Title
Teaching Data Science for Social Justice to Pre-Service Mathematics Teachers

Author
Pelaez, Kevin

Publication Date
2022

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

SAN DIEGO STATE UNIVERSITY

Teaching Data Science for Social Justice to Pre-Service Mathematics Teachers

A dissertation submitted in partial satisfaction of the requirement for the degree
Doctor of Philosophy

in

Mathematics and Science Education

by

Kevin Pelaez

Committee in charge:

University of California San Diego

Professor Thandeka Chapman
Professor Sherice Clarke

San Diego State University
Professor William Zahner, chair
Professor Richard Levine
Professor Susan Nickerson

2022

The Dissertation of Kevin Pelaez is approved, and it is acceptable in quality and form for publication on microfilm and electronically

Chair

University of California San Diego

San Diego State University

2022

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

Advanced Placement (AP)

Black, Indigenous, and People of Color (BIPOC)

Common Core State Standards for Mathematics (CCSM-M)

Comprehensive Assessment of Outcomes in Statistics (CAOS)

Critical Race Theory (CRT)

College Preparatory Mathematics (CPM)

Design-Based Research (DBR)

Design Feature (DF)

Exploratory Data Analysis (EDA)

Free or Reduced Priced Meals (FRPM)

Google Collaboratory (Colab)

Guidelines for Assessment and Instruction in Statistics Education I (GAISE I)

Guidelines for Assessment and Instruction in Statistics Education II (GAISE II)

Introduction to Data Science (IDS)

Levels of Conceptual Understanding in Statistics (LOCUS)

Mathematical Language Routine (MLR)

National Council for Teachers of Mathematics (NCTM)

Pre-Service Mathematics Teachers (PSMTs)

Predict-Check-Explain (PCE)

Predict-Run-Investigate-Modify-Make (PRIMM)

Problem, Plan, Data, Analysis, and Conclusion (PPDAC)

Quantitative Critical Race Theory (QuantCrit)

Teaching Experiment (TE)

Teaching Mathematics for Social Justice (TMSJ)

Teaching Statistics for Social Justice (TSSJ)

Statistical Education of Teachers (SET)

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

This work would not have been possible without the support of my dissertation advisor, Dr. William Zahner. I sincerely appreciate the time that you took to guide me through this dissertation and your mentorship during this PhD program, especially during this pandemic. Working with you has been one of the highlights of this program. I'm excited to find ways to continue to work with you in the future (and finish those publications!).

I also want to thank everyone on my dissertation committee. Your feedback and mentorship have really helped me in this dissertation. To the UCSD faculty, Dr. Thandeka Chapman and Dr. Sherice Clarke, thank you for the amazing learning opportunities at UCSD. Both of your classes inspired parts of my dissertation. Dr. Chapman, thank you for providing an introduction to education research in your EDS 252 course. That class and your dissertation feedback has helped me unpack the intersection between race and racism with statistics and data science. Dr. Sherice Clarke, thank you for your continuous support, feedback, and introduction to how learning theories may support transformative practices in EDS 251. I appreciate that you pushed me to think about learning theories and critical perspectives in this dissertation. To the SDSU faculty, Dr. Susan Nickerson and Dr. Richard Levine, thank you for providing a supportive and collaborative community. Dr. Nickerson, thank you for your leadership in the program, research rotation experiences, and overall support in this program. I always look forward to running into you at CRMSE and catching up. Finally, Dr. Levine, I really really really appreciate your continuous support that started way back when I first started the master's program. I wouldn't be here without you.

I also want to thank other researchers, peers, family, and friends that helped me in this process. Dr. Rochelle Gutiérrez, thank you for your feedback and comments. You really pushed me to think about teacher education and teaching for social justice. Nick Heyer, thank you for being a co-researcher and helping me collect data over the summer. I am also appreciative of the guidance and resources from the MODULE(S2) team, especially Andrew Ross and Jeremy Strayer. I really enjoyed the conversations that we had as I was developing the course. I also want to give the University of San Diego Student Support Services and McNair Scholars program a huge shoutout for the ongoing support and community.

Nicole Suarez and Antonio Martinez, I really appreciate our check ins and little community that we created. Talia, Isabel, and Kristin, thanks for all going to coffee shops with me, work sessions at CRMSE, and all our other adventures *insert heart with fingers*. Huge thanks to everyone in the MSED community who was ever a sounding board, helped me proofread abstracts and charters, and provided any other forms of support.

To my partner, Jaime Diaz, thanks for making sure that I'm fed, that I'm drinking water, and for making sure that we have food and a home. Nala also says gracias for walking and feeding. Yajaira and Elenai, you both have supported me way more than you know. Thanks for the 10+ years of friendship. Finally, mi mamá, papá, y hermanos. Muchas gracias por todo su apoyo. Los quiero mucho.

