# **Data Analysis in Middle School Students: What is the current performance level?**

By

Cody A. N. Marvel

PLAN B PROJECT

Submitted in partial fulfillment of the requirements for the degree of Masters in Science in Natural Science- concentration in Middle Level Math in the College of Education at the University of Wyoming, 2018

Laramie, Wyoming

Masters Committee:

Alan Buss, Chair

Linda Hutchison, Committee Member

Scott D. Crawford, Committee Member

**Abstract**

This paper presents the results of an investigation of 8th grade science students' quantitative literacy skills. A unit assessment from 36 regular education students at Centennial Middle School in Casper, WY was analyzed using the Quantitative Literacy VALUE Rubric designed by Rhodes (2009) and refined by the Association of American Colleges and Universities. Five skills were investigated (Interpretation, Representation, Application & Analysis, Assumption, and Communication) during this study, leading to a maximum score of 20 for the assessment. The results showed that many students do not perform at an adequate level in Analysis, evidenced by the mean of 9.8, the median of 9.75, and the mode of 11.5. A range of 15.5 was also seen, as the lowest total score was 3, while the highest was 18.5. Of the 36 assessments, only 11 scored high enough to reach Natrona County School District's proficiency standard of 70%. The study highlights the need for more data analysis-related assignments and questions (i.e. graphs, charts, tables, figures, etc.), as well as increased expectations in skill-based learning.

## Acknowledgements

Acknowledgements go to Mrs. Sigrid See for providing my scholarship for the program, Natrona County School District #1 and Converse County School District #2 for support throughout the process, all of my colleagues (especially Mr. Shane Palmer) who lent advice or a helping hand, the participants of my study, my parents (Richard and Roberta) for instilling a drive for education early in my life, and my wife (Hillary), daughter (Brooklyn), and son (Nicholas) for their unwavering support, love, and patience throughout the process.

## Contents

# Chapter 1: Introduction

## Purpose

Data analysis skills are some of the most critical skills for students to obtain for a successful future. Eshach and Kukliansky (2016) claim the cultivation of data analysis skills in students not only gives them an intrinsic value, but also helps them bridge the gap between the data they collect in experiments and the theoretical world they are exposed to in the classroom. Compounding this is the fact that standards set forth by various organizations include data analysis skills.

A common set of standards does not exist that is used nationwide for science or mathematics education. However, there are common themes seen throughout standards from different professional organizations. While unpacking the standards from the National Council of Teachers of Mathematics (NCTM), the National Science Education Standards (NSES), the Common Core math standards, the 2016 Wyoming State Standards, and the Next Generation Science Standards (NGSS) several commonalities were discovered. The first theme is creating relevant questions that can be tested and answered using data (NCTM, Wyoming State Standard, and NGSS). Second, the use of appropriate methods to collect data is contained in the NCTM, Wyoming State Standards, and the NGSS. Third, using the data collected to draw meaningful and appropriate conclusions is included in the NCTM, Wyoming State Standards, and the NGSS. Finally, included in all five different sets of standards is using mathematics to find, analyze, and interpret data.

With so many commonalities between the different standards, it seems logical that many of them will be present on state-issued standardized tests. According to the Wyoming Department of Education Wyoming students definitely have room to grow. Looking back on the PAWS scores for the past four school years, the highest percentage of Wyoming 8$^{th}$ graders that were proficient or advanced on the Math PAWS test was 49.69% in 2013-2014. Similarly, the highest percentage of students proficient or advanced on the Science PAWS test over the past four years was 47.07% in 2013-2014 (Wyoming Dep of Ed, 2018). The fact that over half of Wyoming students are not proficient on the state exam is unacceptable.

A comparison of these scores to national scores also raises some large concerns. According to the National Center for Education Statistics (2018) Wyoming is one of the top performing states. This claim is based on the administration of the NAEP (National Assessment of Educational Progress) test nationwide in 2015 and 2017. In 2015, 35% of Wyoming 8$^{th}$ grade students were at least proficient on the NAEP Math test, which was 3% higher than the national average. As for the 2015 Science NAEP test, 38% of Wyoming students were proficient or higher on the test. This was 5% higher than the national average. Identical science scores were seen in 2015, with Wyoming students performing 5% higher than the national average at 38%.

The concerning part about this data is that it appears as though Wyoming students are performing at a high level when compared to the rest of the nation. This is a very misleading claim, however, because 62% of our students are still not performing at grade-level based on the PAWS results. The No Child Left Behind Act was instituted nationally in 2001 and was aimed at closing the academic gaps of 100% of students in the United States. Fast forward to the present day and we see from the data above that has not happened. Furthermore, we seem to be celebrating achievement scores that are significantly lower than the target goal.

The science department of Centennial Junior High School (CJHS) has made several concerted efforts to address these deficiencies. Every science test the students take follows the same generalized format, where much of the focus is placed on skill development rather than recall information. Additionally, each grade level focuses on a different skill development throughout the school year. The focus in 6th grade revolves around variables and all that goes along with them. The students must recognize the different variables (independent, dependent, and control) within an experiment, as well as know how each may be affected by the manipulation of another. In 7th grade the focus shifts to the analysis of graphs, figures, and data tables. The students are expected to be able to choose the appropriate figure for a given problem, as well as create the figure to accurately represent the data. Furthermore, the students are expected to be able to read, analyze, and draw appropriate conclusions from the figures. Finally, in 8th grade the focus is on assumptions and communication. The students are frequently given scenarios that are then manipulated, and they are required to make justifiable assumptions about the scenario and write their assumptions in a science journal. They are expected to be able to see how hypothetical situations can potentially affect the given scenario, as well as use evidence-based claims to support their assumptions.

These focus skills are additional to the state-mandated standards for Wyoming. The students receive instruction on each of the standards throughout their time at CJHS, which is supposed to prepare them for the next step in their education. Judging by the test scores seen above, as well as those discussed in Chapter 2, it is clear that more efforts must be made in order to effectively prepare our students in Wyoming for the future.

**Problem**

I believe the largest deficiency in our students' problem-solving skillset is their inability to analyze data and the corresponding charts/graphs. In teaching math and science for seven years, I have seen a consistent pattern of kids not being able to analyze data sets and/or the different figures often seen in math and science. This is a problem for numerous reasons. First, it shows that we are under-preparing our students for high school and subsequently any post-secondary education. Second, many of the standardized test questions seen on the WY-TOPP test deal with data analysis skills. Finally, we are not giving our students the best opportunity to be successful in life. As professionals, we understand that many facets of life require adequate data analysis skills. Regardless if our students' plan to attend college, a vocational school, or join the work force after high school, it is our job to provide them with the skills necessary to be successful in life. By not focusing on data analysis skills in school, we are doing our students a disservice. The purpose of my investigation is to gather actual data from my students using a pre-existing unit, and use that data to find deficiencies in quantitative analysis skills within my students' current level of knowledge. From there I will propose potential changes to my current curriculum that can help alleviate those deficiencies.

## Research Questions

Therefore, my focus will be on the following question: What is the current performance level of middle school science students at CJHS as measured by the Quantitative Literacy VALUE Rubric?

## Chapter 2: Literature Review

The term 'data analysis' is commonly used among the scientific and mathematical communities, however, it may not always be defined to those that are expected to perform the analysis. Various different activities may be expected while performing data analysis, some of the most common being collecting data relevant to a problem, modeling data with quantitative methods, and interpreting quantitative findings. According to Grolemund and Wickman (2014), data analysts tend to focus less on the properties of a method and more on the connections between the data, the method, it's results, and reality. They believe data analysis can be summarized as the procedural use of data to extract insights and information about reality. Freedman (2009) feels many current data analysis techniques are actually creating a regression in students' abilities to effectively analyze data because they rely on technical sophistication instead of realistic assumptions.

Data analysis, however, is a subgroup within the larger concept of quantitative literacy. Wilkins (2000) defined the term quantitative literacy as,

> an everyday understanding of mathematics. More explicitly, quantitative literacy includes a knowledge of mathematical content embedded in a contextual framework that promotes an understanding and appreciation of the nature, development, and social impact of its applications. Furthermore, it includes a capacity for reasoning and utility and is further supported by a feeling that one is able to function in a quantitative situation. (p. 406)

Quantitative literacy is a concept that seems to have been formalized in the latter part of the 20$^{th}$ Century, with the 1980s serving as a crux for reform within our educational system. Several reports including A Nation at Risk (National Commission on Excellence in Education, 1983), Mathematics Counts (Cockcroft, 1986), and Everybody Counts (NRC, 1989) advocated that students be educated to function within a quantitatively complex society and be able to apply mathematics in everyday life. According to Wilkins (2000), much discussion has been focused around quantitative literacy since this time. This paper, therefore, focuses on the smaller subset of data analysis as defined above.

