocol was reviewed by the IRB and has been approved.

PI: Bottega, Caroline
Approval Date: 1/22/2024
Title: Authentic Data in High School Marine Science: Impacts on Data Literacy and Real-World Connections

Protocol #: 2024-1140-EXEMPT
Review Type: Exemption
Expiration Date: 1/22/2029

Work described under this protocol may now commence. The PI is responsible for ensuring that the protocol accurately describes research practices being conducted.

> Review Category designation determined by the IRB can be found in the final section of your protocol.
> IRB-stamped active Consent Forms are attached within your protocol where applicable.
> Any changes must be submitted via Amendment prior to implementation.
> Per the Common Rule, research only requires Interim (annual) Review by the IRB if 1) it was reviewed via Full Committee or 2) is regulated by the FDA.
> All research is subject to post approval monitoring.
> All protocol types must be renewed 5 years after approval.
> Inform the IRB once your research is complete so that the protocol may be inactivated.

Please contact your IRB Program Manager with any questions or if you are in need of assistance. Thank you for your diligence in the care of human subjects research participants.

Institutional Review Board for the Protection of Human Subjects | Office of Research Compliance | Montana State University

Access your protocol anytime at https://montanaprod.topazti.net//Elements?emailLink=11%2c102%2c10912.

APPENDIX B

PRE- AND POST-ASSESSMENT: DATA LITERACY BASICS

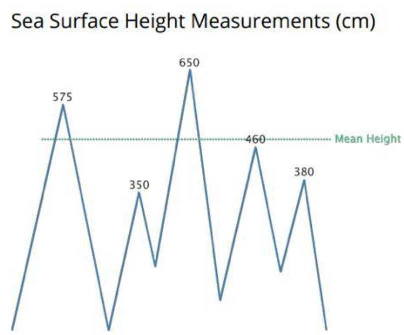
Disclaimer: Please answer the following questions to the best of your ability. Data collected from this assessment is <u>anonymous</u>, and participation in this research is voluntary. Scores for this assessment will not affect your grade or class standing in any way.

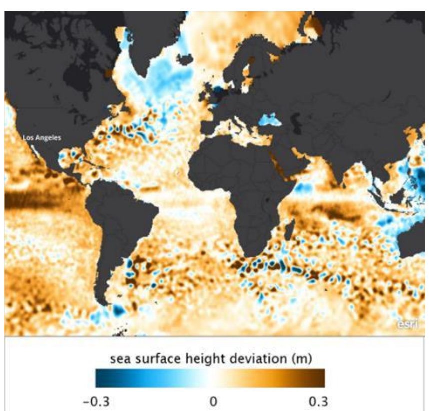
Section 1: Data Literacy Essentials

#	**Question**	**Skill Assessed**
1	______If you are investigating how changes in water temperature impact the growth of coral reefs, what is the dependent (responding) variable? a. Water temperature b. The size of the coral reefs c. The types of fish in the area d. The depth of the ocean	**SKILL #1:** Ability to identify scientific variables (independent, dependent, control)
2	______Which type of graph is suitable for representing the changes in the pH levels of seawater over a period of time? a. Bar graph b. Line graph c. Pie chart d. Scatter plot	**SKILL #2A:** Identify the proper format for the graphical representation of data given the particular data type.
3	______You want to compare the population sizes of two different species of marine turtles in various locations along the coast. What type of chart or graph would be most suitable for this comparison? a. Bar chart b. Pie chart c. Scatter plot d. Line graph	**SKILL #2A:** Identify the proper format for the graphical representation of data given the particular data type.
4	______Which of the following is an example of an independent (manipulated) variable in a marine science experiment? a. The number of fish species found in a coral reef b. The salinity of the ocean water measured off the back of a boat c. The amount of time spent surveying a plot of coastline d. The abundance of kelp found on the coast	**SKILL #1:** Ability to identify scientific variables (independent, dependent, control)
5	______You want to represent the change in ocean acidity levels over time in a line graph. Which type of data should be plotted on the y-axis? a. Time (in months) b. Ocean acidity levels c. Latitude and longitude coordinates d. Depth of the ocean	**SKILL #1:** Ability to identify scientific variables (independent, dependent, control)