VITA

2014	Bachelor of Arts, University of San Diego
2018	Master of Science, San Diego State University
2022	Doctor of Philosophy, University of California San Diego, & San Diego State University

ABSTRACT OF THE DISSERTATION

Teaching Data Science for Social Justice to Pre-Service Mathematics Teachers

by

Kevin Pelaez

Doctor of Philosophy in Mathematics and Science Education

University of California San Diego, 2022
San Diego State University, 2022

Professor William Zahner, Chair

The emerging field of data science has brought attention to how we teach statistics and data science (Bargagliotti et al., 2020; Franklin et al., 2007) and prepare the next generation of statistics and data science teachers (Franklin et al., 2013). To realize the full potential of statistics and data science, researchers have also called for using data to guide conversations about race and racism (Philip et al., 2016, 2017), especially given the Black Lives Matter movement, climate change, and public health.

In this dissertation, I drew on the Teaching Mathematics for Social Justice (Gutstein, 2006), Quantitative Critical Race Theory (Castillo & Gillborn, 2022; Crawford et al., 2018; Gillborn et al., 2018), and Habits of Mind (Cuoco et al., 1996) frameworks to study the potential of using a social justice-oriented approach to teaching data science for preservice mathematics teachers, highlighting the intersectionality of race and racism with statistics and data science.

Data comes from a credit-bearing course taken by 14 students during the Summer 2021 term at a four-year public institution in the US-Mexico borderlands of Southern California. Data included pre- and post-assessments, pre- and post-task-based interviews, and classroom data (e.g., student work, whole-class recordings, field notes).

There were four research questions and analyses in this dissertation. First, there was a qualitative description of the features used to design the course centered around Freire's (1998) notion of critical consciousness and praxis with illustrations of how the design features were enacted. Second, there was a quantitative analysis of pre- and post-assessments that aimed to measure the students' statistical and data scientific content knowledge. Third, there was a qualitative analysis of pre- and post-task-based interviews that aimed to capture development of students' critical statistical and data scientific practices. Finally, elements of a focusing phenomenon framework were used to coordinate how aspects of the classroom environment (e.g., design features, tasks, tools, and the teacher) directed students' attention towards understandings of race and racism in the context of data science.

Chapter 1: Introduction

The widespread availability of big data, machine learning, and computational power has amplified the relevance of data in our day-to-day lives. For example, it is almost impossible to make it through a day without reading the news, scrolling through Facebook, or reading a policy report that does not generate data or reference or result from data-informed decision-making. As a result, mathematics policy reforms and curriculum developers have begun exploring how data science can be introduced at the K-12 level (e.g., Gould et al. 2016) .

At the same time, secondary mathematics teachers have reported feeling uncomfortable with their knowledge about statistics content and practices (Batanero et al., 2011; Franklin et al., 2015). This may be partially attributed to how the statistical education of pre-service mathematics teachers (PSMTs) is often mathematics-oriented (Burill & Biehler, 2011), although mathematics and statistics are two distinct fields. This is reflected in preparation of science and mathematics teachers. In California, secondary science teachers prepare for a single-subject credential in a specific sub-discipline (e.g., physics, chemistry, biology), but secondary mathematics teachers are expected to teach all disciplines (e.g., algebra, geometry, statistics, computer science). Furthermore, introductory statistics courses often focus on applying formulas and algorithms to simplified sets of data (Bargagliotti & Franklin, 2015; Franklin et al., 2013; Garfield & Ben-Zvi, 2008), providing students with a limited view of what it means to engage in authentic statistical investigations. As a result, Franklin et al. (2013) stated, “[t]eacher preparation in statistics must become a priority for our teacher preparation colleges and for professional development” (pp. 9-10). This preparation may include providing experiences for PSMTs to engage with statistics as a pathway to data science (Gould et al., 2017).

To realize the full potential of statistics and data science, researchers have also found that using critical perspectives in data science courses may afford opportunities to develop racial data literacy, calling for a need to prepare teachers to use data to guide conversations about race and racism in data science courses (Philip et al., 2016, 2017) and foregrounding the role of race and racism in data analysis (Crawford et al., 2018; Garcia et al. 2018; Gillborn et al., 2018; Weiland et al., 2017). This is especially important given the impact of data-informed decisions and centrality of contexts in statistics (Cobb & Moore, 1997), affording opportunities to use statistics to learn about, identify, and challenge social and racial injustices (Brantlinger, 2013).

Thus, reframing Franklin et al.'s (2013) statement, I argue that teacher preparation in *critical* statistics must become a priority for our teacher preparation colleges and for professional development for pre- and in-service mathematics teachers. I use critical to refer to making sense of our world and how data are situated within a social, cultural, historical, and political contexts. Under this critical perspective, we may shift our view of statistics and data science as apolitical tools that focus on applying formulas and algorithms to one where statistics and data science can be used as a sociopolitical tool to understand injustices and advance social change.