Assessing the standards discussed in Chapter 1 is extremely important for students, teachers, and parents in understanding how much learning is happening and what the best course of action is for the future. These assessments can come in many different forms, some of which are formal, and others that are very informal. Warm-ups, exit slips, quizzes, standardized tests, and many other forms of assessments exist. Though each of these assessments has value within our educational system, many look to the state standardized tests to evaluate how a state, district, or school is preparing their students.

According to the Wyoming Department of Education (2018), 35.85% CJHS students scored proficient or advanced on the PAWS math test in 2013-2014, while 45.64% of district students were proficient or advanced. In 2014-2015 CJHS students scored 33.19% proficient or advanced, whereas the district scored 41.06%. This trend worsened in 2015-2016, where CJHS students scored a 24.66% proficient or advanced compared to the district's 41.26%. Finally, in 2016-2017 CJHS students scored a 37.7% proficient or advanced, while the district scored a 44.81%. This gap only deepens when comparing CJHS scores to the state scores, with proficiencies of 49.69%, 47.44%, 47.88%, and 48.58% for the respective years listed above.

**Table 1.** Comparison of percent of CJHS 8$^{th}$ Graders, NCSD #1 8$^{th}$ Graders, and Wyoming 8$^{th}$ Graders that are proficient or advanced on Math PAWS

| | 2013-2014 | 2014-2015 | 2015-2016 | 2016-2017 |
|---|---|---|---|---|
| Proficiency Level | State= 49.69%<br>NCSD= 45.64% | State=47.44%<br>NCSD= 41.06% | State= 47.88%<br>NCSD= 41.26% | State= 48.58%<br>NCSD= 44.81% |

| | | | | |
|---|---|---|---|---|
| | CJHS= 35.86% | CJHS= 33.19% | CJHS= 24.66% | CJHS= 37.7% |

A similar (but less drastic) trend is seen when comparing the science PAWS scores of CJHS students to those across the state. For the 2013-2014 year 45.45% of CJHS students were proficient or advanced, compared to the state proficiency of 47.07%. 2014-2015 saw 37.5% of CJHS students score proficient or advanced, while 41.61% of the state students were at least proficient. 2015-2016 was the lowest for CJHS students, with a proficiency of 35.16% compared to the state with a 41.57% proficiency score. Finally, CJHS students scored a proficiency of 43.21% in 2016-2017 compared to the state proficiency score of 45.31%.

**Table 2.** Comparison of percent of CJHS 8$^{th}$ Graders, NCSD #1 8$^{th}$ Graders, and Wyoming 8$^{th}$ Graders that are proficient or advanced on Science PAWS

| | 2013-2014 | 2014-2015 | 2015-2016 | 2016-2017 |
|---|---|---|---|---|
| Proficiency Level | State= 47.07%<br>NCSD= 44.67%<br>CJHS= 45.45% | State=41.61%<br>NCSD= 38.57%<br>CJHS= 37.5% | State= 41.57%<br>NCSD= 40.38%<br>CJHS= 35.16% | State= 45.31%<br>NCSD= 44.85%<br>CJHS= 43.21% |

Munzur (2014) investigated the importance of summative assessments on the motivation and learning of a group of students in an urban English prep school in Istanbul. The driving force for the study came from Boud and Associates (2010), who claim that assessment is an integral piece of a student's education and is recognized as an important factor in student learning. In taking away the summative assessments for a 16-week semester, Munzur was testing how the students would respond in their learning and motivation. The results showed that most students approached the absence of the assessments with uncertainty, and they were uneasy in the amount of learning they felt was occurring. Of the 220 students that participated in the study, 58% reported that they felt more motivated toward school when they knew summative assessments were involved. Also, 59% of the students reported that they preferred to have summative assessments for their learning (Munzur, 2014). The feelings of the students, coupled with the fact that summative assessments provide teachers and parents with numerical values of learning, give validity for using summative assessments as a driving force for classroom instruction.

Critical thinking and data analysis skills are of the utmost importance to the academic success of students today. Much research has been done recently to help identify which skills are most important for students to have, as well as where the largest deficiencies lie within post-secondary students. As for students at the middle level, not much research has been conducted on this subject. It would be useful to investigate the current data analysis performance level of middle-school students is as measured by the Quantitative Literacy VALUE Rubric.

The Quantitative Literacy VALUE Rubric, which was created by Rhodes in 2009, is a great tool for educators to use in assessing students' data analysis skills. After its creation, teams of faculty experts from colleges and universities across the United States validated and refined its contents. To do this, the Association of American Colleges and Universities (AAC&U) sent a survey to various faculty members from colleges and universities throughout the Unites States, asking what they thought were the most important skills needed in post-secondary education. The survey results showed that employers and educators alike feel that many of today's students will need a wide range of quantitative skills to complete their every-day assignments in the work force. The results of the survey also lead the researchers to discover that it was very difficult to assess a student's level of understanding based only on their math assignments, as the assignments did not let the researchers in on the students' thought processes. Therefore, the AAC&U further developed the Quantitative Literacy VALUE Rubric to help educators assess their students' work that addresses quantitative literacy (QL) in a substantive manner (Rhodes, 2009).

The AAC&U define QL as a habit of mind, competency, and comfort in working with numerical data (Rhodes, 2009). It is comprised of six different skills, and each of the skills is evaluated on a four-point scale. The skills included are: Interpretation, Representation, Calculation, Application/Analysis, Assumption, and Communication. The lowest mark a student can receive is the Benchmark (1) level, which consists of the student attempting to explain information found in mathematical data, but coming to incorrect conclusion. The next two levels are found in the Milestones (2 and 3). In level 2 the students may draw some accurate conclusions while others are incorrect. Level 3 sees the students coming to the correct conclusions while using a set of mathematical data, but not going any further. Finally, the Capstone level (4) is given to students who not only come to the correct conclusion, but also provide appropriate inferences for future use based on the information from the data (Rhodes, 2009).

The first skill, interpretation, is defined by the Quantitative Literacy VALUE Rubric as the, "Ability to explain information provided in mathematical form (e.g. equations, graphs, diagrams, tables, words)" (2009, n.p). Bravo et al. published an article in 2010 that highlighted their research in environmental science courses. They were attempting to create a semester-long project aimed at better developing the quantitative and analytical skills of their undergraduate students. The basis for this was developed from more than just their classroom data. Coil, Wenderoth, Cunningham, and Dirks (2010) surveyed 159 faculty members from a wide variety of institutions around the country about the skills they felt were most important for undergraduate students to acquire. 60% of the respondents said that ‘Interpreting Data’ was one of the skills necessary for students to gain before graduation.

Furthermore, the faculty members rated 22 different skills from one (unimportant) to five (very important). Two skills on the list earned a score of 4.9, and one of those skills was ‘Interpreting Data: tables and graphs’ (Coil et al. 2010). Gormally, Brickman, and Lutz (2012) surveyed a different set of faculty members throughout the United States, asking similar questions to those of Coil et al. Though the results from their survey were more generalized than those of Coil et al., they did find that 150 of the faculty members cited skills in inquiry and quantitative analysis as being important to undergraduate students.

According to the Quantitative Literacy VALUE Rubric (2009), representation is the, “Ability to convert relevant information into various mathematical forms (e.g. equations, graphs, diagrams, tables, words)” (2009, n.p.). In the study by Coil et al. (2010) where they asked 159 faculty members from across the country to rate 22 skills from one to five, they found that ‘Creating the appropriate graph from data’ received a score of 4.7. This score was behind only three other skills, all of which are addressed in the Quantitative Literacy VALUE Rubric. Eshach and Kukliansky (2016) feel that graphs are the cornerstone of the laboratory data analysis process, and show the relationships between continuous variables in pictorial form. Furthermore, graphs allow students to see potential trends in data that cannot be easily recognized through a data table alone. Bowen and Roth (2005) claim that scatter-plots, best-fit-functions, and other types of graphs created using coordinate planes are effective for representing the continuous co-variation between two variables that would otherwise be very difficult to express in words.

Gormally et al. 2012 used the information gathered in their survey to create a test specifically designed to gather information about students' scientific literacy skills. They developed, validated, and tested the Test of Scientific Literacy Skills (TOSLS), which was given to students in five general education biology classes at three different undergraduate institutions. They tested nine different skills in the TOSLS, one of which was for the students to create graphical representations of data. This skill was focused on, at least in part, because of work done by Bray Speth et al. in 2010 where they found students have difficulty representing quantitative data on graphs, including the correct labeling of axes and choosing the correct type of graph to use for a given data set.

