THE USE OF AUTHENTIC MESSY DATA AND CASE STUDIES TO IMPROVE DATA LITERACY SKILLS IN HIGH SCHOOL STUDENTS

by

Julie Kathryn Ekhoff

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

of

Masters of Science

in

Science Education

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2022

DEDICATION

I would like to dedicate this to my family, especially my mom and my 2021-2022 Research 1 students who participated in the project and allowed to be a student with them.

TABLE OF CONTENTS

TABLE OF CONTENTS CONTINUED

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The purpose of the is study was to examine if analyzing and interpreting authentic messy data had any effect on my students' data literacy skills. Additionally, the study examined if case studies were an effective means of communicating authentic data. For this project students were given case studies that presented data using various graphs. Students were then asked to describe, analyze and reach conclusions about the data in the case study. The non-treatment group received case studies with clean data, while the treatment group received case studies with messy data. The two groups were compared using pre- and post-assessment, surveys and student interviews. The results showed that the use of messy data did not impact student data literacy skills, but using messy may increase student ability to think of data critically. Using case studies allowed the students to incorporate information on the subject and data collection methods into their data-based conclusions.

INTRODUCTION AND BACKGROUND

The purpose of this project was to examine the effectiveness of using case studies to improve data literacy skills in students taking high school science courses. It was proposed that the use of authentic messy data incorporated into case studies would increase student ability to evaluate data-based arguments and form their own data-based conclusions. The data literacy skills being assessed were those that pertained to describing and interpreting data displays. This focus came from observing my students' data literacy skills and conversations I had with them.

Context of Study

I work at a stem magnet school within the Boise School District, in Boise Idaho. Boise is a metropolitan area that has experienced a 14.6% growth rate over the last 10 years and now has a population of 335,000 people, of which nearly 22% are below the age of 18 (KTVB, 2021; US Census, 2019). Several technology companies such as Micron, Hewlett Packard (HP), and Clearwater Analytics are based in Boise. They chose the Boise area due to the low costs to run a business, the lower cost of living when compared to other states and the potential work force (Kohl, 2018). 41.6% of Boise residents have a bachelor's degree and the Boise School District sees 94% of its students graduate (Kohl, 2018; U.S. Census, 2019). The school I teach at, Treasure Valley Mathematics and Science Center (TVMSC), was established 2004 with the help of Micron and HP to serve students who accel in math and science.

TVMSC is a half-day program in which students travel to the school either in the morning or afternoon for advance science and math courses. TVMSC is open to students throughout the Boise District, neighboring districts and to students who home school. The students at TVMSC are grouped into cohorts, seventh to 12th, based on the math class in which they are enrolled. For example, most of the students in the ninth cohort take Integrate Math III (Algebra 2), with some exceptions every year. This year the students that comprise the 9^{th} cohort range in actual grade level from eighth to 10^{th} grade, and within actual grade levels it is not uncommon to have students that have skipped a grade. Nearly every graduating senior at TVMSC goes onto college, with many eventually pursuing graduate degrees. With a highly motivated student population who are typically ahead of their peers in science and math one would think that data literacy is not a skill that the students would lack, but in recent years I have been noticing gaps in their skill set, particularly when it comes to the evaluation of data that they did not collect.

I have taught at TVMSC for 15 years, teaching several different classes, two of which are Advanced Placement (AP) Biology and AP Environmental Science, both of which have students evaluating data in order to form their own conclusions and those made by authors. Typically, the first step in this evaluation has students describing the data set. The descriptions that my students often give are simple and don't address any fluctuations within the data. Students are then asked to interpret the data display, to form conclusions involving the variables represented or evaluate conclusions formed by others. Interpreting the data displays is made more difficult when the descriptions are very generalized. A majority of my AP students have taken another class I teach, Research 1, which contains a basic statistics unit, this is in addition to the statistics taught in their math classes. Some of the students have also taken AP Statistics. The question I was left with was "why weren't my students accurately describing data?" Perhaps it was because they really didn't understand what the data was saying.

When talking with my students about data that they collect compared to data that is presented to them, most preferred to collect their own data. When collecting their own data students know why variations in the data exist and for the most part attributed those variations to the quality of the equipment they are using or that they are “amateur” scientists. These discussions also revealed that when analyzing data that is presented to them, they are uncomfortable with data that has a high degree of variability because they don’t know why it is there. Both instances demonstrate a gap in their data literacy in that most data comes with variability and that variability is not necessarily a result of error. I had been working with the assumption that if the students could do data analysis and create displays for their data that they would be able to evaluate data generated by others, including data with increased complexity. I would liken this assumption to teaching students how to write a story and then expecting them to be able to analyze Shakespeare. The result of these conversations and what students would produce when asked to describe data had me thinking that this may be due to their lack of experience with complex data displays. Their inexperience could be caused by students working primarily with data that they generated, which lacks numerous data points, and with them attributing any variability to low tech equipment or human error. Secondarily, when presented with data generated by professional researchers it is typically “cleaned up”, with outliers and variability lessen, this is to reinforce the scientific concept being taught and is how data is commonly presented in educational media.

Focus Question

My focus question was, How does the use of authentic messy data sets affect student understanding of data, as seen in their ability to read and interpret graphs? In order to introduce messy data to my students, the strategy of using case studies was determined to be a viable method, as it provides context for the data and increases critical thinking.

My sub-questions include the following:

1. What effect does the use of data-based case studies have on data literacy?
2. How does the use of messy data affect students' abilities to describe graphs, including the ability to describe variation within data?
3. How does the use of messy data affect students' abilities to reach data-based conclusions and evaluate the conclusions formed by others?

CONCEPTUAL FRAMEWORK

In order to investigate the use of case studies to present authentic messy data to improve data literacy skills several avenues of research were done. These avenues included the topic of data literacy, the use of case-based learning, and the complexity of authentic data sets.

The Need for Data Literacy

Data literacy, also known as statistical literacy, is the ability to analyze and evaluate data in order to form data-based conclusions. Herried (2014) described data literacy (quantitative reasoning) as the ability "to read and interpret data, graphs and statics" (p. 1). Forbes et al. (2011) expands on this definition to include the ability to communicate using data. While Weiland (2017) takes the definition one step further by adding the ability to make informed decisions based on "statistical arguments" (p. 33). All these definitions result to two skills sets. The first is the ability to apply statistical analysis to raw data and create appropriate displays to communicate the data. The second skill is the ability to evaluate data and reach a data-based conclusion.

In science the use of data is extensive as it is the evidence used in scientific arguments (Gibson & Mourad, 2018). Students need to be able to support their research with data-based evidence and be able to evaluate evidence presented by others (Gibson & Mourad, 2018). Using data-based evidence in arguments is part of the Next Gen Science Standards (NGSS) Framework (NGSS Lead States, 2013) and evaluating data-based evidence is a science practice skill seen in AP science courses (Science Practices, n.d.). In our world the evaluation of data-based evidence is not limited to scientific research read in journal articles or heard at conferences. People are presented with claims every day, claims that are seemingly supported by data, and these claims need to be looked at critically (Vahey et al., 2012). We are currently living in a data rich environment and this data is meant to inform, but that can only happen if the individual can understand the data being presented, with the Covid-19 pandemic serving as an example (Oslington, 2020). Topics such as climate change, cost of living compared to wages, polling results, and racial inequality all serve as recent examples of issues in our society that are often presented along with data.

The importance of being able to evaluate complex data is required in the careers that current high school students will be pursuing along with the responsibility of becoming a fully participating citizen in their communities. As our world becomes more data rich and companies turn to data analysis in their decision-making processes a skill gap has developed in the workforce (Hemo, 2020). Companies have found that employees can work to create data and apply appropriate statistical analysis, but fall short in the ability to interpret the data (Hemo, 2020). Outside of school and the workforce, good data literacy skills make for good citizens, who play an active role in their community (Weiland, 2017). The need for data literacy is so great that in 2018 the Canadian government authored a report on data literacy in public service, recognizing that data drives better decision making (Bonikowska, 2019). If students are going to be asked to make informed decisions based off data, then students need the opportunity to develop the data literacy skills called for, and the only way to develop those skills is to expose students to real-world data (Weiland, 2017).

Data Literacy and Schools

Traditionally K-12 education placed data literacy, called statistical literacy by math educators, in the hands of a statistics course that students could take in high school, this has changed with the introduction of the NGSS framework and Common Core State Standards (CCSS) in math. CCSS math standards have statistics at every level of the math curriculum starting in the sixth grade, with earlier grades having standards referred to as measurements and data (Common Core, 2010). The NGSS emphasizes data literacy by having students collect, analyze and evaluate data throughout their K-12 education by including it in content areas (NGSS Lead States, 2013). The NGSS also emphasizes the ability of students to engage in scientific arguments using evidence, including the validating of claims made by others and supporting their own (NGSS Lead States, 2013). The College Board which develops Advance Placement (AP) curriculum has included data literacy as part of their science practices, standards seen in every AP science course (Science Practices, n.d.). Example standards from each of these three entities are seen in Table 1.

Table 1. Examples of high school common core state standards for math content (Common Core, 2010), NGSS data literacy standards (National Research Council, 2013) and AP Science Skills (CED for AP Physics 1, AP Biology, and AP Chemistry).

	Standard
NGSS	HS-LS3-3. Apply concepts of statistics and probability to explain the variation and distribution of expressed traits in a population. HS-ESS2-2. Analyze geoscience data to make the claim that one change to earth's surface can create feedbacks that cause changes to other earth systems. HS-PS2-1. Analyze data to support the claim that Newton's second law of motion describes the mathematical relationship among the net force on a microscopic, its mass and its acceleration.
CCSS Math Content	HSS.IC.B4 Use data from a sample survey to estimate a population mean or proportion; develop a margin of error through the use of simulation models for random sampling. HSS.IC. A1 Understand statistics as a process for making inferences about population parameters based on a random sample from that population. HSS.IC.B6 Evaluate reports based on data
AP Science Practices	AP Physics 1 - 6.1 Students can justify claims with evidence AP Biology 6.C Provide reasoning to justify a claim by connecting evidence to biological theories. AP Chemistry 6.B Support a claim with evidence from experimental data

Assessing Data Literacy

Data literacy standards, like those seen in Table 1, can be assessed using multiple choice, constructed response items such as fill in the blank or solving a problem, and performance tasks (Zalles, 2005). The assessment used depends on the wanted outcome of the standard. For the examples used in Table 1, all three methods of assessment could be used for standards with the verbs: understand, use, apply and analyze, given the assessments are properly created. For standards with the direction to evaluate; constructed responses and performance tasks would be more appropriate. The AP science practice standards using language like justify and provide reasoning are routinely assessed in the free response portion of AP exams with short answer constructed responses.

When assessing data literacy skills, students will use three forms of reasoning: quantitative reasoning (numerical literacy), content reasoning, and logic (Hoffman, 2016). Students with strong content knowledge may be able to use that knowledge to overcome a lack in quantitative reasoning (Hoffman, 2016). Noting some of examples in Table 1 again, many of the standards are closely tied to content. This poses an issue if the goal is to develop data analysis and interpretation skills that can be used in many different situations. In order to bolster data literacy skills that will be needed outside of a specific content area, data literacy should be taught with an interdisciplinary approach (Erwin, 2015; Vahey et al., 2012; Weiland, 2017).

Teaching Data Literacy

Best practices for teaching data literacy skills that pertain to the interpretation of data, forming data-based conclusions or arguments and the evaluation of arguments made by others include:

1. The use of context (Herreid, 2014; Kahn et al., 2015)
2. Interdisciplinary approach (Casey, 2010; Herreid et al., 2012)
3. Constructivism (McLeod, 2019)
4. Use of authentic messy data (Bowen & Bartley, 2014; Kjelvik & Shchultheis, 2019)

One method for the incorporation of these practices is the use case studies. The use of case studies to teach or reinforce data literacy skills opens the door for an interdisciplinary approach and the use of complex real-life data, both of which add relevance and context to the curriculum being taught (Erwin, 2015).