Purpose of the Study

The purpose of this dissertation was to study the potential that using a social-justice oriented approach to teaching content courses for pre-service mathematics teachers (PSMTs) may have on their understandings of statistics and data science as well social justice. For the purpose of this dissertation, I define data science as the intersection between statistics, computer science, and the contexts that give rise to data. In particular, statistics provides a foundation for the methods that are commonly used in data science (e.g., understanding the sampling and randomization process, A/B experimental testing, hypothesis testing, and regression that is often the foundation for more complex models). Computational abilities have allowed us to extend statistics to include maintaining and performing large-scale data analysis, using machine and deep learning methods, and new tools for visualizations. Finally, statistics and data science are always situated within a context, such as using data to gain customer insights, inform financial decisions, streamline processes, or bring awareness to social and racial injustices. I define social justice as a process (instead of an outcome) that is rooted in challenging oppression to provide equitable resources, opportunities, and responsibilities for everyone, to empower communities, and to build solidarity and collaborative action. Racial justice extends this by having an explicit focus on racial oppression. Furthermore, in this dissertation, I addressed the need to provide preservice mathematics teachers (PSMTs) with experiences to develop knowledge of statistics and social justice (Thanheiser, Harper, et al., 2020), challenged a perceived dichotomy that content (e.g., mathematics, statistics, or data science) is mutually exclusive from social and racial justice, and provided illustrations of the intersectionality of statistics and data science with social and racial justice.

This study was in alignment with the goals of statistics education reform documents, such as the Guidelines for Assessment and Instruction in Statistics Education I (GAISE I; Franklin et al., 2007) and II (GAISE II; Bargagliotti et al., 2020), the Statistical Education of Teachers (SET; Bargagliotti and Franklin, 2016), and the Common Core State Standards for Mathematics (CCSS-M; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) that call for an emphasis on using real data and technology, developing conceptual understandings, and fostering statistical thinking and practices. This project aimed to help prepare the next generation of teachers who are critical statisticians and data scientists that are able to read (i.e., identify and understand social and racial injustices) and write (i.e., engage in actions that advance social and racial justice) the world with data.

In terms of teacher education, this project began the development of a statistics and data science course for PSMTs that includes statistical and data scientific pedagogical goals as well as social justice pedagogical goals, where the social justice topics motivated the data analysis and were the learning outcomes. In terms of theoretical contributions, this study also provided practical contributions through the development of two units and design principles for designing data science classrooms for PSMTs that used a social justice-oriented approach to teaching. Finally, in terms of policy work, this study continued conversations about incorporating data science curriculum in secondary schools and promoted the idea that social justice can be included in all disciplines.

Researcher Positionality and Motivation

This dissertation is motivated by my experiences as a student, educator, statistician, and statistics education researcher. A central theme across all these experiences is the notion of *Nepantla*. Particular, in *Borderlands: La Frontera the New Mestiza*, Anzaldúa (2012) defines *Nepantla* as

> And I now call it *Nepantla*, which is a Nahuatl word for the **space between two bodies of water, the space between two worlds.** It is a limited space, a space where you are not this or that but where **you are changing**. You haven't got into the new identity yet and haven't left the old identity behind either—**you are in a kind of transition**. And that is what *Nepantla* stands for. **It is very awkward, uncomfortable and frustrating** to be in that *Nepantla* because you are **in the midst of transformation** (p. 276).

That is, *Nepantla* is the space of living between multiple worlds (borderlands), a space that is often awkward, uncomfortable, and frustrating. However, transformation occurs when these multiple worlds clash and, more importantly, are embraced. This transformation is often characterized by a framing and reframing of ourselves in relation to our world, resulting in a new set of consciousness.

I proudly identify as a bilingual (English and Spanish) first-generation Chicano. My parents are Mexican and Peruvian immigrants, but I specifically choose the term “Chicano” to recognize living between multiple racial, ethnic, and political worlds. I also use “Chicano” to emphasize the political nature of Latinidad in the USA (especially in the Southwestern borderlands of USA and México), to express political empowerment and solidarity, to recognize my Whiteness, and also to identify with my indigenous ancestry and oppose White supremacy.

For example, as a student, I was interested in mathematics and statistics as well as social and racial justice. Particularly, on one hand, I experienced mathematics in a way that people may describe as a neutral, universal, and/or apolitical (Gutiérrez, 2013; Martin, 2009; Shah, 2017). On the other hand, I was actively involved in race and ethnic-based organizations across my high school and undergraduate schooling that were centered around social justice (e.g., the Movimiento Estudiantil Chicanx de Aztlán, MEChA). I had few opportunities to explore what the intersection of mathematics and social justice looked like. In fact, I remember being told that I should consider switching majors and leaving mathematics if I wanted to pursue something more “political” or “social.” Admittedly, in my first few years in college, I internalized the idea that mathematics and social justice were mutually exclusive. As a result, I always felt like there was “Math Kevin” and “Social Justice Kevin,” but struggled to find spaces where I could blend all my passions and be myself.