Gormally et al. then piloted the TOSLS to roughly 1,000 students in various general education biology courses over the course of three semesters. The test was administered to the students at the beginning of each semester and then again at the end. The students were then asked to answer a survey about the TOSLS, which contained questions about the difficulty of each of the skills (both pre- and post-test). The data gathered from the pilots was then analyzed in an effort to identify any trends.

One trend found was the students gave the data representation questions a rating of 0.4 (pre-test) and 0.5 (post-test). On the scale used, which was developed by Feldt in 1993, higher numbers represent easier questions, while lower numbers represent more difficult questions. Test items ranging from 0.3 to 0.8 are considered to be acceptable questions to gain accurate representations of student understanding. Of the 28 questions contained in the TOSLS, only three other questions received a lower post-test score. This validates the thought that data representation is viewed as one of the more difficult quantitative literacy skills to students.

Not only is this an integral data analysis skill, but it is also contained in the 8th grade Common Core Math Standards. According to the National Governors Association Center for Best Practices in May 2018 (NGACBP), standard CCSS.MATH.CONTENT.8.SP.A.1 states that 8th grade students should be able to accurately create bivariate scatter plots and interpret them to see patterns between two sets of data. The Representation skill was also assessed using questions 29-30 on the unit assessment. These questions required the students to draw a line of best fit that accurately showed the population growth of each species over time. This skill is also part of the Common Core Math Standards, specifically standard CCSS.MATH.CONTENT.8.SP.A.2. According to the NGACBP, 8th grade students should be able to fit a straight line to two quantitative variables and assess the model fit by judging the closeness of the data points to the line.

Another skill focused on in data analysis is 'Calculation'. This skill is fairly self-explanatory, as there is no description given in the rubric. Students simply must be able to compute numerical data given to them in a variety of different forms. Gormally et al. (2012) included calculation in the TOSLS, showing agreement with the authors of the VALUE rubric. They felt students must be able to solve problems using quantitative skills, including probability and statistics. Three of the questions on the TOSLS are dedicated specifically to evaluating students' calculation skills. In post-test interviews with the students, they found that many students had low self-efficacy or they guessed on this question. Given that this is happening at the post-secondary level, it is reasonable to assume the same is happening in the middle and high schools.

Blaylock and Kopf (2012) focused on the relationship between the success on higher order questions and the ability to solve computational problems by hand. They designed a study that tested higher education students' abilities to select an appropriate quantitative tool, create a quantitative model, and interpret the results of a quantitative analysis based on their pre-existing computational skills. Their results suggest that there is a significant relationship between a student's ability to both select an appropriate quantitative tool and to create a quantitative model with their computational skills in mathematics. They did not find, however, a relationship between a student's computational skills and his/her ability to interpret the results of a quantitative analysis. This study supports, at least at the correlation level, the belief that a student's ability to calculate math problems by hand does affect his/her performance on advanced work that requires higher-order quantitative reasoning skills.

Application/Analysis is another skill found in the Quantitative Literacy VALUE Rubric. This skill is the ability to make judgments and draw appropriate conclusions based on the quantitative analysis of data, while recognizing the limits of this analysis. Gormally et al. (2012) included application and analysis in the TOSLS because they feel that students must be able to, justify inferences, predictions, and conclusions based on quantitative data. Three of the questions within the TOSLS are dedicated to these skills. Two of the three questions include hypothetical scenarios, while the third question used actual biological data. They found the students performed well on the hypothetical questions (item difficulty scores of ~0.65 and 0.62), but when given an actual graph with biological data, they struggled to come up with the correct response (item difficulty score of ~0.3).

Goldstein and Flynn (2011) further support the findings of Gormally et al. The authors taught a revised version of their laboratory class to 141 students; however, they changed six of the 11 labs to include multiple data analysis opportunities with actual biological data. In the pre-test they found that a median of 1 of 4 students responded correctly to the data analysis questions. Conversely, on the post-test they found that a median of 3 of 4 students responded to the data analysis questions correctly. They also found that there was no correlation between students that had prior preparation in statistics and those with no experience. Therefore, students who had previously learned basic quantitative analysis skills had trouble applying their knowledge to actual biological data sets.

Making assumptions is another skill focused on in the rubric and is the ability to make and evaluate important assumptions in estimating, modeling, and data analysis. This skill can be applied in many different ways. One example comes from Gormally et al. (2012) in the three application/analysis questions seen in the TOSLS. Each of these three questions requires the students to make judgments and draw conclusions based on data, but it also requires them to take those judgments and make assumptions based on their analysis. These assumptions often ask the student to take their conclusions and apply them to a future situation. Much research has been done regarding students' abilities to find trends in data and apply them to a future situation.

Kilpatrick and Silver (2000) believe that data analysis is a crucial piece to developing new theories and ideas, and that without these skills students become stagnant. The students must pay close attention to patterns, find relationships between or within data sets, postulate why outliers exist within data sets, and determine the external factors that may affect data sets. In doing this, the students will develop a much deeper understanding of the difference between evidence and theory, which is often lost among students. In studying data and statistics, students can also learn that solutions to some problems depend on assumptions and have some degree of uncertainty. The kind of reasoning used in probability and statistics is not always intuitive, and so students will not necessarily develop it if it is not included in the curriculum.

The final skill in the Quantitative Literacy VALUE Rubric is communication, which is the expression of quantitative evidence in support of the argument or purpose of the work. Coil et al. share the importance of this skill in their 2012 study. They included two different forms of communication in their survey of 159 faculty members, one being written and the other being oral. The results of the survey gave written communication a score of 4.7 out of a possible 5, while oral communication received a score of 4.6.

Another interesting aspect of communication comes from a presentation given by Ledbetter in 2011 and published in 2012. She views communication as one of the important competencies, which are skills that can be developed as a means of practicing math and science. These skills can eventually develop a much more in-depth way of looking at the world using mathematics and science. She reports that students need to be able to communicate their content and results in a manner that allows them to collaborate with other disciplines. They need to recognize that no one individual can master the knowledge necessary in all areas to address complex questions, but rather they will need to collaborate with others. Furthermore, they need to recognize that effective collaboration skills require interpersonal skills, particularly in oral and written communication.

# Chapter 3: Methods

## Introduction

I am currently in my eighth year of teaching, and this entire study was spurred by my overwhelming feeling that the average middle level learner cannot effectively analyze data. It has been my experience that when faced with problems dealing with quantitative analysis, interpreting or creating graphs/charts, or creating data tables, kids struggle. For this reason I wanted to design a unit centered around the development of these skills, with a pre- and posttest acting as the data for my results. However, it was later decided that I should instead investigate if my feelings about my students' abilities had any substance to them, or if these feelings were nothing more than a hunch. Therefore, I decided to teach a pre-existing unit as I always have, followed by a detailed analysis of the assessment using the Quantitative Literacy VALUE Rubric as the driving force. The goal of the study was to evaluate the current performance level of middle school science students, pertaining specifically to data analysis.

## Context

The study was conducted at Centennial Junior High School (CJHS) in Casper, Wyoming. CJHS is a school of approximately 750 students in grades 6-8, and is located in a rather affluent area in Casper. This can be a bit misleading, however, as the school district (Natrona County School District #1) is a school-of-choice district. This means that there are no physical or imaginary boundaries for school assignments. Students and parents are allowed to attend any school they prefer as long as they correctly follow the application process. Due to this, the student population of CJHS exceeds 50% free-and-reduced lunch according to the Teacher Cancellation Low Income Directory (2018). There are about 250 8$^{th}$ graders at CJHS, and all of them are required to take life science during their 8$^{th}$ grade year. Personally, I teach one advanced class of life science and five regular classes, with each of them ranging anywhere between 20 and 26 students.