Section 2: Data in the real world: *How is Sea Level Calculated?*
Adapted from NOAA Ocean Data Exploration (2022)

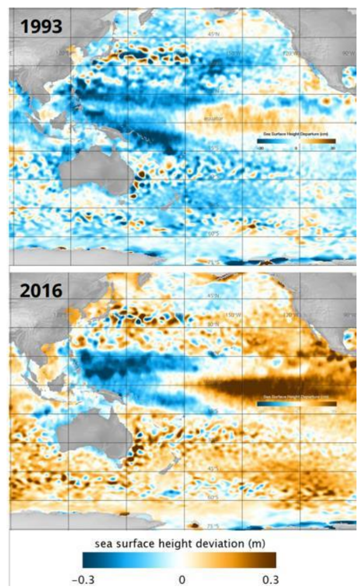
6	_______ To understand sea-level data, it is important to know the concept of the mean average. How do you calculate mean in a data set? a. Arrange the numbers in numerical order, find the middle number b. Add up all the numbers, then divide by how many numbers there are c. Find out which number appears the most often d. Take the greatest value and subtract it from the smallest value	**SKILL #3:** Ability to read and interpret graphical representations of data
7	_______Imagine that at a particular location, the sea level (known as sea surface height) was measured 5 times. These 5 measurements are shown in the diagram. What is the mean of these five values? a. 380 b. 300 c. 483 d. 460	**SKILL #3:** Ability to read and interpret graphical representations of data





Sea surface height deviation (SSHD) is the difference between the historical mean and the sea surface measurement for a particular date. Examine a data map that shows sea surface height deviations (SSHD) over the entire planet from January 2016.

8	_______What do the yellow and orange colors on the map represent? a. The sea surface is warmer than historic measurements b. The sea surface is higher than the historic mean c. The sea surface is rougher than historic measurements d. The sea surface is lower than the historic mean	**SKILL #3:** Ability to read and interpret graphical representations of data
9	_______In January 2016, what was the approximate SSHD off the coast of Los Angeles, USA? a. 0.1 m b. - 0.2 m c. 0 m d. -0.1 m	**SKILL #3:** Ability to read and interpret graphical representations of data

10	______What unit is the SSHD measured in? a. Centimeters b. Millimeters c. Kilometers d. Meters	**SKILL 2B:** Implement the standards outlined in Graphing Checklist (Appendix D)
11	______How much have sea levels changed over time? Compare the SSHD data maps from December 1993 and December 2016 to find out. Which statement best describes how have sea levels changed globally, from 1993 to 2016? a. Sea levels have decreased from 1993 to 2016. b. Sea levels have not changed from 1993 to 2016. c. Sea levels have increased from 1993 to 2016. d. Sea level range has grown smaller from 1993 to 2016. 	**SKILL 3B:** Ability to draw conclusions about a study
12	______Consider the changes identified in the data maps. Which of the following is a real-world consequence of these changes? a. Receding shorelines b. Decreased habitat c. Increased wave height d. Shorter boating season	**SKILL #5:** Ability to connect scientific findings to the larger body of scientific knowledge
13	______You are tasked with creating a line graph to show the changes in mean sea surface height in a coastal area over the past 50 years. What should be on the x-axis (horizontal axis) of your graph? a. Years b. Time of Day c. Depth of the Ocean d. Temperature	**SKILL #2:** Ability to create graphical representations of data
14	______When analyzing mean sea surface height data, why is it important to consider data over a long time period rather than just a few days or weeks? a. Long-term data provides a more accurate average. b. Short-term data is easier to collect. c. Short-term data is more reliable. d. Long-term data is irrelevant to sea surface height analysis.	**SKILL #4:** Ability to communicate data to a general audience **SKILL #5:** Ability to connect scientific findings to the larger body of scientific knowledge

15	______If you discovered a significant increase in mean sea surface height in a coastal region, what could be a potential real-world application of this data for the scientific community? a. Identifying the best surfing spots. b. Predicting potential flooding risks. c. Developing new seafood recipes. d. Promoting beach tourism.	**SKILL #5:** Ability to connect scientific findings to the larger body of scientific knowledge