Case-Based Learning

Case-based learning is the use of real-world scenarios to provide students the opportunity to expand their knowledge in a content area or apply it through the use of critical thinking (What is Case-Based Learning, 2020). The real-world story in a case study enriches data being presented and the data can enrich the story (Herreid, 2014). Think about major "stories" from our history that mean more when data is used and the data means more when the story behind it is understood. Stories like the battle of Gettysburg or the Holocaust. The result of this enrichment is a deeper understanding of topics and increased active learning for all students. Case-based learning reinforces prior learning, promotes critical thinking, and develops problems-solving skills (Mostert, 2007). Kahn et al. (2015) discussed the benefits of case-based learning with medical students, stating that increased analytical thinking was due to students driving the learning during a case study, with teachers becoming guides, and that case studies provided context for knowledge learned. The context provided in case studies can be from one discipline but in many instances, they provide the opportunity for students to problem-solve and form conclusions using information from more than one subject.

Case-Based Learning: An Interdisciplinary Approach

The use of statistical analysis for a scientific problem is in itself an interdisciplinary approach, with math and science working together. Case studies allow for more disciplines to enter the story, further enriching the data. For example, in a case study for high school students presented in Science Stories: Using Case Studies to Teach Critical Thinking, students are tasked with examining the factor a toxin may have played in the Salem Witch Trials (Herreid et al., 2012). In this case study students read about the Salem witch trials including the dynamics of the community and they are given information about a potential toxin produced by a fungus that grows on grains. Students then evaluate the historical medical records to determine what role if any the toxin may have played in the mass hysteria that led to the trials. Having students work with two different disciplines, history and science, as well as data forces students to critically think about all aspects of the problem they are presented with. An interdisciplinary approach, like the one seen in the example case study, is necessary in many fields of science. Environmental science is a great example of an interdisciplinary field which brings together the sciences of geology, biology, and physics, along with government oversight and ethics. The integration of multiple disciplines leads to improved critical thinking and an increase in retention of content knowledge (Casey, 2010).

Case-Based Learning and Constructivism

Case studies make use of constructivism, as students use their current knowledge base to build new knowledge. The building of new knowledge means the learner is not a passive participant in the process but an active one (McLeod, 2019). The current knowledge a student brings to a case study is not limited to curriculum but also includes their personal experience. The addition of a student's experiences results in that student having their own interpretation to any given case study (Cognitive Constructivism, 2021). The interdisciplinary approach and the building of knowledge make case-based learning a vehicle for students to practice data literacy with the student leading the learning.

Implementing Case Studies

Case studies can be classified by difficulty and method of use (Herreid et al., 2014). In Science Stories You Can Count On, the difficulty classification is broken down into three contributing factors; conceptual, presentation, and analysis (Herreid et al., 2014). Conceptual speaks to the depth of knowledge in the content area a student needs to be successful with the case study. How the information is presented is the presentational aspect and the analytical piece pertains to what the students have to do with the information presented. The analytical aspect can be as simple as presenting both the problem and a solution to students with the students evaluating the solution, to presenting students with information and having them define the problem and then working to find a solution. The appropriate adjustment of these factors can help to mitigate a student's reliance on their content knowledge in order to push the development of quantitative reasoning and data literacy skills.

Using Messy Real-World Data

The data most often seen, analyzed and interpreted by students is clean and lacking in variation. Bowen and Bartley (2014) describe clean, or sanitized, data as unambiguous and as the data most often presented in textbooks. They add that this lack of variation in data leads students to believe that unless the data is perfect that the relationships between variables do not exist, when in fact they might. The use of messy data adds an authenticity to the data and in the inquiry it brings. Gould et al. (2022) noted that when high school students worked with messy data provided from satellites, it led to discussions involving which data points to include in analysis and which to ignore, including the criteria that should be used to make that decision. For the students working with the satellite data, an observation was also made that the model they were using might not best fit the data they were seeing. The messy data provided an additional level of analysis and critical thinking.

Just as Herreid classified case studies, data sets themselves can be classified. Kjelvik and Shchultheis (2019) discuss the classification of authentic (real-life) data sets and the benefits of using authentic data in “Getting Messy with Authentic Data.” The classification scheme has 5 features; scope, selection, curation, size, and messiness. They define messiness as “the presence of variability, outliers, missing values and unexpected trends” (p. 4). Kjelvik and Shchultheis explain that students must understand that variability is inherent in data. As Kjelvik and Shchultheis point out the need for scaffolding, as students cannot jump right into a highly complex data set.

<u>Examining Data Literacy Skills of Students</u>

Due to the importance of data literacy skills, the examination of student data literacy skills and effective methods of classroom instruction have been carried out with students from all levels, elementary through college. There are some commonalities in studies that investigate instruction methods and student data literacy.

Established skill levels generally follow a pattern of moving from the ability to accomplish a statistical calculation to answering a statistical question. A statistical calculation involves reaching the single correct answer, while answering a statistical question has the student making claims, which there may be more than one, by interpreting what the calculations means within context (Tran & Lee, 2015). In a study working with middle school students, the move from calculations to questions was classified in 4 phases as they pertained to the construction of graphs: translating numbers from tables to graphs, extracting and organizing data to create tables and graphs, construction of different types of graphs, and finally interpreting graphs (Blagdanic & Chinnappan, 2013).

Another commonality is the use of open-ended questions. This is seen in a variety of ways, such as giving students a data set and then asking them to make a graph, but not instructing them as to the type of graph to make (Oslington, 2020). When studying students' ability to answer statistical questions, forming conclusions and supporting them, open-ended questions are the only way to assess to their ability. Due to the nature of open-ended questions, in which a student may answer some but not the whole of the question, partial credit is often given when assessing the success of a strategy (Blagdanic & Chinnappan, 2013).

Data Literacy, Case Studies and Authentic Data

In conclusion, the need for data literacy extends beyond a classroom setting and beyond the ability to perform statistical operations and the creation of graphical displays. As students leave the classroom and become participating members of society the analysis and interpretation of data in order to make informed decisions is a required skill set. The critical thinking and data literacy needed for this task as well as those needed in the fields of science to evaluate evidence may be improved through the use of case studies and real-world data. Case studies offer the interdisciplinary approach that many of our data-based decisions as citizens require and brings conceptualism to the learning process, which increases content knowledge. The use of authentic messy data presents students with genuine data seen within various fields, verses data presented in educational material or generated by the student.

METHODOLOGY

This purpose of this study was to see if using authentic messy data effected students' ability to describe and interpret data seen in graphs, asking the question, "How does the use of authentic messy data sets affect a student's understanding of data, as seen in their ability to read and interpret graphs?" This question was broken into 3 sub-questions. A treatment was implemented and data collected over the course of 18 weeks, data collection included: assessments, surveys and interviews. Table 2 outlines the instruments used for each sub-research question of the project.

Table 2. Triangulation matrix for data collection of each sub-research question.

Sub-research Questions	Instruments used
1. What effect does the use of data-based case studies have on data literacy?	Pre- and post-assessments Pre- and post-Likert surveys Focus Group
2. How does the use of authentic messy data affect students' ability to describe graphs, including the ability to describe variation within data sets?	Pre- and post-assessments Pre- and post-Likert surveys Focus Group
3. How does the use of authentic messy data affect student's ability to reach data-based conclusions and evaluate the conclusions formed by others?	Pre- and post-assessments Pre- and post-Likert surveys Focus Group

A comparison between the treatment and non-treatment group using post-assessments, were used to see if the use of authentic messy data had any impact on data literacy, allowing for the evaluation of messy data. In order to gain insight into how the students felt about their own data literacy skills, Likert surveys and focus groups were used. Once again, the treatment and non-treatment group was compared. To investigate the impact of the use of case studies, the same pre- and post-assessments, Likert surveys, and focus group interviews were used to see if there was skill growth within each group. The research methodology for this project received an exemption by Montana State University's Institutional Review Board and compliance for work with human subjects was maintained (Appendix A).

Demographics

This project focused on my three class periods of Research 1 students at TVMSC. Students travel to TVMSC from other schools around the Boise School District and from outside the district. The Research 1 class contains ninth cohort students, comprised of mostly ninth graders, with a few eighth graders in each period. Since the class is a mix of grades and home districts, students in the class may come from middle schools (6^{th}-8^{th} grade), junior highs (7^{th}-9^{th} grade), and two different types of high schools (9^{th}-12^{th} or 10^{th}-12^{th} grade). For their science course, most students are enrolled in Accelerated Chemistry, but a few in each class are taking Biology. As seen with the demographics presented in Table 3, periods one and six were the most similar with a majority of the students taking Integrated Math III (Algebra 2) and a couple of students taking Integrated Math II (Geometry).

Table 3. Research 1 student demographics.

	Period 1	Period 2	Period 6
Total Students	23	25	20
Gender Distribution			
• Female	8	9	8
• Male	15	19	12
Grade Distribution			
• 8^{th}	4	5	4
• 9^{th}	19	20	16
Current Math Class			
• Geometry (Integrated II)	1	1	1
• Algebra 2 (Integrated III)	22	12	19
• Pre-calculus	0	10	0
• AP Calculus AB	0	2	0

Parents and guardians were notified of the study before any treatment started to ask for permission to include their student. While no parent or guardian removed their student from the study, not all students were in focus group pool. The focus groups consisted of six randomly selected students from each group, with three females and three males. The focus groups happened during class time, which can cause stress for some students. Some students were never added to the pool of potential interviewees due to Individual Learning Plans (IEPs) that asked to limit changes in their classroom routine. Additionally, if students were randomly selected to be in the focus group but did not feel comfortable losing work time they could opt out. For these reasons the focus group random selection pool narrowed slightly.

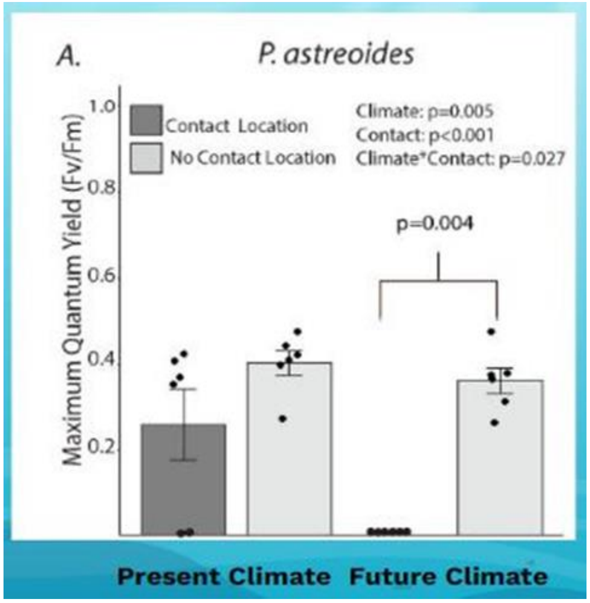
Treatment and Data Collection

For this project the treatment consisted of creating case studies that contained messy real-life data, while the non-treatment group received clean real-life data (Appendix B). The entire treatment was broken into focus periods where students worked with case studies for a period of time that had one type of data display (graph). In total there were three focus periods: box-and-whisker plots, scatter plots and bar graphs with error bars. Keeping in mind the need to scaffold when working with complex data sets, students started with two display types they were most familiar with, box-and-whisker plots and scatter plots. During the focus period students were then given a series of case studies with the graph of focus. While working on case studies students were allowed to converse with one another, but each student completed their own work. Students completed a pre-assessment, a case study with clean data, and a Likert survey in which they rated their confidence before the start of each focus period. At the end of the focus period the students took the same Likert survey again and completed a post-assessment, once again a case study with clean data. Open-ended questions were used on the assessments and partial credit was given on student responses.

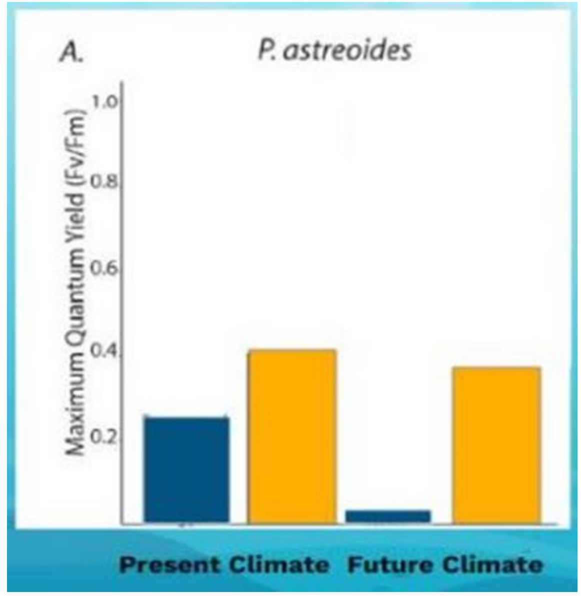
A slight switch was made with the final focus period. Since both groups were assessed with clean data, it became apparent that perhaps any effect the authentic messy data was having was not being seen. For the focus period on bar graphs with error bars, half of the individuals in the treatment group and half of the individuals in non-treatment group were assessed with authentic messy data.