Throughout the school year my teaching partner and I teach eight major units. We also have several short units that are review, such as the scientific method, the metric system, and laboratory safety. We do these small units the first three weeks of the school year. After they are completed we get into the units that are based on the Wyoming State Standards. We begin with classification, which revolves around the seven levels scientists use to classify all living organisms. After this we lead into evolution, where the students learn about Charles Darwin, natural selection, and how adaptations and mutations lead to the evolution of a species. Next is cell structure and function, where the students learn about the different organelles within a plant or animal cell and what each of them does. Upon completion of this unit we move into cell processes. This unit contains many of the different processes our cells use to complete their functions, such as osmosis, diffusion, endo- and exocytosis, etcetera. Next is our unit on cell reproduction, where we explain the processes of mitosis and meiosis. This leads into the genetics unit, where we explain the different concepts that lead to heredity. This unit also contains information on how to accurately predict the outcome of various crosses using ideas discovered by past geneticists. Finally, we get into the human body systems in our last unit. Within this unit we investigate the skeletal, muscular, cardiovascular, respiratory, and circulatory systems. Throughout each of these units the students are required to gather and analyze data, but the evolution unit has them do this more substantially.

I chose to use my evolution unit for this study. I did so because my teaching partner and I recently revised the unit, complete with a summative assessment that contains quantitative analysis questions. The unit is designed to last five to six 80-minute blocks, or roughly 12 45-

minute classes. It contains a mixture of several different types of instructional strategies, but none of them directly teach data analysis skills. I chose to continue teaching it this way to get a more accurate idea of how well my students can currently analyze data. I also chose this unit because it falls at the halfway point of the school year. This would allow me to make necessary changes to my instructional practices based on the results from the assessment.

**Participants**

All of the participants are 8th grade life science students. At the time of the study they were all 13 or 14 years of age, and had consent to participate from their parent/guardian. Those students who were 14 at the time of the study were also required to give their assent to participate in the study as required by the Institutional Review Board (IRB). Thirty-six total students participated in the study, and they came from five different regular classes I teach. I chose to do this because I wanted to obtain the most accurate data possible for the data analysis skills of an average 8th grade student. Also excluded were any students with an individualized education plan (IEP). I chose to exclude them from the study because many of them are required to be in small group testing, and not every classroom aid provides the correct level of guidance to satisfy the IEP. Therefore, I could not guarantee that the tests were entirely the students' own work.

The students that chose to participate in the study did not receive any benefit or detriment for their participation. All of the students were expected to participate in the unit as if there was not a study being conducted. If a student decided to play a part in the study he/she was simply expected to participate and perform as normal as possible. The students also had the option to remove themselves from the study at any time during the unit, and no reasoning was required for the dismissal of their consent. The only remaining artifacts from the study are the numbered tests, consent forms, and assent forms. All of this will remain locked in a file cabinet for three years from the completion of the study.

## Procedure

### Evolution Unit.

The design of the unit lends itself to an almost seamless sequence of progression. It begins with a brief discussion as to what the students feel evolution actually is. This leads into the development of the major vocabulary terms that will be used throughout the unit, such as natural selection, Charles Darwin, adaptation vs. mutation, etc. Once the foundation is set with the vocabulary we move into a natural selection simulation, complete with several data analysis questions that inspired the assessment questions. Following the natural selection simulation we move into a quick formative assessment worksheet to see where the students are performing. This then leads into an activity that has the students give an existing organism some form of adaptation, complete with a description of what the possible impacts would be from the adaptation. Next, we complete a unit review with all of the major components included. Finally, the students take the unit assessment.

### Instrument.

Two different instruments were used in this study to evaluate the students’ data analysis skills. First, the Evolution Assessment was used due to its design and ability to be dissected for individual data analysis skills. Second, the Quantitative Literacy VALUE Rubric was used for several reasons. Not only does the rubric break down each skill specifically and give adequate reasoning for each of the levels, but it is also peer-reviewed and accredited.

***Evolution Assessment.***

Item 18 of the Evolution Assessment required the students to take the data seen in **Figure 1** and create two different scatter plots.  Two seven-by-seven grids were provided for the creation of the scatter plots, which can be seen in **Figure 2**.  These questions were used to evaluate the representation skill from the Quantitative Literacy VALUE Rubric.

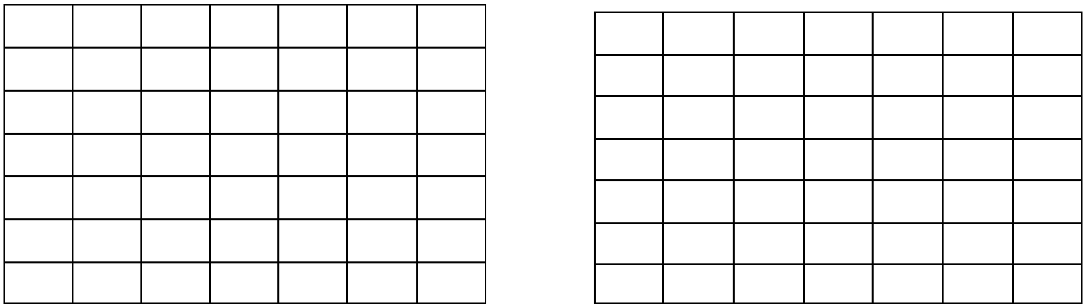
**Figure 1.**  Excerpt from Item 18 of the Evolution Assessment

Elk vs. Wolf Population in Northwestern Wyoming

| Year | # Elk | # Wolves |
|---|---|---|
| 2000 | 1,536 | 12 |
| 2004 | 1,782 | 25 |
| 2008 | 1,698 | 33 |
| 2012 | 1,701 | 75 |
| 2016 | 1,649 | 49 |
| 2020 | 1,996 | 55 |

**Figure x**. *Item 18*

*Item 18*- Create a scatter plot for both the elk and wolf populations.



Item 19 asked the students to manually draw a line-of-best-fit for each of the scatter plots. They were given a ruler to complete this if they chose to use it. These two questions were used to evaluate the representation skill, as well.

Item 20 required the students to use the lines-of-best-fit they had drawn in Item 19 to predict future populations. They were specifically asked to predict the populations of both the wolves and elk in the year 2030. This item was used to evaluate the application/analysis skill and the communication skill.

Item 21 had the students describe any trends they saw in the elk and wolf populations between 2000 and 2020. They were not prompted to use the data or the scatters plots exclusively, nor were they prompted to use a combination of the two. Rather, they were required to use previous knowledge to find relationships in the data. These questions were used to evaluate the interpretation skill.

Item 22 gave a fictional scenario in which the number of elk licenses issued in the area being studied were doubled for two years. Given this new information, the students were asked to give a logical prediction of the effects this would have on the wolf population. These questions were used to evaluate the assumptions skill.

Item 23 followed suit with questions 35-37, where another fictional scenario was given to the students. This time a wolf-hunting season was opened in the studied area, with 50 licenses hypothetically being issued for four consecutive years. The students were required to give a prediction for the effects of this hunting season on both the wolf and elk populations. These questions were also used to assess the assumptions skill.

After completion of the test, I pulled tests from students that had returned consent forms signed by their parents and/or signed the assent form. After compiling all of the qualified tests I made copies of each of them, followed by the return of the original tests to the students. I then used whiteout to cover the names and any other identifying marks on the tests and assigned each test a number. This was followed by the creation of a second set of copies, this time with only the number assigned to each test at the top. This was the set of tests that I evaluated using the QL VALUE Rubric, so that I did not have any indication of whom the test belonged to or what numerical grade was assigned to the test.

***<u>Quantitative Literacy VALUE Rubric.</u>***

Using the QL VALUE Rubric, the evolution assessment was analyzed and each of the items in the data analysis section was assigned to the appropriate skill listed in the QL VALUE Rubric. With the help of Dr. Alan Buss (the faculty advisor of this study) it was determined that the following items from the evolution test represented the associated skill on the rubric: items 18 and 19= Representation; item 20= Application/Analysis and Communication; item 21= Interpretation; and items 22 and 23= Assumptions.

Items 18 and 19 were chosen to evaluate the representation skill due to their direct correlation to data representation. By requiring the students to take a given data set and create figures from it, they are essentially providing evidence of their skill level.

While assessing the representation skill using the rubric, certain items were necessary in order to receive a score of four. First, the graphs needed to be scatter plots and not something different (like a line graph or bar chart). Second, the intervals on the y-axes needed to be appropriate for the particular graph. If the student chose to skip the bottom of the graph on the elk example he/she needed to correctly use the scale marker to signify that some of the graph had been omitted. Third, the years had to be accurately defined on the x-axes, complete with a continuation through 2030 to accurately estimate the populations for that year. Fourth, an accurate line of best fit was necessary for the students to correctly estimate the populations in 2030. Lastly, the graphs had to include the usual components seen in a complete graph, such as titles, labels, and accurate data representation. If each of these items were present the student received a four, and for every missing piece the score went down.