As an educator, these two worlds started to clash. I was fortunate to teach mathematics at high school where I shared similar racial, ethnic, and linguistic backgrounds with my students. Given that the school was named after two social and racial justice leaders (Dr. Martin Luther

King Jr. and Cesar Chávez), I saw an opportunity to create and experiment with different curriculums that empowered Black, Indigenous, and Students of Color. This is where I first learned about teaching mathematics for social justice (TMSJ) and Freire's (1998) *Pedagogy of the Oppressed*. I was attracted to the idea of using mathematics to learn about and challenge social injustices, a perspective that I did not have in most of my own schooling. But, I quickly realized some of the challenges that come with designing and implementing a TMSJ curriculum: administrator buy-in (Gonzalez, 2009), time commitment in class dedicated to providing prior knowledge about the social justice problem context (Bartell, 2013; Gutstein, 2006), time commitment outside of class dedicated to lesson planning (Gregson, 2013), pressures from administration to meet pacing guides and prepare students for standardized assessments (Brantlinger, 2013; Gonzalez, 2009), and avoidance of reinforcing deficit narratives (Bartell, 2013; Giroux, 2001; Gonzalez, 2009) and stereotype threat (Brantlinger 2013; Rubel et al. 2016).

Nonetheless, I continued to learn as an educator by collaborating with other teachers at my school and the community. For example, I co-created a unit with the Ethnic Studies teacher to develop interdisciplinary units related to educational achievement (similar to the unit used in this dissertation), where the mathematics unit supplemented what our high school students were learning in the Ethnic Studies class. This resulted in posters and papers that used mathematics to tell stories about social and racial injustices related to climate change, homelessness, and policing. In many ways, I would say that my first experience as an education researcher started in those after school meetings where we co-created the lessons about educational achievement and found ways to highlight the intersection between social and racial justice with mathematics.

I eventually returned to graduate school and earned a master's degree in statistics. As a statistician, I recognize the importance of quantifying data, using machine learning algorithms to predict an attribute of interest, and the role that data-based arguments have in decision-making. At the same time, I also noticed that "many of these models encoded human prejudice, misunderstandings, and bias into the software systems that increasingly managed our lives...and they tended to punish the poor and the oppressed in our society, while making the rich richer" (O'Neil, 2017, p. 3). In this dissertation, I focus specifically on the role of race and racism in data science, where I believe that without carefully attending to how race and racism influence data collection, analysis, and conclusions, data may be an oppressor despite how well-intentioned an algorithm and analysis are. Data (e.g., quantifiable phenomena, symbols, images) are the result of a social process and, as a result, are susceptible to social biases. For a simple example, when I make rubrics for an assignment, I make decisions on how many points are assigned to a particular part of the task, if and when partial points are given, or how much each question is worth and why. Thus, my "objective" score on a test is the result of a social process.

As another example, Buolamwini (2017) notes how some facial recognition datasets (e.g., those used to create filters on social media applications like Snapchat and Instagram) overwhelmingly overrepresent people identified as lighter skinned (about 80% to 86% of the entire dataset) and men (about 75% of the entire data). As a result of the overrepresentation in the dataset, people identified as darker women were 32 times more likely to be misclassified (i.e., the facial recognition algorithms did not recognize faces) than people identified as lighter men. It is worth noting that this overrepresentation bias in the dataset may have not been done on purpose. However, one possible explanation for this bias is that the groups or organizations collecting the data and creating the facial recognition algorithms had an implicit bias towards men with lighter faces or blindness to their bias (which is ironic since facial recognition is part of computer vision!). Another possibility is perhaps the people that were used to pilot the facial recognition algorithm were not representative of the target population, so misclassification may have not been as evident in the initial stages of the algorithm.

Fortunately, it was possible to reduce the misclassification rate in the facial recognition data by sampling a more diverse population and running the algorithms. Furthermore, in the facial recognition example, there may appear to be no major consequences for misclassification besides not recognizing everyone's face. However, what could happen if (or when) similar datasets are used for predictive policing? In particular, we know that there are racial biases in policing, where Black, Indigenous, and People of Color (BIPOC) are more likely to be overcriminalized than their White counterparts (Staats et al., 2017). Then, if we use past data on crimes to predict future crime and allocate resources, then we run the risk of overcriminalization BIPOC people in the future. That is, algorithms are only as good as the data that is used, but if those data are the result of a social process, then those algorithms run the risk of encoding social biases.