Creating the Case Studies and Treatment Groups

Each case study focused on one type of graph and asked students to identify and describe aspects of the data, to form conclusions about the data and to evaluate the author's claim regarding the data. As the focus of case studies was to evaluate data literacy skills and not science concepts, none of the case studies involved topics they were currently studying, yet they required a low level of conceptual understanding of the case study's subject. This allowed for the data literacy skills to be assessed while lowering the chance that a student may be relying on their content knowledge. Case studies were created by either collecting published data from several sources and combining it, summarizing published scholarly articles, or using published case studies meant for students. An example of authentic messy data and the clean data version can be seen in Figure 1, these examples are from a case on coral bleaching, a complete case study example can be found in Appendix B.



A



B

Figure 1. Data from *Effects of future climate on coral-coral competition* by Johnston et al. (2020). Graph A is from the original published article (authentic messy data) and graph B is the clean version.

After a case study was created or selected, the graphs were then modified to clean up the data. For example, in some cases sections of outliers were removed or best fit lines added to scatter plots.

The three Research 1 class periods were divided into two groups. The treatment group, period one, received case studies with unaltered graphs. Period one was chosen due to its demographic similarities with the other two class periods. The non-treatment group, periods two and six, received case studies with the cleaned-up graphs to resemble graphs typically seen in education material. Other than the graphs the case studies were identical.

<u>Examining Sub-question 1</u>

The analysis of data started with sub-question 1, “What effect does the use of data-based case studies have on data literacy?” At the beginning of each focus period students completed a Likert survey that recorded their confidence when working with the graph of that focus period

(Appendix C). The students also completed a pre-assessment case study, which were similar in style to those they would work with during the treatment period. When the focus period was complete students once again took the Likert survey regarding their confidence when working with that style of graph and they completed a post-assessment case study, once again these were similar in style to those completed during the focus period.

The summary statistics and a paired t-test, with $p<.05$, was used to compare the pre- and post-assessments within each treatment group in order to determine what effect, if any, the case studies had on students' abilities to describe, interpret and reach data-based conclusions. The pre- and post-Likert confidence surveys were also analyzed using a paired t-test, $p<.5$. After analysis was complete within each treatment group, the two groups were compared.

In addition to the assessments and Likert surveys, students were interviewed in a focus group about their experiences working with the case studies. The students that were randomly selected for the focus groups were interviewed with a pre-determined set of questions (Appendix D). Most of the initial questions were "yes or no" or "would you rather" questions that were then followed up with "why" and "how" questions. At the beginning of the focus group the students were shown an example of a clean data display and its authentic messy counterpart for comparison. During the focus group notes were taken by hand and sound from the interview was recorded using a cell phone to reference later. After the completion of the focus groups the answers were organized in a document and examined for themes. Answers were first organized by the response to the initial question ("yes or no" and "would you rather"), then commonalities within response groups were examined. During the examination some common language appeared, such as: outliers, context, and reliability. Reasonings behind initial responses and treatments groups were compared.

Examining Sub-question 2 and 3

The results of the two treatment groups were compared in order to examine the sub-questions: “How does the use of authentic messy data affect students’ ability to describe graphs, including the ability to describe variation within data sets?” and “How does the use of authentic messy data affect students’ ability to reach data-based conclusions and evaluate the conclusions formed by others?”

Along with assessment totals, the post-assessments were broken into two categories: describe and interpret. The describe category contained open-ended questions that asked students to identify or describe aspects of the display. The interpret portion asked students to form conclusions and evaluate the conclusions formed by the author of the original data set. Comparisons of the post-assessment data between the treatment group and non-treatment group was made using a sample two-tailed t-test ($p<.05$) and 2xSE (standard error). The post-Likert surveys were also compared using a sample two-tailed t-test, $p<.05$. The results from the focus groups were compared for any similarities or differences in overall answers and themes.

DATA ANALYSIS

The results of the data analysis indicated that the use of authentic messy data had no impact on student data literacy skills, but that the repeated exposure to both messy and clean data did result in increased confidence of data literacy skills. Students from both groups did say that the use of case studies and the context it provided for the data made the evaluation of data easier. Finally, student answers during the focus groups suggest that the use of authentic messy data may lead students to think more critically about data analysis.

Authentic Data and Data Literacy Skills

Multiple assessments, Likert-surveys and focus groups were used to assess the sub-questions, "How does the use of authentic messy data affect the data analysis skills of students, including describing and comparing data sets?" and "How does the use of authentic messy data affect student ability to reach data-based conclusions and evaluate the conclusions formed by others?" The results of the assessments for the data displays conflicted with those of the Likert-surveys and focus group.

Assessment Results

With box-and-whisker plots, the treatment group performed better on the post-assessment (M = 85%, SE = 0.13) than the non-treatment group (M = 78%, SE = 0.17), when comparing using standard error (Figure 2). The comparison using s sample t-test showed that this difference was insignificant (p>.05). When comparing the assessment categories, describe and interpret, there was also no significant difference (p>.05) between the treatment and non-treatment group.

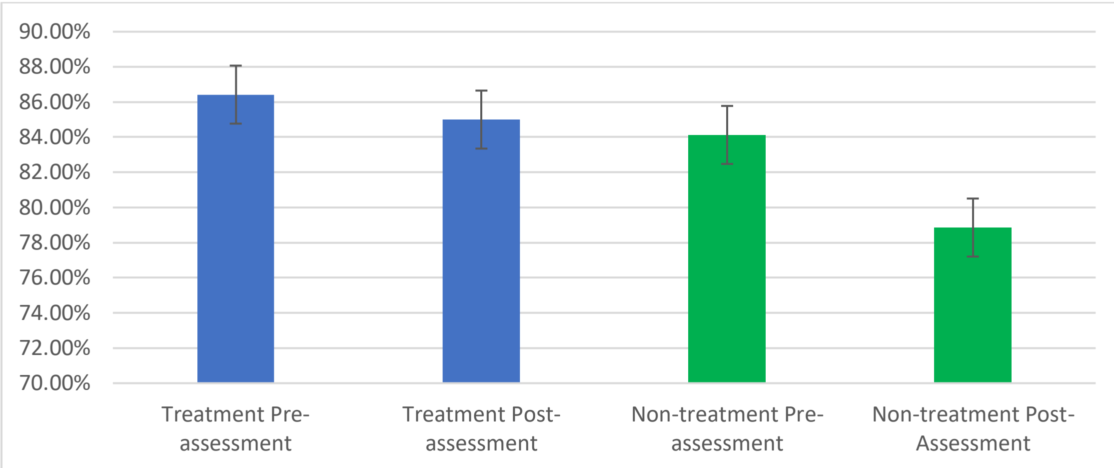


Figure 2. The pre- and post- box-and-whisker plot assessment averages for the treatment (*n*=18 and non-treatment group (*n*=35).

The focus period on scatter plots yielded results which saw the two groups performing almost identically on the post-assessment (Figure 3), with no significant difference between the post-assessment results of the two groups ($p>.05$). A comparison of data literacy skills asking students to describe or identify aspects of the display on the post-assessment showed no difference between the two groups ($p>.05$), but a comparison of the interpret skills showed that the non-treatment group performed better ($p>.05$).

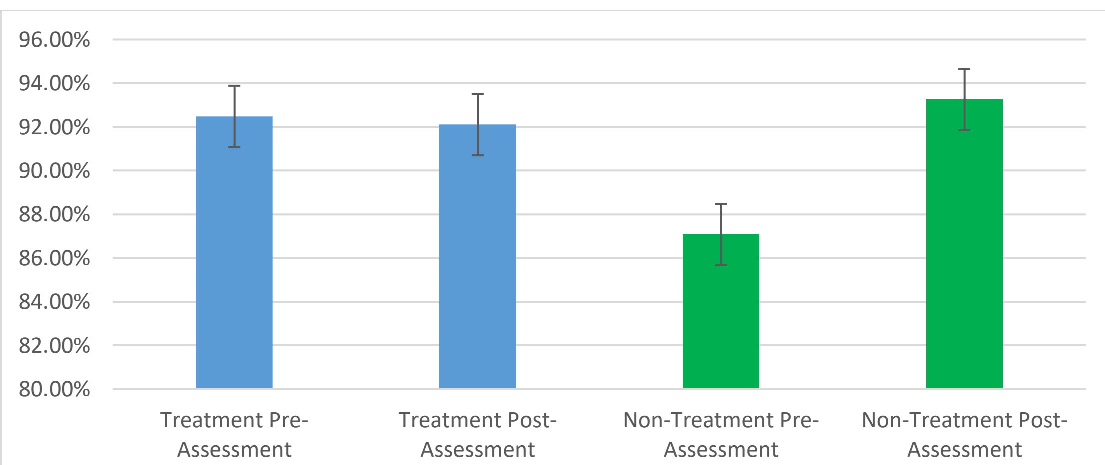


Figure 3. The pre- and post- scatter plot assessment averages for the treatment (*n*=19) and non-treatment group (*n*=42).

The lack of gains in scatter plot and in the box-and-whisker plot assessments could be due to prior experience with both display types before the start of the treatment period. The experience is most likely the cause of somewhat high pre-assessment scores, creating a possible ceiling effect. Additionally, the overall high assessment scores may have impacted any effect the use of authentic messy data had on treatment group, resulting in insignificant differences between the two groups.

The last display type used for assessing whether or not the use of authentic messy data effects student data literacy skills were bar graphs with error bars. While the students had experience prior to the treatment period with bar graphs and histograms they had little to no experience using error bars with those displays. Along with a more unknown data display, a different assessment strategy was used for this treatment period. Half the treatment group assessed with authentic messy data, while the other half assessed with clean data. Within the non-treatment group, half were assessed with authentic messy data and half with clean data. This was done to adjust for the potential that perhaps the repeat exposure to data was affecting results and not the use of authentic messy data. In the previous display focus periods, everyone was assessed with clean data.

The treatment and non-treatment group, regardless of the data type on the assessment improved from the pre- to the post-assessment. Those in the treatment group that were assessed using clean data saw the smallest increase in assessment average, from 80% (*SD* = 0.14) to 83% (*SD* = 0.12). This slight change proved to be insignificant with a paired two-tailed t-test (p>.05).