Item 20 of the unit assessment provided an opportunity to evaluate the application/analysis skill as well as the communication skill. Both of these skills could be assessed using this item due to the fact that the students must report quantities based on the analysis of a figure. The questions had the students predict the population of each species in 2030 based on the line of best fit from item 19. This prediction relies on the students being able to correctly apply the line of best fit to a future situation. They must also make thoughtful judgments and draw qualified conclusions from the carefully placed line of best fit. In providing potential populations for 2030, the students are also providing quantitative evidence to support the purpose of the work. In this case the purpose is to predict how a scenario will affect the wolf and elk populations in an area over time. The quantities given by the students show their understanding of the skills through their application of the data and the communication of their conclusions.

The assessment of each of these skills relied on the students' ability to correctly create the figures in items 18 and 19. Both of the skills were graded based on the numerical values the students gave for the populations in 2030 and the evidence used to support those values. The accepted values for each of the populations had to be assigned a range as they were based on the intervals used for the y-axes and the slope of the line drawn in the data. Therefore, answers between 2,000 and 2,300 were accepted for the elk population, while the range for the wolf population was from 85 to 100. A score of four was assigned to any answer that not only had a numerical value within the accepted range, but also had the x-axis correctly labeled through 2030, had an interval on the y-axis that continued high enough to assign a value, and had a point on the line of best fit that represented the population for that given year. Answers that were missing pieces of evidence, had incorrect values, were guesses based on no real evidence, or were not answered with a population value were assigned less points based on the severity of the missing items.

The interpretation skill was assessed using item 21. This item was chosen for the interpretation skill because it required the students to find trends in data presented in mathematical form. The students were expected to find trends in the given data using the data table, the graphs they created, or a combination of both. While looking through the data the students were expected to see that both populations grew over time and had an overall positive correlation. They were also expected to realize that the two populations were interconnected. For instance, a significant increase in the wolf population would likely lead to a decrease in the elk population over the following four-year span. Conversely, a large spike in the elk population would likely lead to an increase in the wolf population over the next four years as they would have more food to hunt. Students that were able to note the overall positive trends, see that the two populations were interconnected, and give evidence from the data to support their claims were awarded a score of four. Again, lower scores were awarded to the answers that were missing pieces or lacked evidence.

The final skill assessed using the rubric was assumptions, and for this items 22 and 23 were used. These items were used for the assumptions skill because they both required the students to make meaningful predictions and assumptions about data sets based on new variables compared to the original data set. Both of these items gave a different hypothetical scenario, and the students were required to analyze the scenario using information they had already gathered. They were then expected to make an assumption about the potential effects on the populations.

In item 22, the number of elk licenses issued in the area being studied was hypothetically doubled for a two-year period. The students were tasked with taking that information and formulating a hypothesis about the effects that would have on the wolf population. A complete answer included the assumptions that the elk population would likely decrease due to more human hunting, the wolf population would likely see a slight dip due to less food availability, and both populations would likely rebound from their changes after several years. This would result in a score of four, with lower scores being issued to incomplete answers.

As for item 23, the students were given the scenario that a new wolf-hunting season would be introduced, and 50 licenses would be issued for a four-year period. They were expected to assume that the wolf population would see a decrease due to the human hunting, but also understand that not all of the wolves would be killed and survivors would reproduce. This idea was based directly on the Natural Selection Lab that we did in class during the unit. The students were also expected to understand that the elk population would likely increase due to less predation, but that both populations would find a balance of nature after a period of time and level out. Otherwise, it would be logical to assume that the trends they propose would continue year-after-year. This, too, was derived from the Natural Selection Lab performed during the unit. A score of four was issued to an answer containing all of the aspects above.

The assumptions for items 22 and 23 are based on the ideas presented on the Natural Selection Lab. The main concepts were that increased predation would result in more deaths, but that the stronger (or more well-adapted) individuals would survive. These individuals would then reproduce the following year, creating a population better suited to survive future predation. The overall population, however, would be smaller for a short time frame due to the increased predation. Eventually, this would lead to a bounce-back effect for the population being hunted, as the individuals being born each year would have the genetics better suited for survival. This would positively affect the predator species, as they would have more prey to hunt.

**Data Analysis**

The final piece of the study came in the form of organizing and analyzing the collected data. Not only did I go through all 36 tests and assign each skill a score, but my teaching partner also did the same. He took the numbered copies of the tests, and using the guidelines listed above for each of the skills, independently assigned each of the skills a score. After completing all of the tests, the two of us went through each of the scores for each test and compared our results. If there were any discrepancies between the assigned scores we would look at the specific test and come to a consensus on what score should be final. This step was very important as it ensured that the study had inter-rater reliability. The goal of inter-rater reliability is to increase the validity of the results by decreasing the effect any biases may have had on them. After a consensus decision was made for each of the tests, the results were finalized and further assessed for Chapter 4.

# Chapter 4: Results

## Scoring

Data analysis is undoubtedly one of the most important skills necessary to adequately prepare students for the future, yet it is a skill that is often sacrificed for the completion of content standards. To assess this standpoint, 36 unit tests from my 8$^{th}$ grade science classes were analyzed using a peer-reviewed rubric designed to assess data analysis skills. The results, as a whole, were not outstanding, nor were they surprising.

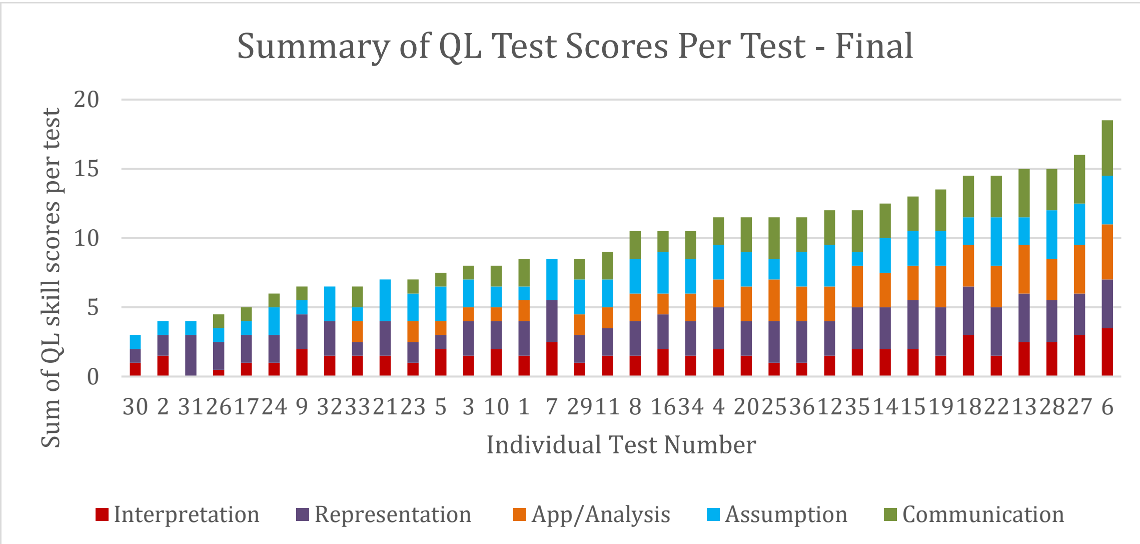
Using the four-point scale determined by the designers of the rubric as well as the unit assessment designed by my teaching partner and myself, I was able to assess five of the six skills listed in the QL Rubric. The only skill that I was not able to assess was 'Calculation', which unfortunately was left out due to the lack of any actual calculations on the assessment. This meant that each student had the potential to score up to 20 for a rubric score. After both my teaching partner and myself analyzed each of the tests and subsequently came to an agreement on the scoring of each test, I can say that none of the tests received a perfect score. **Figure 3** shows the total score for each of the tests based on the above 20-point scale.

The highest score achieved on all three sets of data came from test number six, with a final score of 18.5 being given to that student. Conversely, the lowest score of three was seen on test 30. A fairly wide range of scores was seen between the two extremes.

There were 21 tests that scored above nine, while the remaining 15 scored below nine. The mean test score was 9.8, the median test score was 9.75, and the mode test score was 11.5. All of this put together tells me that the validity of the test was proficient, as there was not a large discrepancy between any two of the values (mean, median, and mode). The range, however, was quite large. Given the highest test score was 18.5 and the lowest was 3, a range of

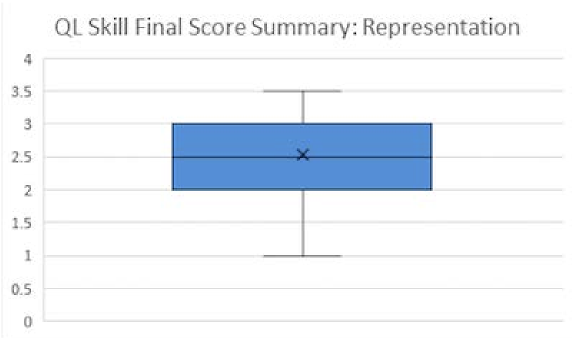
15.5 was assigned to the data. This, coupled with the fact that 3 was not an outlier due to the slow progression of scores, shows that the students did actually perform at all different levels on the test.