APPENDIX C

DATA LITERACY SKILLS

SKILL #1: Ability to identify scientific variables (independent, dependent, control)
SKILL #2: Ability to create graphical representations of data • **2A:** Identify the proper format for the graphical representation of data given the particular data type. • **2B:** Implement the standards outlined in the Graphing Checklist (Appendix D)
SKILL #3: Ability to read and interpret graphical representations of data • **3B:** Ability to draw conclusions about a study
SKILL #4: Ability to communicate data to a general audience
SKILL #5: Ability to connect scientific findings to the larger body of scientific knowledge

APPENDIX D

GRAPHING CHECKLIST

- The correct type of graph is made for the type of data presented (**line graph, bar graph, scatter plot, etc**)
- Graph is **neatly constructed, organized**, and makes **good use of space**.
- **Specific title** is included. The title indicates what data is presented.
  - The **Y axis** is labeled with the **Responding** Variable and **units**
  - The **X axis** is labeled with the **Changed/Manipulated** Variable and **units**
  - The **data is spread out evenly**, not clumped
  - X and Y axis **intervals are consistent** and correct
  - All points are **plotted clearly and correctly.**
  - If needed, a **key** is included on the graph, describing
  - If needed, **best fit lines** are added to the graph to show trends or relationships

Remember DRY MIX

Dependent
Responding
Y-axis

Manipulated
Independent
X-axis

APPENDIX E

PRE- AND POST ASSESSMENT: CONFIDENCE IN DATA LITERACY

Disclaimer: Please answer the following questions to the best of your ability. Circle one answer when using the disagree-agree scale. Data collected from this assessment is anonymous, and participation in this research is voluntary.

Your honest responses are valuable in helping me understand how you feel about your confidence in your data literacy skills.

#	**Question**
1	______I feel confident in my ability to organize data for analysis. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
2	______I am comfortable with interpreting data and drawing conclusions from it. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
3	______I have the skills to create data visualizations, such as charts, graphs, and other tools. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
4	______I am confident in my ability to communicate the results of data analysis clearly to others. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
5	______I feel well-prepared to handle real-world situations that require data analysis and interpretation. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
6	______I believe that data skills are essential for success in various fields, not just in science and mathematics. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree

7	______I find data analysis and interpretation to be an enjoyable and rewarding activity. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
8	______I think my teachers have provided me with adequate resources and guidance to improve my data skills. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
9	______I feel that I can use data to make informed decisions in my personal life. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
10	______I am confident in my ability to work with and communicate data effectively in group projects or collaborative settings. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree

APPENDIX F

PRE- AND POST- ASSESSMENT: CONNECTION TO SCIENCE

Disclaimer: Please answer the following questions to the best of your ability. Circle one answer when using the disagree-agree scale. Data collected from this assessment is anonymous, and participation in this research is voluntary.

Your honest responses are valuable in helping me understand how you feel about your understanding of science.

#	**Question**
1	______I believe that the knowledge gained in science class is relevant to my everyday life. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
2	______I can see the connection between science and the real world in my daily experiences. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
3	______Science is essential for solving global challenges, such as climate change and disease outbreaks. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
4	______Science teachers at my school effectively relate scientific concepts to practical, real-life situations. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
5	______I believe that the skills and knowledge gained in science class will be valuable in my future endeavors, whether I pursue a scientific career or not. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
6	______I believe that science can provide solutions to real-world issues and problems. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree

7	______I believe that my science classes help me to develop environmental awareness and a sense of responsibility toward our planet. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
8	______My science teachers provide clear examples of the ways in which science has the potential to bring positive change to my local and global community. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
9	______I feel that science is accessible to all students, regardless of their background or prior knowledge. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree
10	______I see how science classes teach me skills and lessons that I can apply to my future and everyday life, even if I do not plan to pursue science. A. Strongly Agree B. Agree C. Disagree D. Strongly Disagree

APPENDIX G

OPEN-ENDED SURVEY: DATA VISUALIZATION TOOLS

Disclaimer: Please answer the following questions to the best of your ability. Data collected from this assessment is anonymous, and participation in this research is voluntary. Scores for this assessment will not affect your grade or class standing in any way.