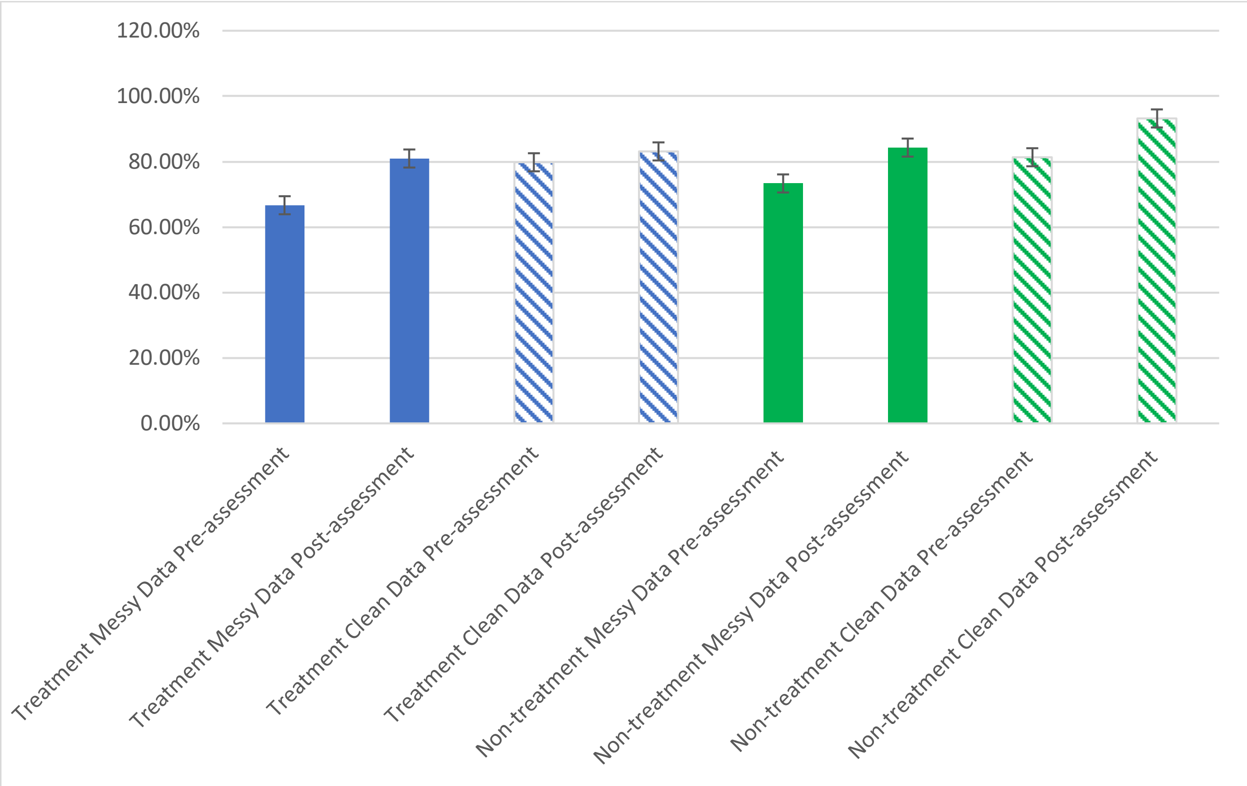


Figure 4. The pre- and post-assessment the focus period on error bars: treatment group assessed with messy data (*n*=9), treatment group assessed with clean data (*n*=11), non-treatment group assessed with messy data (*n*=20) and the non-treatment group assessed with clean data (*n*=21).

For the three other groups: treatment assessed with messy data, non-treatment assessed with messy data, and non-treatment assessed with clean data, the increase between the pre- and post-assessment was large enough to fall outside the range of standard error (Figure 4). The increase in the post-assessment scores was only significant for the students who were only given clean data, in the cases studies and on the assessments ($p<.05$).

A series of sample two-tailed t-tests were performed to compare the two groups and the assessment types within each group. A comparison within the treatment group, between those assessed with messy and those assessed with clean data resulted in no significant difference ($p>.05$), when examining assessment totals and question type of describe and interpret.

The lack of difference was also seen in the non-treatment group, for those assessed with messy data and those assessed with clean ($p>.05$), when comparing the total assessment scores and those pertaining to the ability to described the displays. A comparison of assessment scores pertaining to the interpretation data, had those that were assessed with clean data performing significantly ($p<.05$) better than those that those that assessed with messy.

A comparison between treatment and non-treatment group, also yielded insignificant ($p>.05$) results in regard to total assessment scores and describe assessment scores. The non-treatment group that assessed with clean data did outperformed the treatment group that was assessed with clean data ($p<.05$).

Overall, the assessment data shows that the use of case studies and/or repeated exposure to displays had minimal effect on data literacy skills when assessing skills related to familiar data displays. The use of authentic messy data had no statistically significant effect on data literacy skills regardless of familiarity. The standout group were those in the non-treatment group that worked with and assessed with only clean data, who improved significantly in two out of three display types.

Likert Survey Results

At the start and close of a data display focus period students were asked to complete a Likert survey to assess their confidence in their data literacy skills. The Likert survey options of Strongly Agree (4), Agree (3), Disagree (2), and Strongly Disagree (1), were used. The results were used to measure whether or not the students thought that they had improved.

Each of the three focus periods saw an increase in overall student confidence in their data literacy skills. A comparison of the post-Likert survey totals using a sample two-tailed t-test between the treatment and non-treatment group revealed that there was no significant difference in their confidence levels regarding scatter plots, $t(61) = -0.52$, $p = .61$ and the use of error bars, $t(60) = -0.68$, $p = .50$. The difference in the box-and-whisker plots post survey total did prove significant, $t(38) = -2.46$, $p = .02$, with the non-treatment group having higher overall confidence in their skills. While both groups grew in their confidence, the difference in confidence levels was marginal, suggesting that the use of authentic messy data did not affect confidence, and even may have hindered it.

Focus Group Results

Six students from each group were randomly selected to participate in a focus group at the end of the treatment period. During the focus group they were asked questions about which type of data was easier to evaluate. These questions were designed to see how the two groups viewed the different types of data.

When asked which type of data, messy or clean, was easier to describe, all but two students, both from the non-treatment group, chose clean data (Appendix D Tables 1 and 2). One student responded that clean data “looked better,” while another stated “it is easier for the general public.” The two students that chose messy data, said that it had more information to work with.

When asked which type of data was easier to speculate as the cause of variation, all of the students in the treatment group chose the messy data. Their reasoning stemmed from being able to see the whole picture and therefore any trends that were present. Half of the students in the non-treatment group chose clean data as easier, with one student saying that “clean data is easier to understand”, a statement that the other two agreed with.

The last question asked the students if messy or clean was easier to form conclusions from. For this question, all but two stated clean data, with everyone in the treatment group choosing clean. While the treatment group all said clean data was easier, they went on to have a discussion about how clean data needed to be looked at critically. One student stated that the data "might have been cleaned too much", another said that clean data was "less reliable" and they all agreed the messy data would lead to a more accurate conclusion. Those is the non-treatment group that chose clean data only mentioned that it was easier to understand. Students from the non-treatment group that chose messy data, agree with the treatment group in that there would be more information to base a conclusion on.

Overall, there was no significant difference between the data literacy skills of students who worked with authentic messy data and those that did not. A difference was seen during the focus group when the treatment group discussed the reliability of clean data.

Case Studies and Data Literacy

The sub-question "What effect does the use of data-based case studies have on data literacy?" was addressed using assessments, surveys and focus groups. Every time the students worked with data, messy or clean, the data was presented in a case study. The case study included background information on the topic, the question posed by the researchers, data collection methodology and the results. The only thing that differed was the results section, with the treatment group receiving the authentic messy data and the non-treatment group receiving clean data. The effectiveness of using case studies on data literacy were mixed, with assessments showing no impact, while surveys and interviews suggest that they may have had an impact.

Assessment Results

A paired two-tailed t-test was used to compare pre- and post-assessment results for each group and for each display type. The only group that showed a significant increase between the pre- and post-assessment was from the non-treatment group when working with scatter plot displays, $t(41) = -2.13$, $p = .04$, and for those that assessed with clean data involving error bars, $t(20) = 2.76$, $p = .01$. When using gain of averages to look for growth between the pre- and post-assessments, the gains were <0.001 for both groups on every display type. Once again, these results may be caused by the high pre-assessment averages, at least for box-and-whisker plots and scatter plots.

Likert Survey Results

The results of Likert surveys on confidence differed from those of the assessments. For this evaluation the pre- and post-survey data from the two groups were combined. When a paired two-tailed t-test was used to compare the pre- and post-Likert survey results, there were significant increases in confidence in both groups and with all three display types ($p < .05$). The gains in confidence were most notable in the areas of interpreting data, as seen in the forming of conclusions, and evaluation of an author's claim.

For box-and-whisker plot displays, prior to the treatment period, students were the most confident with describing a single box plot (Strongly Agree = 53.45%, n=58). Students were least confident in evaluating an author's claim (Strongly Agree = 10.34%, n=58) and speculating on variation (Strongly Agree = 8.62%, n=58). Strongly Disagree was only chosen 3 times throughout the survey (Figure 5).

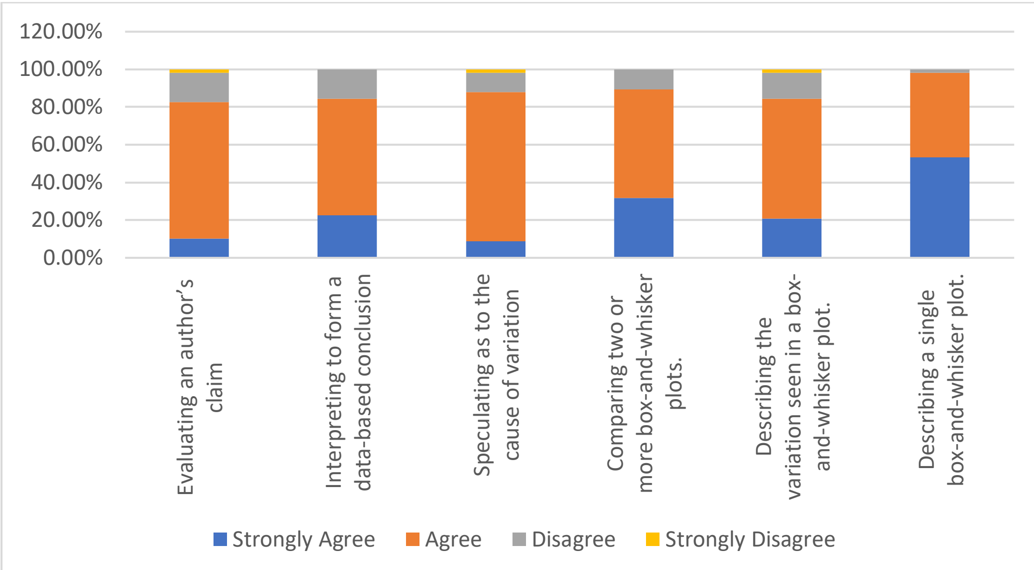


Figure 5. Pre-treatment Likert survey for box-and-whisker plots responding to statements of "I am confident…" (*n*=58).

After the treatment period the average score on the Likert survey increased from 77.68% to 83.69 (*n*= 58), this increase showed to be significant with a paired two-tailed t-test, t(57) = -3.57, $p < .001$). Significant increases in student confidence regarding evaluating an author's claim ($p <.001$) and in speculating on variation occurred ($p < .05$) (Figure 6).

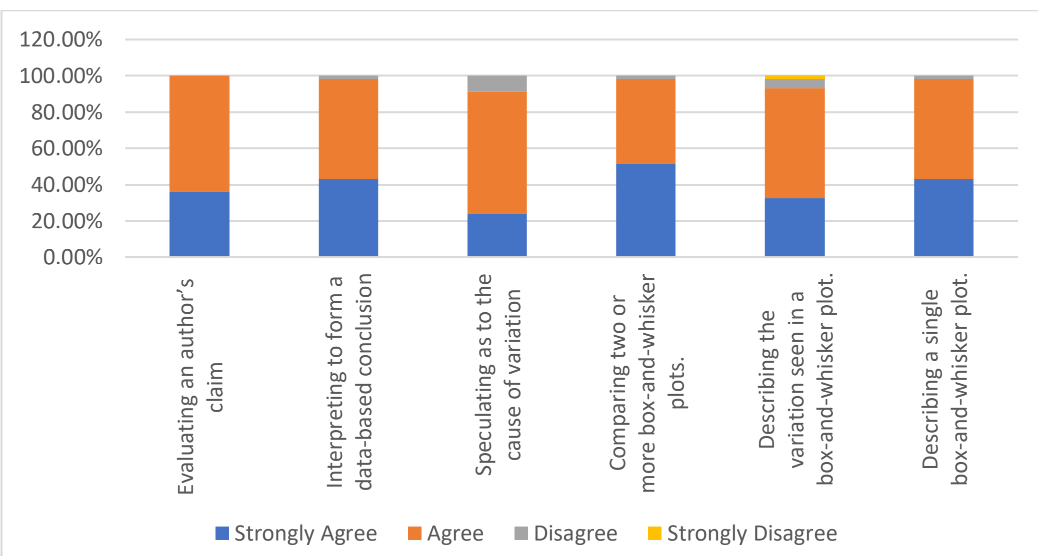


Figure 6. Post-treatment Likert survey for box-and-whisker plots responding to statements of "I am confident…" (*n*=58).