**Figure 3**- Final consensus test scores broken down by individual skills.



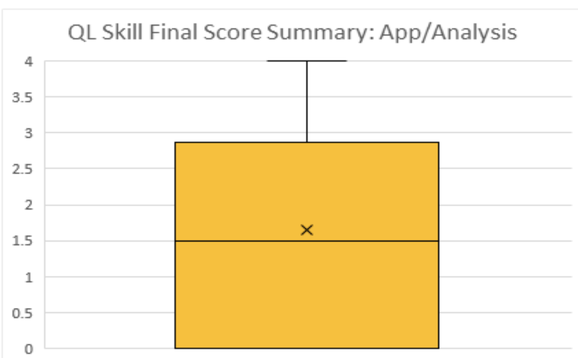

Based on the four-point scale from the QL VALUE Rubric, each individual skill was able to be analysed. **Figure 4** gives a great visual for the representation skill. The figure shows that the range for 'Representation' was from 1 to 3.5, with 50% of the total scores falling in quartiles two and three. This, coincidentally, fell between the rubric scores of 2 and 3. 25% of the scores were contained in the first quartile between the scores of 1 and 2, while the final 25% of the scores belong to the fourth quartile between the scores of 3 and 3.5. It also shows that there were no outliers for this particular skill.

**Figure 4.** Box and whisker plot of the student scores for the representation skill



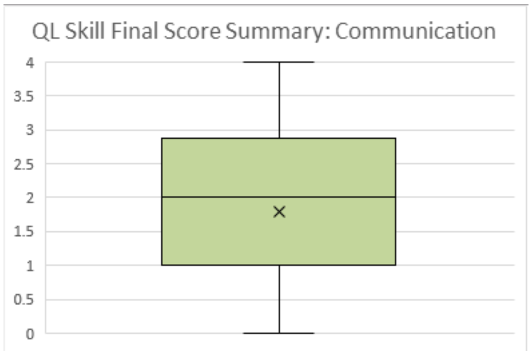

The application/analysis skill was the next to be analyzed individually. **Figure 5** shows the range for this skill reaching all the way across the spectrum, going from 0 to 4. 75% of the scores fell within the yellow portion of the graph, ranging from 0 to 2.85, while the remaining 25% of the scores fell into the fourth quartile. This ranged from 2.85 to 4. Given that the range spanned the entire spectrum and the box is so wide, there were no outliers for this skill.

**Figure 5.** Box and whisker plot of the student scores for the application/analysis skill



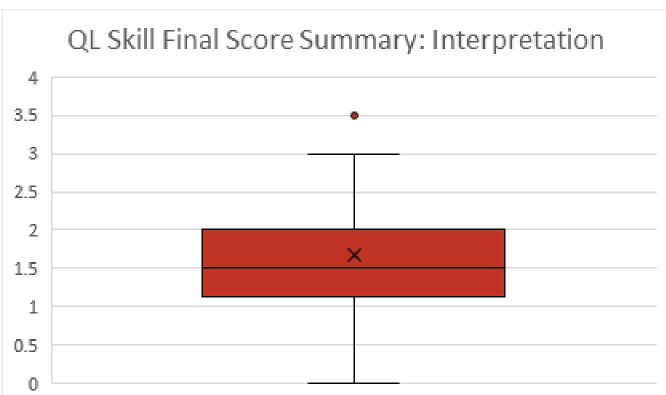

The communication scores were also analyzed individually, and **Figure 6** shows the results from that analysis. Quartile one ranged from zero to one, and it contains 25% of the scores. 50% of the scores were contained in quartiles two and three, and they ranged from scores of 1 to 2.85. Finally, quartile four contains the the remaining 25% of the data for the communication skill and ranges from 2.85 to 4. Given that the data spanned the entire range of possible scores no outliers exist.

**Figure 6**. Box and whisker plot of the student scores for the communication skill




Interpretation was the next skill analyzed**. Figure 7** gives a visual representation of the scores for this skill. 25% of the scores fell into quartile one, which ranges from zero to 1.15. Quartiles two and three contain 50% of the data for this skill. These quartiles range from 1.15 to 2. Quartile four contains almost all of the remaining 25% of the data points for interpretation, though an outlier does exist at 3.5. The data in quartile four ranges from 2 to 3.

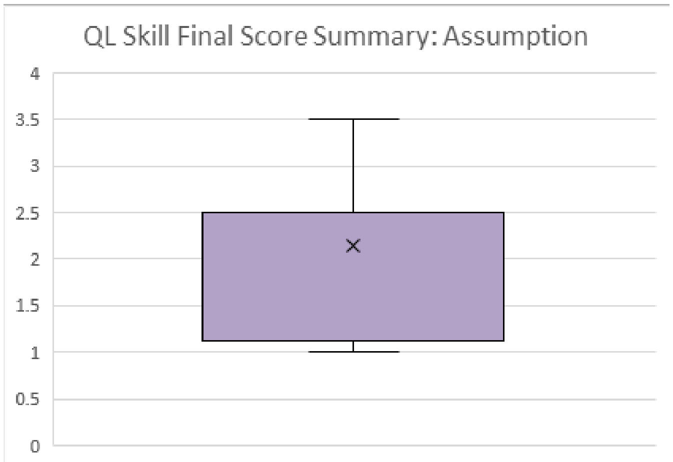
**Figure 7**. Box and whisker plot of the student scores for the interpretation skill




The final skill to be analyzed individually was assumption, which can be seen in **Figure 8.** Quartile one ranges from scores of 1 to 1.15, and it contains 25% of the scores for this skill. 50% of the scores fall into quartiles two and three, and they range from 1.15 to 2.5. The final

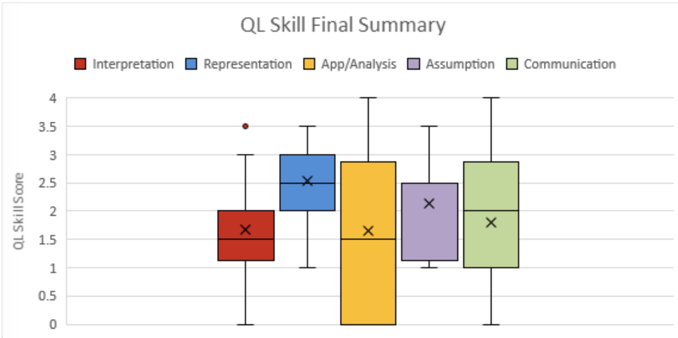
25% of the scores fall into the fourth quartile, and the scores range from 2.5 to 3.5. There are no outliers associated with the assumption skill.

**Figure 8.** Box and whisker plot of the student scores for the assumption skill




Some interesting trends were also seen when analyzing the averages for each of the skills. Representation was the highest rated skill across all three data sets, with the final average being 2.5 out of 4. The second-highest rated skill across all three data sets was assumptions, with an average of 2.1 out of 4. Communication was the third highest rated skill, with an average of 1.8 out of 4 on all three data sets. From there, however, the remaining two skills fluctuated in their average across the different data sets. In the end both interpretation and application/analysis had averages of 1.7 out of 4, giving them the distinction of having the lowest averages of the five skills assessed. **Figure 9** shows all of these trends in a compartive manner.

**Figure 9.** Box-and-whisker plots for all five of the skills analyzed using the QL VALUE Rubric.



## Chapter 5: Discussion

At the onset of this project I set out to get an accurate depiction of my students' abilities to effectively analyze data. I chose to investigate this after teaching in a public middle school for seven years and continually feeling as though my students were not leaving my class adequately equipped for future experiences. Due to this hunch, I developed a plan to collect data using the kids currently in my classes to provide evidence that would either support or deny my claim. After the completion of the study and subsequent analysis of the data collected, I would say that my hunch was in fact correct.