Turning to this dissertation, *Nepantla* helps frame how I view the intersection of racial justice, statistics, and data science. Particularly, I believe that not only does data science have a significant overlap with social and racial justice, but that data has potential for playing a role in advancing social and racial justice. For example, research and analyses on achievement gaps in educational outcomes may help identify social injustices. In fact, for me, identifying social injustices through data played a pivotal role in learning about social injustices. However, stopping at identifying social injustices always left me with a "now what" feeling or like I was gap-gazing (Gutiérrez, 2008), especially when the data story may reify deficit narratives (e.g., statements that suggest that Students of Color do not perform as well as their White counterparts on standardized assessments). Thus, I think that data should be situated in its social, cultural, historical, and political contexts and also that data should be used to guide action. In the words of Freire (1998), "there is no transformation without action" (p. 68). As a result, I am interested in finding ways to use statistics and data science that go beyond identifying social and racial injustices and also include using statistics and data science to advance social and racial justice.

Finally, I end this section by discussing that *Nepantla* can often be painful, but that pain is often what leads to change. For me, I think of all the microaggressions and racist interactions where people have made me feel like I do not belong in mathematics. I think specifically of instances like when I was an undergraduate student and a professor told me to consider switching majors to sociology or ethnic studies, experiences as a statistician or data scientist where I have been told to not make everything about social justice or race because we assume that data is objective and politically neutral, or more recent experiences in my doctoral program where my peers would tell me that I should consider transferring to another program because it is clear that I am interested in social and racial justice. While these comments were often told as a joke, they perpetuated a myth that mathematics, statistics, and data science are separated from social and racial justice. To me, these comments sent the message that I did not belong in mathematics, statistics, or data science. However, I hope that this dissertation shows that there is a large and important intersection between mathematics, statistics, and data science with social and racial justice. More importantly, I hope that this dissertation helps show students (especially Black, Indigenous, and Students of Color) across all grade levels and career paths that they are welcomed and in fact needed in mathematics, statistics, and data science.

Advance Organizer

Here, I present an advance organizer for the dissertation. Chapter 2 presents the frameworks used in this study: (a) Teaching Mathematics for Social Justice, (b) Quantitative

Critical Race Theory, and (c) Habits of Mind. Chapter 3 presents the methods used in this study, including a description of design-based research and how it is applied in this study, the setting and participants, a description of the teaching experiment, data collected, and how the data was analyzed. Chapters 4 through 7 include the results. In particular, Chapter 4 presents the design features used to teach data science for social justice to pre-service mathematics teachers. Chapter 5 presents a quantitative analysis of pre- and post-curriculum-aligned assessments. Chapter 6 presents a qualitative analysis of the practices that emerged pre- and post-task-based-interviews. Chapter 7 presents a qualitative analysis of how aspects of the classroom learning environment (e.g., design features, task, tools, and the teacher) directed students' attention towards understandings of race and racism in the context of data science. Finally, Chapter 8 is the discussion and Chapter 9 is the conclusion which includes a summary of the study, limitations, and avenues for future work.

Chapter 2: Frameworks

In this chapter, I begin by presenting the three frameworks used in this study: (a) Teaching Mathematics for Social Justice, (b) Quantitative Critical Race Theory, and (c) Habits of Mind. Teaching Mathematics for Social Justice (TMSJ) was used as the design framework to guide justice-oriented instruction that interweaves content goals that focus statistical literacies and social justice goals that focus critical literacies. Notably, TMSJ is not specifically about race and racism but is broadly about social justice. Quantitative Critical Race Theory (QuantCrit) draws on elements of Critical Race Theory and builds on the assumption that data are not objective and is, instead, the result of a social and racialized process. QuantCrit was used to extend TMSJ by centering, examining, and transforming how race and racism undergird data collection, analysis, and conclusions. Finally, the Habits of Mind framework was used to describe the different statistical and critical practices that we may engage with to advance social justice in the context of a content course for pre-service mathematics teachers.

Design Framework: Teaching Mathematics for Social Justice

Teaching Mathematics for Social Justice (TMSJ) is used to teach statistics and data science for social and racial justice. My conceptualization of TMSJ draws from Paulo Freire's (1998) *Pedagogy of the Oppressed*, where "the solution is not to 'integrate' them [students] into the structure of oppression, but to transform that structure so that they can become 'beings for themselves'" (p. 55). Thus, the goal of TMSJ is not to integrate students into an existing society or, similarly, integrate social justice into an already established curriculum. Rather, TMSJ is about creating a social justice curriculum that uses mathematics to identify, analyze, and challenge social injustices.