The Likert surveys for scatter plot displays had similar results, once again in response to the statement "I am confident evaluating an author's claim…", Strongly Agree had the lowest percentage (19.35%, n=62). Unlike the box-and-whisker plot survey fewer students were confident in their ability to describe the display (Strongly Agree = 25.81%, n= 62). Additionally, only 27.42% (n= 62) of students expressed strong confidence in their ability to draw a line of best fit and then predict values from that line (Figure 7).

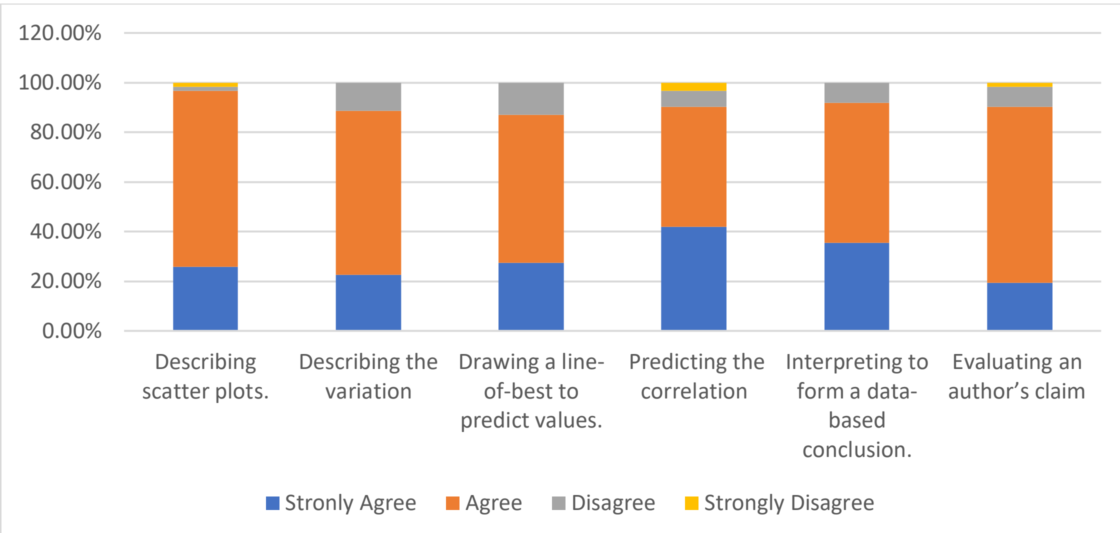


Figure 7. Pre-treatment Likert survey for scatter plots responding to statements of "I am confident…" (n=62).

The students' confidence significantly increased in all areas post treatment period, as the survey average went from 79.63% to 83.93%, $t(61) = -3.79$, $p < .01$. The three areas where students had fewest number of Strongly Agrees, all had increases. Two of the increases showed to be significant with confidence for describing variation and drawing a best fit line increasing significantly (p,.05), while evaluating an author's claim was not ($p > .05$) (Figure 8).

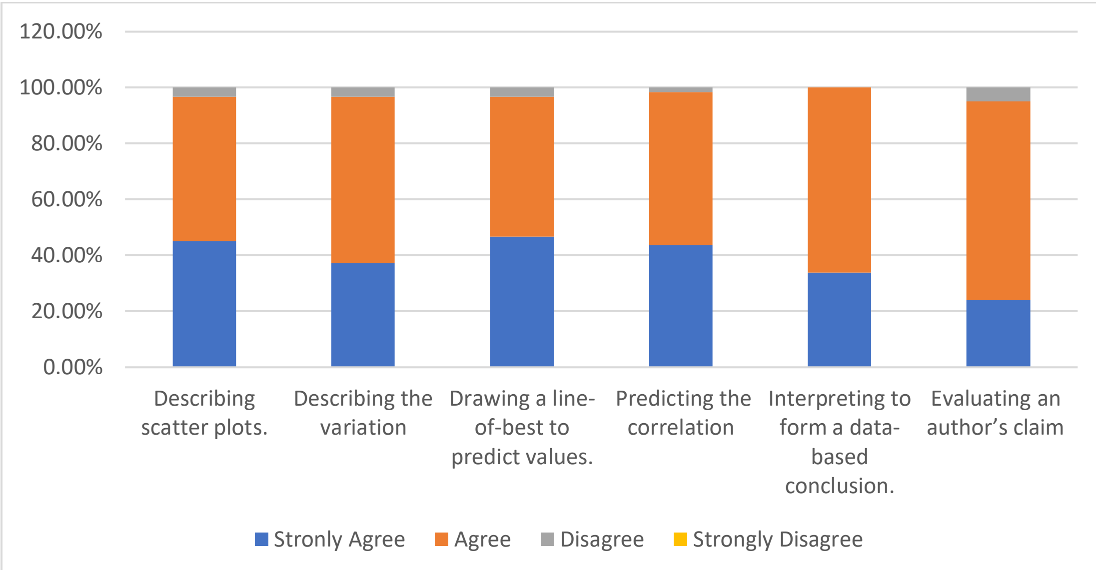


Figure 8. Post-treatment Likert survey for scatter plots responding to statements of "I am confident…" (*n*=62).

Unlike the previous two displays the students had no experience within the scope of the class with error bars, but they did have extensive experience with bars graphs. As seen in Figures 9 and 10, both classes did not respond with Strongly Agree to the confidence statements as often, compared to the previous display types. For the statement, "I am confident in my ability in using error bars in bar graphs" and "…drawing conclusions from bar graphs with error bars", had more students responding with Disagree when compared to any other Likert survey question.

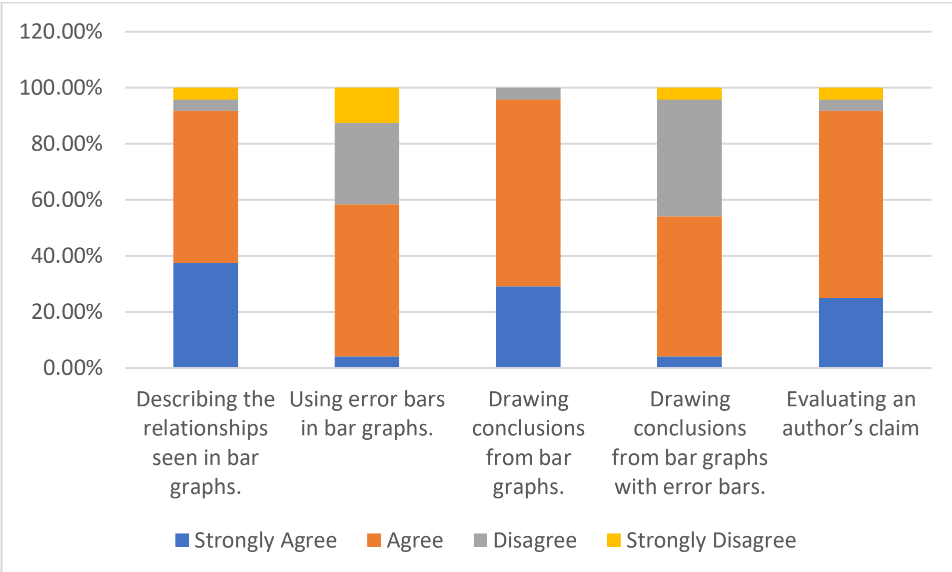


Figure 9. Treatment group's pre-treatment Likert survey for error responding to statements of "I am confident…" (*n*=24).

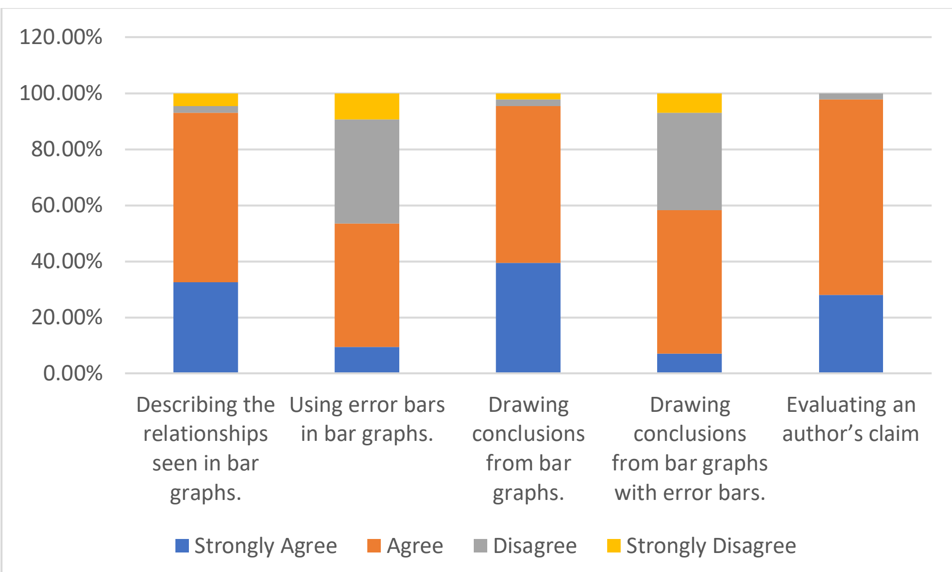


Figure 10. Non-treatment group's pre-treatment Likert survey for scatter plots responding to statements of "I am confident…" (*n*=38).

The post survey resulted in overall increase for both groups (Figures 11 & 12), the treatment group went from 14.66 to 16.7 out of 24, while the non-treatment went from 14.9 to 16.48. The increases seen were significant when compared with a paired two-tailed t-test (treatment p <.01 and non-treatment p < .001). This is interesting to note because the non-treatment group did not work with error bars, yet stated that their confidence increased.

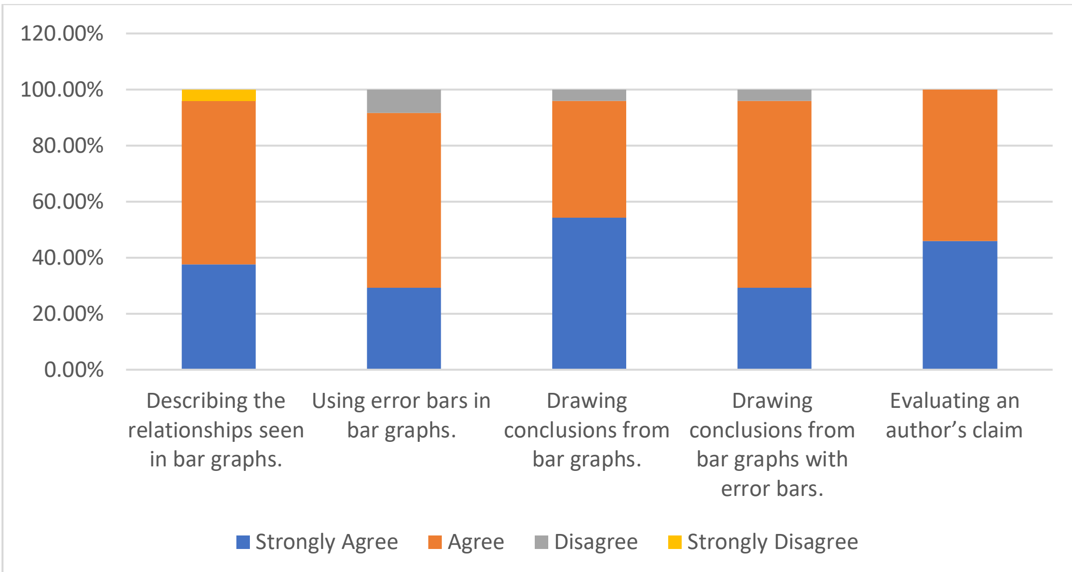


Figure 11. Treatment group's post-treatment Likert survey for error responding to statements of "I am confident…" (n=24).