I found the averages for each of the skills valuable for several reasons. First, the fact that only two out of the five skills received a score over two is surprising. I hypothesized that my students were lacking in the data analysis skills, but to have three skills below the basic level of performance was not something I expected. Second, the Application/Analysis score of 1.7 was surprising given that it was a generalized line-of-best-fit, without any actual calculation of slope or exact fit being needed. Many of the kids were able to construct decent graphs, but struggled with fitting a line to the data. Finally, the Interpretation score of 1.7 was surprising due to the nature of the question. We had completed several questions in class that were very similar in that the students were expected to analyze a data set and describe the trends. During the unit the students performed well on these types of questions, so I expected them to do the same on the assessment. The low average, coupled with the fact that only one student received a zero for this particular skill, tells me that the students performed poorly on this skill across the board. This is in contrast to Application/Analysis, which shared the average of 1.7 but had ten students receive a score of zero.

While analyzing the total test scores on the final data set I also found some intriguing data. Given that the averages for each of the skills were relatively low, I expected the overall test scores to be low as well. In fact, I thought there would be more test scores in the bottom half than there would be in the top half. I found that to be true, as 19 of the 36 tests had an overall score of 10 or lower. This leaves 17 tests with an overall score above half of the total available points from the rubric.

The district I work in (Natrona County School District #1) has a general rule that students must at least meet a 70% proficiency on a specific standard to be able to move on to the next standard. For this study that means the students would be required to score a 14 out of 20 or better to be considered proficient on the standard being addressed. Based on the data collected in this study, only 11 of the 36 tests analyzed would meet that criteria. The remaining 25 students would go through a series of re-teaching assignments and formative assessments, followed by a different form of the summative assessment.

To go along with the low overall scores, the average from the 36 tests collected was 9.8 out of a possible 20. Given that the overall average was below a 50%, I chose to accept my hypothesis. Along with this same idea, 3 of the 5 individual skills examined had averages below 50% (Interpretation, Application/Analysis, and Communication), further solidifying the conclusion that my students do not possess adequate data analysis skills.

The acceptance of my hypothesis does not, however, come without the realization that limitations did exist within the study. This study has a relatively low sample size, both in the number of students tested (36) and the number of questions evaluated (5). In fact, I was not even able to analyze all of the skills on the QL VALUE Rubric due to the low number of questions used. It would be a much more valid study if each of the students was evaluated over several different tests using the QL Rubric. This would give them multiple opportunities to showcase their skills, as well as give me more data to analyze and base assumptions upon. However, I do feel that the five questions used in this study give a decent indication of the students' data analysis skills in a general aspect.

The five questions used in this study provided me with the ability to analyze certain aspects of my students' abilities, while also leaving some major gaps. I was able to obtain a generalized idea of each student's overall data analysis abilities by looking at each individual score combined with the total score for each student. Lower overall scores were generally associated with low individual skill scores, but that was not always the case. For example, test number 7 had an overall score of 9.7, which was just shy of the average of 9.8. This would make it seem like that student was struggling across the board, yet when further analyzing it I discovered that only three of the five skills were actually represented. This student may have not understood test item used to evaluate the remaining two skills, or he/she simply could have overlooked the question entirely. Had the other two skills been represented on this test the score would have likely gone from below average to meeting the district standard.

I was also able to get an idea of each student's thought process in addressing each of the skills by analyzing the answers to each of the short response items. If a student provided a logical answer to a specific item and included all of the pieces described in Chapter 4, I assume he/she has a pretty solid grasp of the skill. If a logical answer was provided but one piece was missing, I assume the foundation for the skill is there but more refinement is needed. If multiple pieces were missing or a completely illogical answer was provided I assume the student has significant deficiencies in the skill and requires remediation. Finally, if an answer was given that I did not anticipate but showed logical reasoning, I assume the students may not necessarily have a grasp of the skill, but he/she does possess the ability to make sense of something new or foreign to him/her.

The difficult part of only using one test with five questions is the extremely small amount of opportunity the students had to show their skills. For example, I required the students to prove their representation skills by creating two scatterplots. However, they were not given any other opportunities to showcase that skill using any other form of graph. This is important because a student that received a score of one or two on this skill due to the scatterplot requirement may have received a three or four if given another type of graph to create. By using multiple tests and a variety of different requirements a more data set could be compiled and analyzed.

As with most scientific investigations, mistakes and points-for-improvement were discovered through the implementation of the study. The first item that I would change came in the creation of the scatter plots. I provided the students with a grid for each of the scatter plots, but did not give them enough room on the grid provided to make a prediction in 2030. This is something that I did purposely with the intention of making them think critically and extending the graph for the situation, however, after grading the tests I think it was more misleading than helpful. If I administer the same test in the future I will either extend the grid far enough for the students to label to 2030 or take the grid out entirely.

Second, I found that many students failed to give a long-term outlook for questions 35-37 and 38-40, despite the fact that every question we had answered throughout the unit that was similar to them required one, as well as me giving them verbal instructions to address the long-term effect for the populations. This is a large part of the conundrum I am facing as a teacher, where an extreme lack of attention inhibits many students' ability to be successful. For future administration I could revise the questions to include a prompt that gives the students an immediate reminder to address that part of the questions. I would do this begrudgingly, however, because I feel as though this further enables them to be simplistic thinkers without motivation to increase their skillsets.

The best course of action would be to better prepare them for questions of this sort throughout the unit so the analysis skills are engrained in them. I believe this would be much easier said than done, though I also believe that it is possible. To start I can create more warm-ups, assignments, quizzes, exit slips, etcetera that contain questions requiring this type of thinking. This will give them numerous opportunities to read, hear, and answer questions dealing with potential trends to a given data set. Beyond this, and possibly more importantly, I will discuss with my future students the importance of data analysis skills and share with them my knowledge gained from this study. My hope is that using real data from a sample set that is meaningful to them (e.g. kids from their actual school district) they will take more from it and find the importance in building these skills.

The final limitation I found in the study came in my analysis of the data. I vastly underestimated the amount of time and energy it would take to effectively assign scores to each of the skills. This led to me sitting down several different times to assess the tests, which I feel decreased the validity of my scores. I say this because I was not in the same mindset each time I sat down to grade the tests, and that could have easily affected the grades I was assigning to each of the skills. It would have been better if I could have sat down and assigned the scores in one or two sittings, but life and time did not allow for this to happen. I do find some solace in the inter-rater portion of the study, though. Many of the potential mistakes were discovered through the conversations between Mr. Palmer and myself, which in turn increased the validity of the overall study. This is not something that I can change for future administrations of the test, but I felt it was important to address for future studies.

In the event that changes could be made to the study for future administration, I would address more than the limitations discussed above. First, I would include students from all of my classes, including my advanced class. This would likely change the overall average for the study, as well as give a more realistic representation of all students' abilities. I feel as though if I had performed this study with only my advanced class the average would likely be above the district standard (70%), and the overwhelming majority of the scores would show my students have amazing data analysis skills. Obviously this would not be an accurate representation, but I would like to see how the scores changed if advanced students were included. I would also change the items used to evaluate the Representation skill to include a menu of sorts. This means that I would allow the students to create a graph of their choice to represent the data, rather than forcing them to make a predetermined graph. Finally, I would adjust the sample size taken to include more students. Dr. Buss and I agreed that 35-40 students was an adequate number to get a decent representation of the students' skills, but a larger sample size would only increase the validity of the data collected.

Aside from accepting my hypothesis and finding limitations within the study, I also found some interesting trends in the data that could affect future students in my classroom. While assessing the data I realized that I could break the final scores of each test into three categories: 1.) Remediation, 2.) Grade-level Support, and 3.) Enrichment. The remediation group would consist of any students who tested in the bottom third of the sample size, or those that failed to address two or more of the skills being assessed. This group would likely need complete re-teaching using a different technique than had been used during the unit. Given that this re-teaching will likely have to occur during regularly scheduled class periods, the remediation group will likely be fairly small. In lowering the student-to-teacher ratio for the re-teaching I will be able to differentiate the instruction for each student much better than I could during the initial unit. Also, I will be able to spend more time with each student in this group than I was able to during the unit.

As for the grade-level support group, they will be tasked with completing an assignment dealing with the unit that reinforces the data analysis skills I am hoping they acquire. This assignment would likely be some sort of lab or activity that is both engaging and educational. It will include some form of a data collection piece, as well as various different questions or scenarios that provide opportunities for the students to practice and refine their data analysis skills. The students in this group will be those that performed in the middle third of the sample size. This usually equates to the students that are performing at the grade-level expectations, though it is not an exact science and may need to be more precisely examined for each specific assessment.