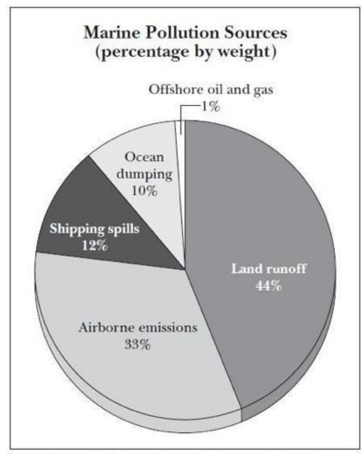
Section 1: Data Literacy Essentials

Below, you'll find various types of graphs and data displays. Choose the option that most accurately reflects your familiarity with each graph and elaborate on your understanding.

When interpreting the graph, clarify the significance of the variables, identify the information presented, and provide additional insights about the specific type of graph.

1 ______Choose the answer that most accurately reflects your understanding of the graph, and describe what aspects you do and do not understand.

a. I have never seen this type of graph before, and I am not sure how to interpret it.
b. I have never seen this type of graph before, but I interpret it as…..
c. I have seen this type of graph before, but I am not sure how to interpret it.
d. I have seen this type of graph before, and I interpret it as…..
e. I have seen and worked with this type of graph before, and I interpret it as…..

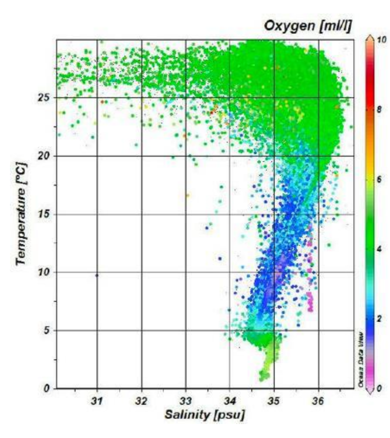


Source: Data taken from United Nations Environment Programme, *The State of the Marine Environment.* Oxford, England: Blackwell Scientific Publications, 1991.

Elaborate on your understanding by describing your interpretation, explaining the variables, identifying the information depicted in the graph, and sharing any additional insights you have about this particular type of graph.

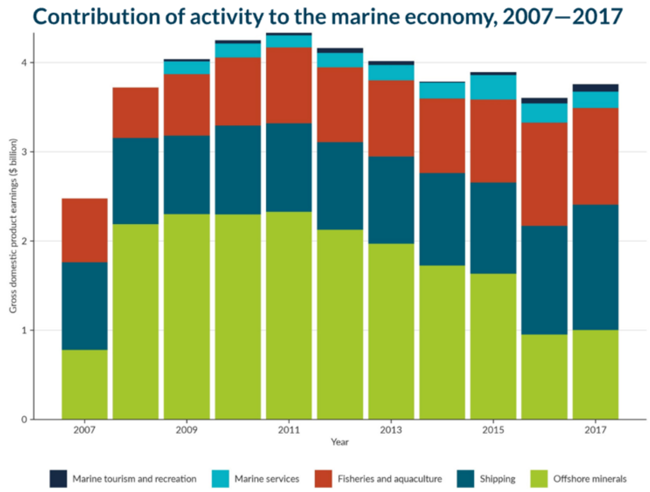
2 ______Choose the answer that most accurately reflects your understanding of the graph, and describe what aspects you do and do not understand. (Brown, 2012).

a. I have never seen this type of graph before, and I am not sure how to interpret it.
b. I have never seen this type of graph before, but I interpret it as…..
c. I have seen this type of graph before, but I am not sure how to interpret it.
d. I have seen this type of graph before, and I interpret it as…..
e. I have seen and worked with this type of graph before, and I interpret it as…..



Elaborate on your understanding by describing your interpretation, explaining the variables, identifying the information depicted in the graph, and sharing any additional insights you have about this particular type of graph.