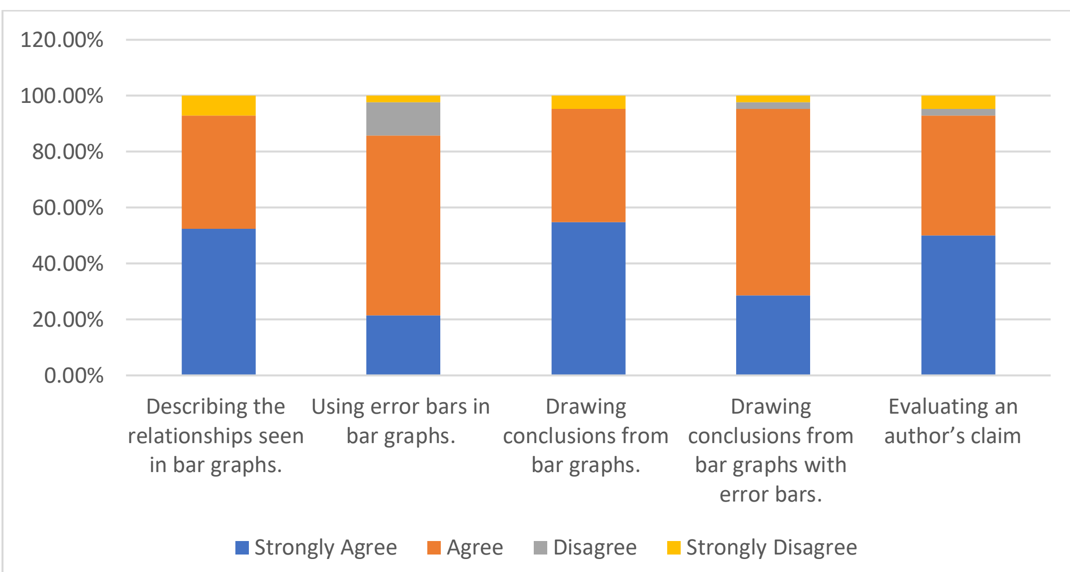


Figure 12. Non-treatment group's post-treatment Likert survey for scatter plots responding to statements of "I am confident…" (n=38).

Focus Group Results

This increase in confidence seen in the Likert surveys could have been a result of repeated exposure to data displays. Although, the results of the Likert surveys were echoed by the focus group, which were asked questions pertaining to how they felt about the case studies and if they thought their skills had improved throughout the semester. The first question posed was, “What were your thoughts on getting data with background information and the data collection methods?” All the students, except one, stated that the background information was helpful, particularly when forming conclusions (Appendix D Table 3). The additional information, including the methods used by the researchers, “provided depth” and showed “why something was the way it was”. Two students agreed that case study information was helpful but stated that it wasn’t necessary for describing the displays.

The second question posed to the focus group asked the students “Compare your data literacy skills from before the use of case studies to now.” All the students responded that their data literacy skills had improved from the start of the treatment period (Appendix D Table 4). Several students mentioned that the case studies aided them when looking for patterns in the data, drawing conclusions, and understanding context. Two students mentioned that they thought the case studies should have increased in difficulty.

Results Summary

The use of messy data within case studies did not improve the data literacy skills within the treatment group, although those who were exposed to clean data and assessed with clean data did see improvement in two of the three data display types. When interviewed those that worked with messy data were the only students to discuss data critically, asking how data was cleaned and if it could be cleaned too much. Suggesting that the students had realized that clean displays may not be telling the whole data story.

The use of case studies did bring about greater confidence in student data literacy skills, but their confidence was not seen in their performance. This may be tied partially to a ceiling effect of high pre-assessment scores on some of the display types. Student focus groups shared that the information provide in the case studies made it easier for them to form conclusions about the data and evaluate those formed by others. The students noted, regardless of the group they were in, that how the data was collected and information about the population of interest impacted how they then viewed the final results.

Claims From the Study

The answer to the project focus question of, “How does the use of authentic messy data sets affect student understanding of data, as seen in their ability to read and interpret graphs?” was mixed. The use of authentic messy data did not have a significant impact on student data literacy skills, but it may have increased student awareness to think about data critically. While it was not seen in the assessment results if the use of case studies increased data literacy skills, the context that they provided was used by the students when examining the variation seen in the data displays, which impacted the data-based conclusions they formed and the evaluation of authors’ claims.

Use of Messy Data

Sub-questions 2 and 3 focused on the use of messy data in the improvement of student data literacy skills. The assessment and Likert survey results for sub-question 2, “How does the use of messy data affect students’ abilities to describe graphs, including the ability to describe variation within data sets?”, saw no significant difference between the two treatment groups. When comparing assessment prompts associated with describe sample t-test results had *p*-values >.05. For sub-question 3, “How does the use of data affect students’ abilities to reach data-based conclusions and evaluate the conclusions formed by others?”, no difference was also seen when assessing prompts that had the students interpreting data and in Likert surveys.

Student growth in their data literacy skills may have been limited to a ceiling effect due to high total pre-assessment averages, over 85%, for box-and-whisker plots and scatter plots.

This was not the case with displays using error bars, where pre-assessment averages ranged from 66% to 84%. A distinction was seen between the two groups during the focus group interview.

The students from the treatment group who were interviewed discussed the reliability of data. Members of the group stated that clean data, might be too clean, and asked what if important information was removed. In a lesson discussed in a publication of the National Science Teachers Association, students were given a large messy data set that they then had to analyze. The author Gould et al. (2022) noted that students had discussions on which data to keep and what not to keep for the final product, and what criteria would be used. During the focus group the students that worked with messy data wondered the same thing about clean data, what had been removed and how was that decided. This discussion on the reliability of the data suggests that the students from the treatment group were taking a more critical look at the data they had been presented with.

The discussion on reliability stemmed from how data displays were constructed and the degree of variability that they could show. In response to sub-question 2, the focus groups showed that authentic messy data may impact how students view variability in data. Just as Kjelvik and Shchultheis (2019) described, the students were becoming aware that data has inherent variation. Since their view of variability was impacted, then in response to sub-question 3, their ability to form conclusions may also have been impacted. Students in the treatment group were reaching the same conclusions regarding the case studies that the non-treatment group were reaching. This may mean that they were able to seen through the variation and still find the relationships, or lack of relationships, between variables present in the data displays.

Use of Case Studies

Case studies were used as means to present the authentic data displays, messy or clean, to the students, and their use as a method for improving data literacy skills was examined. There was no non-treatment group for the use of case studies, as such student feedback provided the best insight to their effectiveness. In response to sub-question 1, “What effect does the use of data-based case studies have on data literacy?”, case studies allowed students to make more informed conclusions regarding data displays. The Likert surveys saw a significant increase in student confidence in their data literacy skill for all three display types. Every student from the focus groups, stated that the background information and methodology found in the case studies helped them in the interpretation of data, increasing their ability to form conclusions from the data and evaluate the conclusions formed by others. The information provided by the case studies allowed the students to critically think about the data by seeing the data in context.

A study performed by Khan et al. (2015) found that students instructed with case-based learning, had significantly higher post-test scores and that students reported that their ability to analytically reason increased. In another study that used case studies in nursing education, found that their use improved practical skills but also the students’ ability to analyze a situation (Qi et al., 2018). In both of those studies and in my project, constructivism was at play, students were asked to construct new knowledge using previous knowledge and experiences.

In this project students did not have previous experiences with the subjects of the case study, for example: heavy metals found in chocolate, citizen science to predict malaria outbreaks, ages of world cups teams, and extinct birds of New Zealand. In every case study the students had to bring in their own experiences regarding the topics such as malaria, data collection, experimental design or sampling. They were able to use these experiences and apply them to the data they were seeing because the case study provide context for the data. The context gave the data a story (Herreid, 2014). The story provided by the case study allowed students to have a greater understanding for variation within data, or as they said, “know why the data is the way it is”.

Value of the Study and Consideration for Future Research

As member of a community, most of us, including our students, will consume more data than we produce (Hemo, 2020). In that consumption we will be asked to come to reach our own conclusions or we will need to evaluate the conclusions reached by others. Data literacy, comprised of reading and analyzing graphs and viewing data critically, is the skill set needed to come to those decisions (Weiland, 2017). The data that students will be asked to interpret during their lifetime may not look like the nice graphs and tables that they see in their educational materials. Exposing students to authentic messy data gives them the chance to practice a higher level of data literacy skills they may need.

In this action research project, the impact of authentic messy data was not seen in the assessments or surveys, but was seen in discussions with students and when watching them interpret difficult displays. The questions posed by the treatment group regarding outliers, overall variation in the data and questing the process used to “clean up” data, suggest that exposure to authentic messy data did impact how they thought about the data analysis process. The questioning of the data analysis process, which data to use and not use in displays, shows that students were taking the first steps at viewing data critically.

The use of case studies to present data also brought about a change in how students view data variability. Every student in the focus groups mentioned that the information provided in the case study assisted in their ability to form conclusions about the data they were analyzing. They took into account the population being studied and the methods of data collection as potential reasons for variability presented in the data displays. In conversations prior to this project students had mentioned that individual data points within a data set that may not follow the set trend closely could be due to error or faulty technology instead of realizing that data will naturally have variation. The case studies provide the backdrop for the data to sit, giving it meaning and context (Erwin, 2015).

Future studies involving authentic messy data could examine its use in a way that is more aligned with curriculum content, method of presenting the data, and having students start with a large messy data set and then go through the process of "cleaning" it up. A limit to this study was the use of the Research 1 class. The class is composed of accelerated students who are enrolled in different math classes. Additionally, as this was not a specific science class, the case studies did not have a curriculum focus. While this allowed for the testing of data literacy skills without the reliance of content knowledge, the importance of authentic messy data may be highlighted when presented in the context of a curriculum. Since all the data presented was out of the context of the class, the case studies used were limited to the most basic type. Varying case studies may garner more interest from the students. Finally, students may better grasp the need to view data critically if they can see how data can be altered.

Impact of Action Research on the Author

Implementing this action research project in my classroom has brought changes to how I teach, in the form of reflecting on student performance and teaching methods. Prior to this study my mindset regarding measuring the effectiveness of a lesson was limited to pre- and post-assessments. Through doing this project and reading other action research projects, I find myself wanting to collect new types of data. While I did not do it for my project, I would like to monitor how students interact with material, what conversations are they having during a lab or activity. The mindset of different forms of data collection sounds fun and the comparison aspect that has me excited. At the end of each school year, my mind turns to the upcoming year. These thoughts always include a list of things to improve on or try. My thoughts at the end of the school year are different. I already have in mind a new action research project and am thinking of what data I would collect and comparisons I could make.

The project had a fun and quirky impact on my relationships with my students. The students in my Research 1 class learn about sampling variations, control groups, summary statistics, they are asked to create graphs, write papers, make posters, and do presentations of their work. All of which is what I have been asked to do. Going through a project and having them along with me made for some interesting conversations. For example, when creating the focus groups students wanted details in how people were randomly selected, they asked questions on how the treatment group was chosen and of course they wanted to know the results. They would ask how this paper was going, with us all agreeing that the methodology and results sections are the easiest to write and asking what statistical tests I was using. This project placed me in my students' shoes, making me more empathetic to their plights as researchers and increasing my ability to help in their research.

REFERENCES CITED

AP Biology. (2019). *College Board*. https://apcentral.collegeboard.org/pdf/ap-biology-course-and-exam-description-0.pdf?course=ap-biology

AP Chemistry. (2019). *College Board*. https://apcentral.collegeboard.org/pdf/ap-chemistry-course-and-exam-description.pdf?course=ap-chemistry

AP Physics 1. (2019). *College Board*. https://apcentral.collegeboard.org/pdf/ap-physics-1-course-and-exam-description.pdf?course=ap-physics-1-algebra-based

Blagdanic, C., & Chinnappan, M. (2013). Supporting students to make decisions with real-life data. *Australian Mathematics Teacher, 69*(2), 4-11.

Bonicowska, A., Sanmartin, C., & Frenette, M. (2019). *Data literacy: What is it and how to measure it in public service.* Analytical Studies Branch, Statistics Canada. https://www150.statcan.gc.ca/n1/pub/11-633-x/11-633-x2019003-eng.htm

Boone, R. (August 13, 2021). *Census: Meridian one of the fastest-growing cities.* KTVB News. https://www.ktvb.com/article/news/local/growing-idaho/census-meridian-one-of-10-fastest-growing-us-cities-idaho-growth-census/277-5650756a-2c10-43c3-b1b3-87e6c45bb650

Bowen, M., & Bartley, A. (2014). *The basics of data literacy.* National Science Teachers Association.

Casey, J. (2010). Interdisciplinary approach: Advantages, disadvantages, and the future benefits of interdisciplinary studies. *ESSAI, 7*(26), 76-81.