The third and final group will be focusing on an enrichment of the unit that has just been assessed. This group will be comprised of the students that performed in the top third of the sample size, which theoretically constitutes scores that were above the grade-level expectation. The enrichment this group receives could be designed a wide variety of ways, but it must stay within the time frame given which will most likely be one 80 minute class period. This is important because the activity must be a continuance of the learning from the unit while also taking the students' thinking to a higher level. Activities that adhere to these guidelines are often designed to apply concepts learned in class in a practical manner. Though these activities do not tend to be difficult to design, they may be difficult to complete in an 80-minute block. To address this I may have this group complete the same lab/activity that the grade-level group does, but the follow-up questions will be tailored for higher-level thinking and application. This will be a great opportunity to create questions that really test the students' abilities to analyze data and make predictions using the data without being potentially detrimental to their grade.

The implementation of this plan may also be a touch tricky given physical restraints in a classroom. It would be very difficult to administer each of the different assignments in one space with one teacher. However, if two teachers were to combine forces and use both of their classrooms I believe it could be easily done. For example, if Mr. Palmer and I decide to follow this plan next year we can assign one of our rooms as the Enrichment room, while the other room would serve as the Remediation/Grade-level room. Both teachers will have the same guidelines for creating the groups, ensuring an equal experience for all kids. Once the groups have been determined they will disperse to their appropriate classroom. One teacher will facilitate the re-teaching to all of the kids that require it (this includes students from both classroom teachers). These kids will have opportunities to ask questions, see the material again, and eventually prove their mastery of the subject one more time. At the same time the other two groups will be completing an activity that pertains to the subject, with different levels of questions at the end for each of the groups. The two teachers will switch roles for the next unit.

Another point of improvement for this unit in the future can come from using actual biological data sets for the questions, as opposed to using hypothetical data sets. This is important due to Gormally et al.'s (2012) findings using the TOSLS discussed in Chapter 2. As they said, students tend to perform better on questions that use hypothetical data sets as opposed to actual data sets. This is due to hypothetical data sets using "nicer" numbers than those taken in the real world. The problem lies in the fact that nature does not always use "nice" numbers or create easily distinguishable trends. This becomes an issue when students are faced with scenarios that lack uniformity within the data. Therefore, it is vitally important that for future assessments I use actual data sets that have been collected in a scientific study. This will insure my students get an opportunity to use their skills and knowledge in a real-world scenario, not just a controlled setting in my classroom.

The overall goal of this study was to get an accurate understanding of what level 8th grade science students can analyze data. I chose to implement a pre-existing unit assessment that was tweaked slightly to fit the criteria of the study. After analyzing 36 different tests for five individual data analysis skills it was concluded that my students have some large deficiencies in their data analysis skills. These skills are some of the most important assets our students can gain while in our care, and much work needs to be done to reform our current system. The push for content knowledge must take a step back, while a surge in skill development (specifically data analysis skills) is necessary to adequately prepare our students for the future.

Several key components of my teaching will change as a result of this study. First, I will be sure to incorporate many more opportunities for data analysis throughout my units. This can be done in many different ways, like warm-ups, exit tickets, and within assignments. Second, I will increase my expectations for student performance. I have developed assignments and tests in the past that I knew students would perform well on, but this was done at the cost of expectation level. I will be very transparent with my students as to what I expect of them, but I will make sure they have adequate preparation before assessing them. Third, I will make sure my students have ample opportunities for productive struggles. This means that I will guide them through tough assignments, while at the same time allowing them to sort through their own thought processes to achieve success. Finally, I will bring the results of this study to my department for others to use. We could potentially use the QL VALUE Rubric as a department to drive student learning. We could have the students assess their learning throughout any given unit as a way for them to better understand what they need to focus on. Also, the QL VALUE Rubric will provide us as teachers a better opportunity to give meaningful feedback, not just a number grade.

**References**


Association of American Colleges and Universities. (2009). *Quantitative literacy VALUE Rubric.* Retrieved from https://www.aacu.org/value/rubrics/quantitative-literacy

Blaylock, B.K., & Kopf, J.M. (2012). The Impact of Arithmetic Skills on Mastery of Quantitative Analysis. *Education Research International*, *2012*, 1-6. doi:http://dx.doi.org.libproxy.uwyo.edu/10.1155/2012/863286

Boud, D. and Associates (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Sydney: Australian Learning and Teaching Council.

Bowen, G. M., & Roth, W. M. (2005). Data and graph interpretation practices among preservice science teachers. *Journal of Research in Science Teaching*, *42(10)*, 1063-1088.

Bravo, A., Porzecanski, A., Sterling, E., Bynum, N., Cawthorn, M., Fernandez, D.S., Freeman, L., Ketcham, S., Leslie, T., Mull, J., & Vogler, D. (2016). Teaching for higher levels of thinking: developing quantitative and analytical skills in environmental science courses. *Ecosphere 7(4). doi*: e01290.10.1002/ecs2.1290

Bray Speth, E., Momsen, J.L., Moyerbrailean, G.A., Ebert-May, D., Long, T.M., Wyse, S., & Linton, D. (2010). 1, 2, 3, 4: infusing quantitative literacy into introductory biology. *CBE Life Sciences Education 9*, 323-332.

Coil, D., Wenderoth, M.P., Cunningham, M., & Dirks, C. (2010). Teaching the process of science: faculty perceptions and an effective methodology. *CBE-Life Sciences Education 9*, 524-535.

edu.wyoming.gov/educators/standards/. Accessed on September 19, 2017.

Eshach, H., & Kukliansky, I. (2016). Developing of an instrument for assessing students' data analysis skills in the undergraduate physics laboratory. *Canadian Journal of Physics*, *94*(11), 1205+. Retrieved from http://link.galegroup.com.libproxy.uwyo.edu/apps/doc/A470463965/AONE?u=wylrc_wyoming&sid=AONE&xid=fc7175dd

Feldt, L.S. (1993). The relationship between the distribution of item difficulties and test reliability. *Applied Measurement in Education, 6*, 37-48.

Freedman, D. (2009). The Development of Theory Could Hasten the Speed with Which Data Fields Adapt to Emerging Challenges. *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*. New York: Cambridge University Press.

Goldstein, J., & Flynn, D.F.B. (2011). Integrating Active Learning & Quantitative Skills into Undergraduate Introductory Biology Curricula. *The American Biology Teacher 73(8)*, 454-461.

Gormally, C., Brickman, P., & Lutz, M. (2012). Developing a Test of Scientific Literacy (TOSLS): measuring undergraduates' evaluation of scientific information and arguments. *CBE-Life Sciences Education 11*, 364-377.

Grolemund G, Wickham H. A Cognitive Interpretation of Data Analysis. International Statistical Review , 82, 2 184-204.

Kilpatrick J, Silver EA (2000). Unfinished Business: Challenges for Mathematics Educators in the Next Decades. Yearbook (National Council of Teachers of Mathematics). Vol. 2000, 223-235

Ledbetter MLS (2012). Vision and Change in Undergraduate Biology Education: A Call to Action. The Journal of Undergraduate Science Education 11(1) A22-A26.

Munzur Z (2014). Reflections on the Impact of Absence of Summative Assessment on Students' Motivation and Learning. *Journal of Education and Future*. 6, 71-89.

National Council of Teachers of Mathematics. *Principles and Standards*. (n.d.). Retrieved

December 11, 2017 from https://www.nctm.org/Standards-and-Positions/Principles and Standards/

National Governor's Association Center for Best Practices. *Common Core State Standards Initiative.* (n.d.). Retrieved December 10, 2017 from http://www.corestandards.org/Math/Content/8/SP/

National Science Education Standards. National Academy Press. Washington DC. 1996

Rhodes, T. (2009). *Assessing outcomes and improving achievement: Tips and tools for using the rubrics.* Washington, DC: Association of American Colleges and Universities.

*TCLI Directory*. (n.d.). Retrieved July 25, 2018, from Federal Student Aid website https://studentloans.gov/myDirectLoan/tcliDirectorySearch.action

*The Nation's Report Card.* (n.d.). Retrieved May 18, 2018, from National Center for Education Statistics website, https://www.nationsreportcard.gov

Waldo J (2014). Application of the Test of Scientific Literacy Skills in the Assessment of a General Education Natural Science Program. The Journal of General Education vol. 63 num. 1

Wilkins, J. L. M. (2000). Preparing for the 21st Century: The Status of Quantitative Literacy in the United States. *School Science and Mathematics*, *100*(8), 405. Retrieved from http://link.galegroup.com.libproxy.uwyo.edu/apps/doc/A70634149/AONE?u=wylrc_uwyoming&sid=AONE&xid=ee7b639