Drawing from critical pedagogies (e.g., Frankenstein's critical mathematics education and critical mathematical literacy, 1983, 2005, 2009; Freire's liberatory pedagogy, 1998; Ladson-Billings' culturally relevant pedagogy, 1995, 2014), Gutstein (2006) presents a design framework for social justice pedagogy that marries both content pedagogical goals and social justice pedagogical goals (Berry et al., 2020; E Gutstein, 2006; Lesser, 2007). Although initially described for mathematics, Gutstein's model provides direct implications for Teaching *Statistics* for Social Justice (TSSJ; Lesser, 2007), shown in Table 2.1. Statistics classrooms may thus become spaces to learn about social issues, where the social issues are a learning objective and also motivate a need for transforming the world through data analysis. This is in alignment with the QuantCrit themes of considering how race and racism underpin data collection, analysis, and conclusions (Crawford et al., 2018; Garcia et al. 2018, Gillborn et al., 2018). Details for the statistical and social justice pedagogical goals are described in the rest of this section.

Table 2.1: Pedagogical goals for Teaching Statistics for Social Justice, modified from Gutstein (2006, p. 23) Teaching Mathematics for Social Justice

Statistical Pedagogical Goals	Social Justice Pedagogical Goals
Reading the statistical word Succeeding academically in the traditional sense Changing one's orientation to statistics and data science	Reading the world with statistics and data science Writing the world with statistics and data science Developing positive cultural and social identities

Statistical Pedagogical Goals

Statistical pedagogical goals include: (a) reading the statistical word, (b) succeeding academically in a traditional sense, and (c) changing one's orientation to statistics and data science. The first two statistical goals are in relation to normative statistics and data science content knowledge. For example, reading the statistical word often implies providing resources for students to develop statistical literacies and succeeding academically in a traditional sense entails measuring student performance. On the other hand, changing one's orientation to statistics aligns with the critical axis of equity, such as viewing statistics as a sociopolitical tool to learn about and critique society.

Reading the Statistical Word

Similar to Gutstein's (2012) classical knowledge and Freire and Macedo's (1987) dominant knowledge, Gutstein (2006) defines *reading the mathematical word*, synonymous to mathematical power, as "any set of competencies that does not gender the systemic search for the root causes of injustice, but instead leaves unexamined structural inequalities that perpetuate oppression" (p. 7). This includes statistical literacies that do not challenge social injustices (e.g., Frankenstein's functional literacy, 1994) but, nonetheless, are required to understand statistics. Furthermore, although understanding normative notions of statistical knowledge will not necessarily lead to transformative change, it may play a role in using (or perhaps reclaiming) statistics and data science as tools to advance social change.

Examples in mathematics education policy include the National Council of Teachers of Mathematics (NCTM; 2000) *Principles and Standards for School Mathematics* conceptualization of mathematical literacy (engaging in complex mathematical tasks, drawing from a variety of mathematical topics, flexible problem-solving, communicating ideas, valuing mathematics) and the National Research Council's (National Research Council, 2001) *five strands of mathematical proficiency* (conceptual understanding, procedural fluency, strategic competence, productive disposition, and adaptive reasoning). In statistics education, the GAISE I (Franklin et al., 2007) and II (Bargagliotti et al., 2020) documents provide a similar framework for statistical literacies. Furthermore, the draft Mathematics Framework for California Public

Schools: Kindergarten Through Grade Twelve (California State Board of Education, 2022) also includes a list of big ideas in statistics and data science across all primary and secondary levels.

Succeeding Academically in a Traditional Sense

Succeeding academically in a traditional sense includes performing well on standardized tests, school exit exams, attending and graduating from college, or other normative measures of success. Gutstein (2006) acknowledges that these forms of normative success are not the end goal, but they do provide students the *mathematical power* to overcome the gatekeeping mechanisms of mathematics. Ladson-Billings' (1995, 2006a) culturally relevant pedagogy calls for long-term student achievement that goes beyond standardized tests to include what students "are able to do as a result of pedagogical interactions with skilled teachers" (2006a, p. 34). For this dissertation, the current California CCSS-M (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) and GAISE II will guide what is classified as traditional academic success. The CCSS-M are composed of both traditional mathematical content standards that vary by grade and are concept specific (e.g., using a randomized experiment to compare two treatments) as well as mathematical practice standards that are reflective of mathematical Habits of Mind (Cuoco, 1996). Although statistics and mathematics are related fields, mathematics focuses on deterministic ideas whereas statistics focuses on stochastic processes that depend on variation, sampling, and the problem context (Cobb & Moore, 1997). Thus, statistics education researchers have identified statistical practices that are similar to mathematical practices, but specific to statistics and data science.