Cognitive Constructivism. (2021). Berkley graduate division. https://gsi.berkeley.edu/gsi-guide-contents/learning-theory-research/cognitive-constructivism/

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

Forbes, S., Camden, M., Pihama, N., & Pfannkuch, M. (2011). Official statistics and statistical literacy: They need each other. *Statistical Journal of LOAS, (27),* 113-182.

Gibson, J. P. & Mourad, T. (2018). The growing importance of data literacy skills in life science education. *The American Journal of Botany, 12*(105), 1953-1956.

Gould, R, Sunbury, S., & Dussault, M. (2022). In praise of messy data: Lessons from the search for alien worlds. *National Science Teachers Associations, 81*(8), 31-36.

Hemo, I. B. (2020). *Why data literacy should be taught in high schools.* Forbes. https://www.forbes.com/sites/forbestechcouncil/2020/10/14/why-data-literacy-should-be-taught-in-high-schools/?sh=3e6119eb4161

Herreid, C. F., Schiller, N. A., & Herreid K. F. (2014*). Science stories you can count on: 51 case studies with quantitative reasoning in biology*. National Science Teachers Association.

Herreid, C. F., Schiller, N. A., & Herreid K. F. (2012). *Science stories using case studies to teach critical thinking*. National Science Teachers Association.

Johnston, K, Campbell, J., & Paul, V., & Hay, M. (2020), Effects of climate change on coral-coral competition. *PLoS ONE, 15*(8): e0235465. https://doi.org/10.1371/journal.pone.0235465

Khan, M. A, Qamar, K., Khalid, S. Javeed, M. M., Gondal, A., Zafar, H., Shoaib, A., Aman, R., Kahn, A., UlAbideen, A. Z., Bilal, A., & Imtiaz, F. (2007). Comparison of case-based learning with conventional teaching – student's perspective. *Pakistan Armed Forces Medial Journal, (65)*3, 415-419.

Khol, L. (2018). *Why tech companies and tech talent find Boise irresistible.* We Know Boise. https://www.weknowboise.com/blog/boise-tech-industry.html

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. *Life Science Education, (18)*2.

McLeod, S. (2019). *Constructivism as a theory for teaching and learning.* Simply Psychology. https://www.simplypsychology.org/constructivism.html

Mostert, M. (2007). Challenges of cased-based learning. *The Behavioral Analyst Today 8*(4), 434-442.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common core state standards for mathematics.* http://corestandards.org/

National Research Council. (2013). *Next generation science standards: For states, by states*. The National Academies Press. https://doi.org/10.17226/18290.

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. https://www.nextgenscience.org/

Oslington, G. (2020). What's in a graph? How student observations change over time. *Australian Primary Mathematics Classroom, 25*(1), 36-40.

Science Practices. (n.d.) College Board AP Central. https://apcentral.collegeboard.org/courses/resources/science-practices

Tran, D., & Lee, H. S. (2015). The difference between statistics and mathematics. *In Teaching statistics through data investigations MOOC-Ed, Friday Institute for Educational Innovation: NC State University, Raleigh, NC*. https://fi-courses.s3.amazonaws.com/tsdi/unit_2/Essentials/Statvsmath.pdf

US Census Bureau. (2019). *Quick facts: Boise City Idaho.* https://www.census.gov/quickfacts/boisecitycityidaho

Vahey, P., Rafana, K., Patton, C., Swan, K., van't Hoff, M., & Kratcoski, A., Stanford, T. (2012). A cross-disciplinary approach to teaching data literacy and proportionality. *Educ Stud Math 81*, 179-205.

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Education Studies in Mathematics*, *96*, 34-47.

What is Case-based Learning. (2020). *Queen's University: Centre for teaching and learning*. https://www.queensu.ca/ctl/teaching-support/instructional-strategies/case-based-learning#:~:text=What%20is%20Case%2DBased%20Learning,group%20to%20examine%20the%20case

Qi, M., Yi, Q., Mo, M., & Yang, Y. (2018). Applications of case-base learning in instructing clinical skills on nursing undergraduates. *Biomedical Research*, *29*(2), 300-304.

Zalles, D. (2005). *Designs for assessing foundation data literacy*. On the Cutting Edge: Professional Development for Geoscience. http://serc.carleton.edu/files/NAGTWorkshops/assess/ZallesEssay3.pdf

APPENDICES

APPENDIX A

IRB EXEPMTION

MSSE Exempt Request

Tue, Nov 30, 2021 at 8:24 AM

Dear Julie,

Thank you for your application. This email acknowledges receipt of the request for IRB Review and serves as the Approval Letter for your research. Your new **IRB Exempt Protocol # is JE113021-EX**.

Study Title: **Effects of Using Authentic Data in Classroom Instruction on Data Literacy Skills**

As the PI, it is your responsibility to facilitate subject understanding by informing subjects of all aspects of the project, providing an opportunity to ask questions, and describing risks and benefits of participation. Submit any new changes to the research protocol to the IRB via Amendment Form prior to implementing.

The research described in your submission is exempt from the requirement of additional review by the Institutional Review Board in accordance with 45 CFR 690.104(d). The specific paragraph which applies to your research is:

> (1) Research, conducted in established or commonly accepted educational settings, that specifically involves normal educational practices that are not likely to adversely impact students' opportunity to learn required educational content or the assessment of educators who provide instruction. This includes most research on regular and special education instructional strategies, and research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.

[Quoted text hidden]

Ekhoff-Reuer JE113021-EX.pdf
458K

APPENDIX B

CASE STUDY TREATEMENT AND NONTREATENT EXAMPLES

Name____________________

Using Citizen Science to Track Malaria

BACKGROUND

Malaria is a vector-borne disease, which causes fevers, headaches, muscle aches, vomiting and diarrhea. Vector-borne diseases are diseases that are transmitted between hosts through organisms. The plague is one of the most infamous vector-borne diseases, as fleas transmitted it between rats and people. A common vector-borne disease in Idaho is West Nile Virus, which mosquitoes spread. Like West Nile, Malaria is also spread through mosquitoes. Malaria and the mosquitoes that spread it is very common in Africa, where 400,000 people die each year.

Currently there are 12 professional mosquito monitoring stations around the country of Rwanda. These are used to track population sizes and types of mosquitoes in order to prevent Malaria outbreaks through using pesticides and other measures.

The 12 stations are not enough so one group of scientists in Rwanda are asking the public for help.

Science question: Can everyday citizens capture and monitor mosquitoes in order to predict Malaria outbreaks?

METHOD

The scientists' plan was simple: have people in five different villages capture mosquitoes and report how much of a nuisance they are. Using the public to track organisms is not new and is used in the US for bird and butterfly migration. Public involvement in research like these is called Citizen Science.

The villagers were taught how to create homemade traps using yeast and sugar, which they placed outside and inside their homes. To monitor the nuisance level, they were asked to rank how annoying the mosquitoes were on a scale from 1 to 5, with five being the most annoying. Finally they reported any cases of Malaria in their home during the research period.

RESULTS

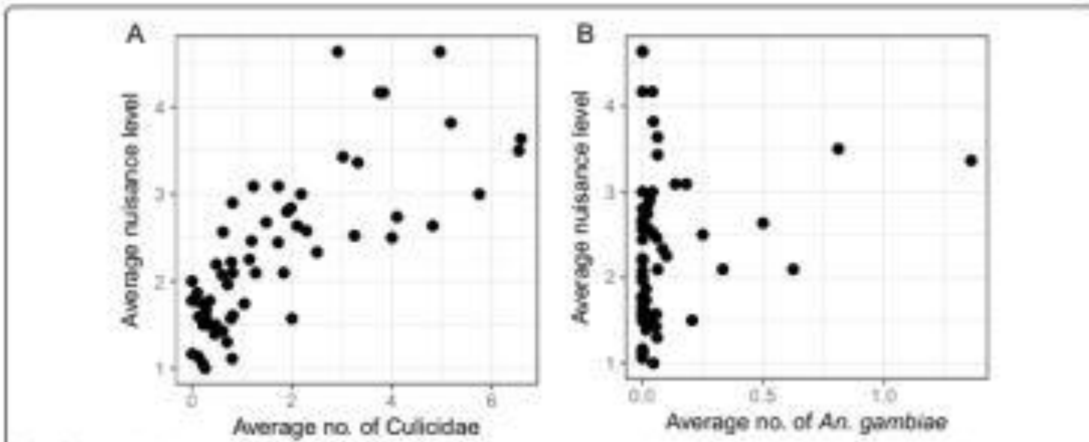


Fig. 6 Scatter plots showing the correlation between perceived mosquito nuisance and number of mosquitoes. The correlation is between average perceived mosquito nuisance per village per month and average number of mosquitoes (Culicidae, **A**) and *An. gambiae* s.l. (**B**) per village per month reported by the volunteers in five selected villages in Ruhuha sector, Rwanda

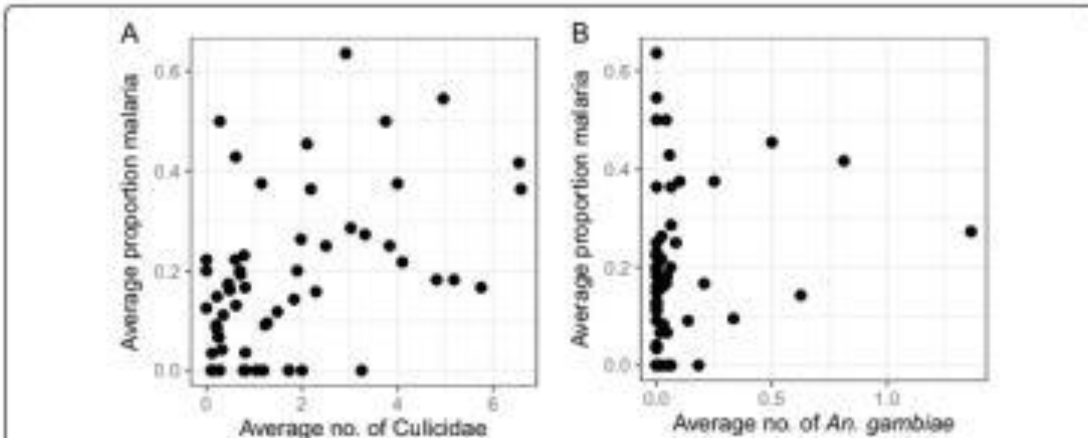


Fig. 8 Correlation between average proportion of malaria in households and number of mosquitoes. Scatter plots showing the correlation between average proportion of malaria in households per village per month and average number of mosquitoes (Culicidae) (**A**) and *An. gambiae* s.l. (**B**) per village per month reported by the volunteers in five selected villages in Ruhuha sector, Rwanda

Culicidae is the general name for mosquitoes. *An. gambiae* is a specific type of mosquito. Averages were found between the five villages in the study.

CONCLUSION

1. Describe figure 6 B (nuisance vs. *An. gambiae*), what type of correlation does it have?

2. Predict the nuisance level when the average number of cuculidae is 6.

3. Which could be better to use to predict the number of Malaria cases, the average number of culicidae (total mosquitoes) or *An. gambiae*? Support your claim with evidence from the graph. Use complete sentences

4. The scientists are hoping that citizen scientists will be able to help monitor mosquitoes in Rwanda that could be used to predict Malaria outbreaks. They claim that a citizen science network could be helpful when other resources are not available or limited. Do you agree with this claim? Can the reporting of nuisance be used to predict Malaria outbreaks and therefore prevent them? Support your reasoning with evidence from the graphs. Use complete sentences.

References

Map of Africa with Rwanda highlighted [online image]. (n.d.). https://www.worldatlas.com/maps/rwanda
Milumbu Murindahabi, M., Hoseni, A., Corné Vreugdenhil, L.C., van Vliet, A., Umupfasoni, J., Mutabazi, A., Hakizimana, E., Poortvliet, P., Mutesa L., Takken, W., and Koenraadt C. (2021). Citizen science for monitoring the spatial and temporal dynamics of malaria vectors in relation to environmental risk factors in Ruhuha, Rwanda. *Malaria Journal* 20(453).
Pirnata, M. (n.d.) Image of a mosquito [online image]. https://www.theatlantic.com/science/archive/2021/06/dengue-mosquitoes-defanged/619161/

Name____________________

Using Citizen Science to Track Malaria

BACKGROUND

Malaria is a vector-borne disease, which causes fevers, headaches, muscle aches, vomiting and diarrhea. Vector-borne diseases are diseases that are transmitted between hosts through organisms. The plague is one of the most infamous vector-borne diseases, as fleas transmitted it between rats and people. A common vector-borne disease in Idaho is West Nile Virus, which mosquitoes spread. Like West Nile, Malaria is also spread through mosquitoes. Malaria and the mosquitoes that spread it is very common in Africa, where 400,000 people die each year.