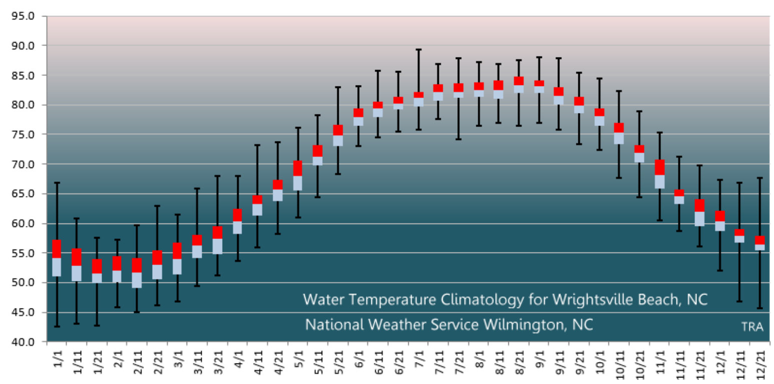
3 ______Choose the answer that most accurately reflects your understanding of the graph, and describe what aspects you do and do not understand. (Ministry for the Environment & Stats NZ, 2019).



a. I have never seen this type of graph before, and I am not sure how to interpret it.
b. I have never seen this type of graph before, but I interpret it as…..
c. I have seen this type of graph before, but I am not sure how to interpret it.
d. I have seen this type of graph before, and I interpret it as…..
e. I have seen and worked with this type of graph before, and I interpret it as…..

Elaborate on your understanding by describing your interpretation, explaining the variables, identifying the information depicted in the graph, and sharing any additional insights you have about this particular type of graph.

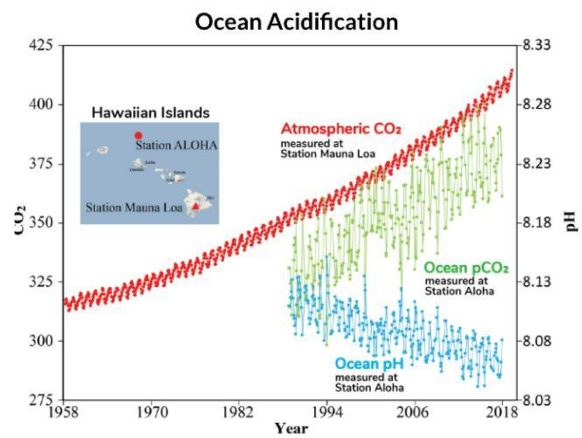
4 ______Choose the answer that most accurately reflects your understanding of the graph, and describe what aspects you do and do not understand. (Armstrong, 2019).



a. I have never seen this type of graph before, and I am not sure how to interpret it.
b. I have never seen this type of graph before, but I interpret it as…..
c. I have seen this type of graph before, but I am not sure how to interpret it.
d. I have seen this type of graph before, and I interpret it as…..
e. I have seen and worked with this type of graph before, and I interpret it as…..

Elaborate on your understanding by describing your interpretation, explaining the variables, identifying the information depicted in the graph, and sharing any additional insights you have about this particular type of graph.

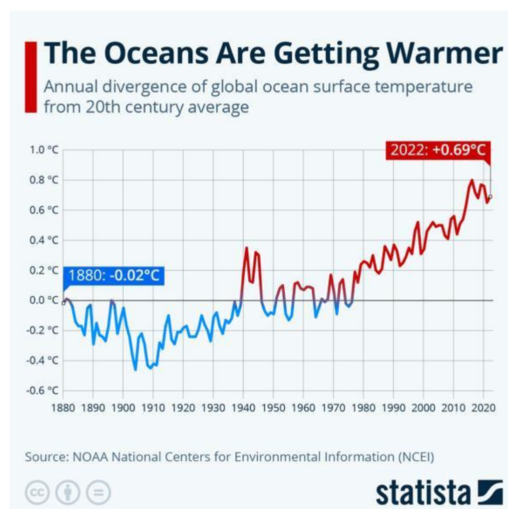
5 ______Choose the answer that most accurately reflects your understanding of the graph, and describe what aspects you do and do not understand. (Fabry et al. 2008).



a. I have never seen this type of graph before, and I am not sure how to interpret it.
b. I have never seen this type of graph before, but I interpret it as…..
c. I have seen this type of graph before, but I am not sure how to interpret it.
d. I have seen this type of graph before, and I interpret it as…..
e. I have seen and worked with this type of graph before, and I interpret it as…..