Changing Our Orientation Towards Statistics and Data Science

Finally, *changing our orientation towards statistics and data science* includes a shift from viewing statistics and data science as isolated apolitical disciplines composed of procedures to viewing them as tools that may be used to identify, analyze, and challenge social injustices. This is especially important in statistics for two reasons. First, statistics courses often focus on pre-established procedures applied to simplified sets of data (Franklin et al., 2015; Franklin, 2013; Garfield & Ben-Zvi, 2008), providing students with a limited view of what it means to engage in authentic statistical investigations. Turning to the National Research Council's (2005) five strands of mathematical proficiencies, this includes developing a productive disposition about statistics so that students can see statistics as "sensible, useful, and worthwhile" (p. 116). Second, "because data are not just numbers, they are numbers with a context" (Cobb & Moore, 1997, p. 801), the problem contexts play an essential role in statistics.

However, rather than using generic problem contexts, statistics presents an opportunity to use problem contexts as a way to explore the sociopolitical factors that lead to social injustices. This requires a utilitarian view of statistics and data science as well as an understanding how statistics are not neutral or objective and how statistics can be used to reveal social injustices (Frankenstein, 1983, 1989, 2001; Gutstein, 2006).

Social Justice Pedagogical Goals

The second half of the TSSJ goals include three social justice pedagogical goals: (a) reading the world with statistics, (b) writing the world with statistics, and (c) developing positive cultural and social identities.

Reading the World with Statistics

First, *reading the world with statistics* draws on Freire and Macedo's (1987) *reading the world*, which refers to understanding the social, cultural, historical, and political (referred to as sociopolitical in this dissertation) conditions of social injustices. This goal parallels Gutstein's (2012) *critical knowledge*, Giroux's (2001) and Frankenstein's (2001) *critical literacies*, and

Ladson-Billings' (1995) goals of sociopolitical and critical consciousness in culturally relevant pedagogies. Thus, reading the world with statistics is using statistics to identify and analyze the sociopolitical conditions that lead to social injustices. Note that this places an emphasis on the sociopolitical conditions as critiquing the larger structures of oppression is an important part of developing critical consciousness, rather than only identifying and analyzing social injustices or a critique of individuals' contexts.

Writing the World with Statistics

Second, *writing the world with statistics* draws on the Freirean notion of writing the world (Freire, 1988; Freire & Macedo, 1987). *Writing the world* refers to engaging in individual or collective actions that may lead to social justice. Reading the world and writing the world complement each other because reading the world provides an understanding of the social, cultural, historical, and underpinnings of social injustice. However, Freire (1998) notes that:

> When a word is deprived of its dimension of action, reflection automatically suffers as well; and the word is transformed into idle chatter, into verbalism, into an alienated and alienating "blah". It becomes an empty word, one which cannot denounce the world for denunciation is impossible without a commitment to transform and there is no transformation without action (p. 68)

Thus, it is important that reading the world is also accompanied by action to have the potential for transformative change.

Developing Positive Cultural and Social Identities

Finally, *developing positive cultural and social identities* draws on Ladson-Billings (1995, 2006a) notion of *cultural competence*, where students honor and maintain cultural competence as part of the curriculum while simultaneously navigating and succeeding academic institutions that were not necessarily designed for them. Similarly, Gutstein (2012) states that this includes community knowledge, or "how people understand their lives, their communities, power, relationships, and society" (p. 110). This community knowledge is similar to the identity dimension of equity (Gutiérrez, 2017), including funds of knowledge (González et al. 2005; Moll et al., 2005) that has been used to identify and leverage household knowledge that can bridge in-school and out-of-school mathematics that enhance learning (e.g., Civil, 2014, 2016; Civil & Planas, 2010).

Freire's Critical Consciousness and Praxis

Central to TMSJ is Freire's notion of critical consciousness and praxis. Freire (1998) defines *critical consciousness* as the social, cultural, historical, and political understanding of our world and an understanding that we are able to contribute to transformational change in the world. My conceptualization of transformative change is guided by Solórzano and Delgado Bernal's (2001) notion of *transformative resistance*, where the highest potential for social change includes both an understanding and critique of social oppression (related to reading the world with data) as well as a motivation to advance social justice (related to writing the world with data). Notably, transformative change may be collective, individual, or a combination of both and may include forms of internal or external resistance. Internal resistance includes moments that appear to "conform to institutional or cultural norms and expectations, however individuals are consciously engaged in a critique of oppression" (Solórzano & Delgado Bernal, 2001, p. 324). External resistance includes moments that involve "a more conspicuous and overt type of behavior, and the behavior does not conform to institutional or cultural norms and expectations" (Solórzano & Delgado Bernal, 2001, p. 325).

Critical consciousness is developed through praxis. Freire described *praxis* as "reflection and action upon the world in order to transform it" (1998, p. 52). *Reflection* includes identifying and understanding the social, cultural, historical, and political understandings of social injustices.

This is similar to reading the world with data. *Action* includes individual or collective action taken to advance social justice. This is similar to reading and writing the world with data. Reflection and action are cyclical because we are in a permanent state of discovery (Freire, 1998), which entails forming knowledge about, reflecting on, healing from, and resisting oppression as well as reflecting on that growth.