Currently there are 12 professional mosquito monitoring stations around the country of Rwanda. These are used to track population sizes and types of mosquitoes in order to prevent Malaria outbreaks through using pesticides and other measures.

The 12 stations are not enough so one group of scientists in Rwanda are asking the public for help.

Science question: Can everyday citizens capture and monitor mosquitoes in order to predict Malaria outbreaks?

METHOD

The scientists' plan was simple: have people in five different villages capture mosquitoes and report how much of a nuisance they are. Using the public to track organisms is not new and is used in the US for bird and butterfly migration. Public involvement in research like these is called Citizen Science.

The villagers were taught how to create homemade traps using yeast and sugar, which they placed outside and inside their homes. To monitor the nuisance level, they were asked to rank how annoying the mosquitoes were on a scale from 1 to 5, with five being the most annoying. Finally they reported any cases of Malaria in their home during the research period.

RESULTS

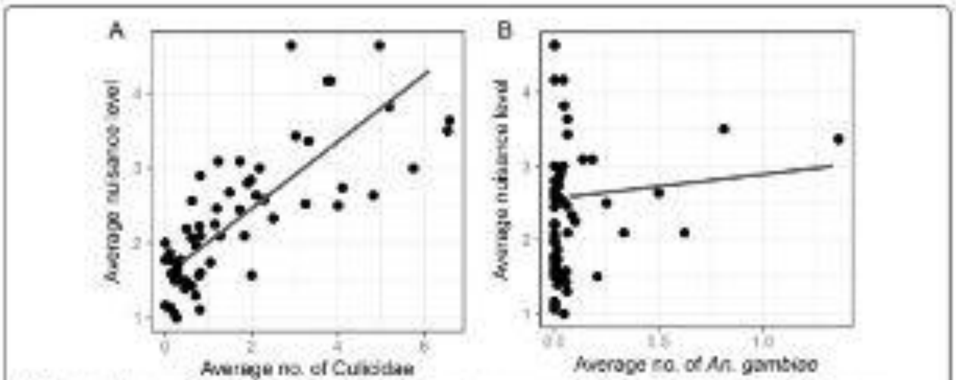


Fig. 6 Scatter plots showing the correlation between perceived mosquito nuisance and number of mosquitoes. The correlation is between average perceived mosquito nuisance per village per month and average number of mosquitoes (Culicidae; **A**) and *An. gambiae* s.l. (**B**) per village per month reported by the volunteers in five selected villages in Ruhuha sector, Rwanda

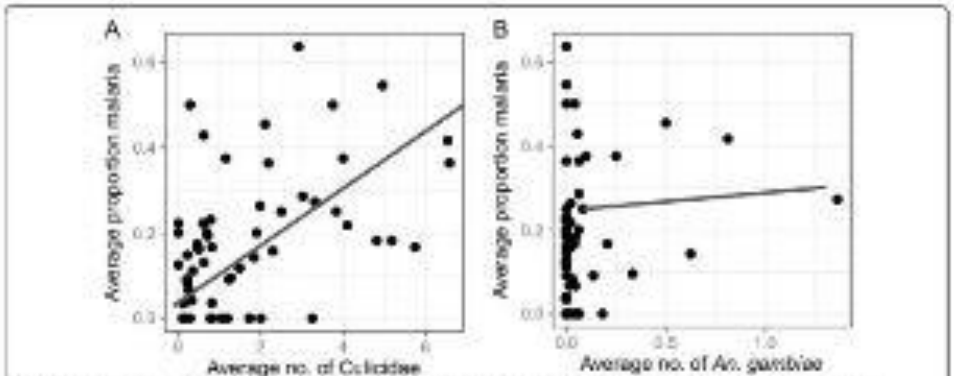


Fig. 8 Correlation between average proportion of malaria in households and number of mosquitoes. Scatter plots showing the correlation between average proportion of malaria in households per village per month and average number of mosquitoes (Culicidae) (**A**) and *An. gambiae* s.l. (**B**) per village per month reported by the volunteers in five selected villages in Ruhuha sector, Rwanda

Culicidae is the general name for mosquitoes. ***An. gambiae*** is a specific type of mosquito. Averages were found between the five villages in the study.

CONCLUSION

1. Describe figure 6 B (nuisance vs. ***An. gambiae***), what type of correlation does it have?

2. Predict the nuisance level when the average number of cuculidae is 6.
3. Which could be better to use to predict the number of Malaria cases, the average number of culicidae (total mosquitoes) or *An. gambiae*? Support your claim with evidence from the graph. Use complete sentences

4. The scientists are hoping that citizen scientists will be able to help monitor mosquitoes in Rwanda that could be used to predict Malaria outbreaks. They claim that a citizen science network could be helpful when other resources are not available or limited. Do you agree with this claim? Can the reporting of nuisance be used to predict Malaria outbreaks and therefore prevent them? Support your reasoning with evidence from the graphs. Use complete sentences.

References

Map of Africa with Rwanda highlighted [online image]. (n.d.). [illegible]

Milumbu Murindahabi, M., Hoseni, A., Corné Vreugdenhil, L.C., van Vliet, A., Umupfasoni, J., Mutabazi, A., Hakizimana, E., Poortvliet, P., Mutesa L., Takken, W., and Koenraadt C. (2021). Citizen science for monitoring the spatial and temporal dynamics of malaria vectors in relation to environmental risk factors in Ruhuha, Rwanda. *Malaria Journal 20*([illegible]).

Pesanic M. (n.d.) Image of a mosquito [online image]. [illegible]

APPENDIX C

PRE- AND POST-LIKERT SURVEY EXAMPLE

Confidence with Scatter Plots

Form description

1. I feel confident describing scatter plots. *

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree

2. I feel confident describing the variation in scatter plots. *

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree

3. I feel confident drawing a line-of-best for a scatter plot in order to predict values. *

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree

4. I feel confident predicting the correlation of a scatter plot. *

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree

5. I feel confident interpreting a scatter plot in order to make a data-based conclusion. *

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree

6. I feel confident evaluating an author's claim (data-based conclusion) through the examination and interpretation of data presented in scatter plots. *

- Strongly Disagree
- Disagree
- Agree
- Strongly Agree

Submit

Clear form

APPENDIX D

FOUCS GROUP RESULTS

Table 1. Treatment group responses to focus group questions regarding working with messy or clean data.

Messy or Clean Data Questions	Treatment Group Student Responses
Which type of data display is easier to describe?	• All students responded with clean data. • One student stated that clean data is easier for the general public
Which type of data display is easier to pose potential causes of variation?	• All students responded with messy data. • One stated that messy data provided a complete picture. • One student said that it was easier to see trends and that it provided a better overview.
Which type of data display is easier to from conclusions from?	• All students responded with clean data. • Two students said conclusions are more accurate with messy data. • One stated that someone might have "cleaned the data too much" and then it might not be reliable.

Table 2. Non-treatment group responses to focus group questions regarding working with messy or clean data.

Messy or Clean Data Questions	Non-Treatment Group Student Responses
Which type of data display is easier to describe? Why?	• Four of the students responded with clean data • One student said that the data looks better • One student said that the data is better organized • Two students said that messy would easier because it had more information
Which type of data display is easier to pose potential causes of variation? Why?	• Three students responded with clean data • Students that responded with clean data stated that it is easier to understand • Three responded with messy data • One student stated that they would be able to see all the outliers and clean data didn't show them
Which type of data display is easier to from conclusions from? Why?	• Two students responded with messy data • Students that responded with messy data said the there was more information to base a conclusion on • Four students responded with clean data • Students stated that clean data was easier to see what was happening • One student said that messy data was too squished • One student stated that with clean data trends "popped out"

Table 3. Treatment group responses to focus group questions regarding confidence with data.

Confidence with Data Questions	Treatment Group Student Responses
Would you rather interpret data that you have collected or interpret data that professional researcher has collected? Why?	• All students responded with data collected by a researcher. • The students responded with that there would be "less human error", "less variation", and "more experience" • Professionals would have less personal bias, that students sometimes try to make their data a certain way. • On student responded that when looking at researchers' data that strange values could be error or something else.
Would you rather from conclusions based on data that you have collected or a that of a professional researcher? Why?	• All students responded with data collected by a researcher. • One student said that they would have a greater understanding of the topic and this would make their data "better" • One student said that students are bias when collecting data for their projects. • Many students echoed the reasons stated for the interpretation question.
Scenario: You have placed data you have collected into a scatter plot and there is a point fairly far from the best fit line. What are your thoughts about that data point?	• One student stated that it was outlier, the others being interviewed then agreed. • One student mentioned that it may be due to natural variation or some other form of variation, the others being interviewed then agree.
Scenario: You have been given a scatter plot of a researcher's data and there is a point fairly far from the best fit line. What are your thoughts about that data point?	• All the students responded that it may be due to natural variation. • One student said that they would want to know how they collected the data. • One student responded that they would question how reliable the data is.

Table 4. Non-treatment group responses to focus group questions regarding confidence with data.

Confidence with Data Questions	Non-Treatment Group Student Responses
Would you rather interpret data that you have collected or interpret data that professional researcher has collected? Why?	• Four students would rather use researchers' data. • Twos students stated that researchers would have better methods and materials for data collection. • Two students said that the results would be more accurate and fewer errors • Two students would rather use their own data. • One student stated that the data would be less confusing and the experiment would be simpler • One student stated that like to have control over the data
Would you rather from conclusions based on data that you have collected or a that of a professional researcher? Why?	• All but one student selected the researcher's data • Several students stated that researcher data had less error and that the data would provide a more accurate conclusion • Two students stated that researchers know the big picture • One student stated that their own data was easier to from conclusions from because it was easier to understand
Scenario: You have placed data you have collected into a scatter plot and there is a point fairly far from the best fit line. What are your thoughts about that data point?	• Some of students stated that the data point might "skew" the results. • Several students stated that it resulted from human error • All the students stated that it would be an outlier and would ignore it • One student said it could be natural variation and should not be discounted
Scenario: You have been given a scatter plot of a researcher's data and there is a point fairly far from the best fit line. What are your thoughts about that data point?	• Most of the students stated that it would be an outlier and would ignore it • One stated that it could be due to a technical error • One student said that it could not be an outlier

Table 5. Treatment group responses to focus group questions regarding the use of case studies.

Using Case Studies Questions	Treatment Group Responses
What were your thoughts on getting data with background information and the data collection methods?	• One student said that it was helpful and they would be lost without it. • Several students said that it made forming conclusions easier. • One student stated that it was helpful with the analysis and with reaching conclusions. • One student stated that it wasn't needed to analyze/interpret the data displays, but was somewhat helpful in forming conclusions. • One student stated that it helped them figure out why there might be an outlier.
Compare your data literacy skills from before the use of case studies to now.	• All the students stated that their data literacy skills • One student said that they were able to see different types of variations and graphs • Several stated that being able to look at many types of graphs was helpful • Several students stated that they were better at drawing conclusions. • One student stated that it helped to know why they were looking at the data (the background information helped)

Table 6. Non-treatment group responses to focus group questions regarding the use of case studies.

Using Case Studies Questions	Non-Treatment Group Responses
What were your thoughts on getting data with background information and the data collection methods?	• One student stated that it was easier to interpret the data when you knew what was going on. Many of the students agreed • One student stated that it provided more depth • One student said that they were able to understand what was going on and "why something was the way it was" • One student said that it helped them to understand the purpose of the graph • One student said that the background was excessive and that it wasn't needed
Compare your data literacy skills from before the use of case studies to now.	• All the students said that they had improved • Three students said that the case studies helped them to understand the context of the data • One student stated that background information helped them to look for patterns in data displays • Two students stated that the case studies should have increased in difficulty