Elaborate on your understanding by describing your interpretation, explaining the variables, identifying the information depicted in the graph, and sharing any additional insights you have about this particular type of graph.

6 ______Choose the answer that most accurately reflects your understanding of the graph, and describe what aspects you do and do not understand. (World Economic Forum, 2022).



a. I have never seen this type of graph before, and I am not sure how to interpret it.
b. I have never seen this type of graph before, but I interpret it as…..
c. I have seen this type of graph before, but I am not sure how to interpret it.
d. I have seen this type of graph before, and I interpret it as…..
e. I have seen and worked with this type of graph before, and I interpret it as…..

Elaborate on your understanding by describing your interpretation, explaining the variables, identifying the information depicted in the graph, and sharing any additional insights you have about this particular type of graph.

APPENDIX H

CER RUBRIC

	Accomplished (3)	Proficient (2)	Developing (1)
CLAIM	Specific language related to the case or argument is used to establish the claim. Claim is clear and concise. Uses complete and grammatically correct sentences	Claim is established, language may be vague or unclear or include some unrelated details	Claim is not established, not accurate or unrelated to the question or case
EVIDENCE	Evidence is provided from the case, article, video or lecture. It is sufficient enough to support the claim	Evidence is provided but lacking in some details or may not be sufficient or lacking in focus.	Evidence provided is insufficient or is not accurate; no details provided
REASONING	Evidence is connected to the claim and to major ideas to principles about the topic. It is complete and thorough	Evidence is tied to the claim, some details missing or not sufficient to establish a connection	Evidence is not tied to the claim, no additional information is used to connect ideas to principles

(Muskopf, 2017)

APPENDIX I

CASE STUDY DATA PRESENTATION RUBRIC

Category	Excellent (4)	Proficient (3)	Basic (2)	Limited (1)	Not Shown (0)
Connection to Real World Science					
Connection Between Classroom Topics to Real-World Research	Very clear connection to classroom topics	Good connection to classroom topics	Some connection to classroom topics	Limited connection to classroom topics	No relevance to classroom topics
Accurate Representation of Case Study	All elements of case study clearly communicated	Most elements of case study clearly communicated	Some elements of case study clearly communicated	Few elements of case study clearly communicated	No elements of case study clearly communicated
Communication of Science	Confidently communicates information to audience and demonstrates strong understanding of scientific topic.	Mostly communicates information to audience and demonstrates proficient understanding of scientific topic.	Communicates some information directly to audience and demonstrates good understanding of scientific topic.	Communicates little information directly to audience and demonstrates limited understanding of scientific topic.	Communicates no information directly to audience and demonstrates no understanding of scientific topic.
Data Literacy Skills					
Data visualization tool accuracy	Proper selection of data literacy tool, all graphing elements included	Proper selection of data literacy tool, some graphing elements included	Somewhat proper selection of data literacy tool, some elements included	Improper selection of data literacy tool, missing many essential elements	Improper selection of data literacy tool, missing essential elements
Effective Data Selection	All data contributed to communication of case study	Most data contributed to communication of case study	Some data contributed to communication of case study	More data needed to clearly communicate case study	Not enough data presented to communicate case study
Data visualization tool accuracy	All graphs align with graphing checklist	Most graphs align with graphing checklist	Some graphs align with graphing checklist	Few graphs align with the graphing checklist	No graphs align with the graphing checklist
Data Interpretation	Student clearly communicates results through data visualizations	Student mostly communicates results through data visualizations	Student somewhat communicates results through data visualizations	Student poorly communicates results through data visualizations	Student does not communicate results through data visualizations
Overall Score	**[Total Score]**	**[Total Score]**	**[Total Score]**	**[Total Score]**	**[Total Score]**