Reflection and action are complementary because action is embedded in reflection and reflection is embedded in action. Particularly, action that is not critiquing social oppression may not necessarily lead to transformational change (Freire, 1998; Solórzano & Delgado Bernal, 2001) if it is not targeting the larger social, political, cultural, or historical structures that lead to that oppression. Similarly, reflection without action may not transform or challenge oppression. Thus, action and reflection do not occur separately but, rather, are complementary.

Theoretical Framework: Quantitative Critical Race Theory (QuantCrit)

A critique of Freire's (1998) critical consciousness and Gutstein's (2006) TMSJ is that it may not explicitly foreground race and racism in social justice. That is, the frameworks are broadly about social justice and not specifically about racial justice. Given the centrality of race and racism in the USA (where this study takes place), I also draw on Quantitative Critical Race Theory (QuantCrit) to foreground the role of race and racism in statistics and data science in pursuit of racial justice.

QuantCrit has its roots in Critical Race Theory. Particularly, in 1899, Du Bois aimed to challenge biological determinism that was used to justify racial health disparities between Black and White communities. Using a combination of qualitative and quantitative methods, Du Bois (1899) conducted an analysis that aimed to account for how race and racism created poor conditions that, in turn, led to an increase of deaths in Black communities. Du Bois' analysis is transformative for two reasons. First, Du Bois presents one of the first documented statistical designs that studied the Black community in the USA. Second, Du Bois' bridges the fields of sociology and statistics by illustrating how data analysis can shift from deficit analysis (biological determinism) that serves the interest of eugenics and White supremacy, to a structural analysis (centrality race and racism) that provided a more accurate representation of racial health disparities by accounting for the relationship between historical, social, and political structures of racial inequalities.

Du Bois' (1899) analysis provides a foundation for Quantitative Critical Race Theory. QuantCrit draws from Critical Race Theory (Crenshaw et al., 1995; Ladson-Billings, 2009; Ladson-Billings & Tate, 1995; Solórzano & Yosso, 2002) and aims to center, examine, and transform how data undergirds race, racism, and power (Castillo & Gillborn, 2022; Crawford et al., 2018; Covarrubias, 2011; Covarrubias et al., 2018; Garcia et al., 2018; Gillborn et al., 2018; Pérez Huber et al., 2018; Sablan, 2019). A guiding principle of QuantCrit is that data is not objective. Under this perspective, "data is no less socially constructed than any other form of research material" (Gillborn et al., 2018, p. 158). In other words, data is a result of a social process that aims to *encode* phenomena, including racial prejudices and biases that may exist in the data collection, analysis, and conclusions.

The QuantCrit tenets and brief descriptions are shown in Table 2.2. The tenets provide direct implications for the ways of doing and engaging with statistics and data science. The first tenet (the centrality of race and racism) notes that without a critical race-conscious perspective, statisticians and data scientists run the risk of reifying deficit narratives. For example, Du Bois (1899) noted how racial health disparities were used to justify eugenics and served the interest of White supremacy. In mathematics education, Gutiérrez (2008) describes how gap-gazing research (e.g., focusing on achievement gaps) perpetuates deficit narratives that justify differences between groups and, in doing so, create racial hierarchies of who is capable of learning and why.

Table 2.2: Tenets of QuantCrit with brief descriptions from Gillborn et al. (2018)

QuantCrit Tenet	Description
1. The centrality of race and racism	QuantCrit recognizes that racism is a complex, fluid and changing characteristic of a society that is neither automatically nor obviously amenable to statistical inquiry. In the absence of a critical race-conscious perspective, quantitative analyses will tend to remake and legitimate existing race inequities (p. 169)
2. Numbers are not neutral	QuantCrit exposes how quantitative data is often gathered and analyzed in ways that reflect the interest, assumptions, and perceptions of White elites. One of the main tasks of QuantCrit, therefore, is to challenge the past and current ways in which quantitative research has served White supremacy (p. 170)
3. Categories are neither 'natural' nor given: for 'race' read 'racism	QuantCrit interrogates the nature and consequences of the categories that are used within quantitative research. In particular, we must always remain sensitive for possibilities of 'categorical alignment' (Artiles, 2011; Epstein, 2007) where complex, historically situated, and contested terms (like race and dis/ability) are normalized and mobilized as labeling, organizing, and controlling devices in research and measurement. Where 'race' is associated with an unequal outcome it is likely to indicate the operation of racism but mainstream interpretations may erroneously impute 'race' as a cause in its own right as if the minoritized group is inherently deficient somehow (p. 171)
4. Voice and insight: data cannot 'speak for itself'	QuantCrit recognizes that data is open to numerous (and conflicting) interpretations and, therefore, QuantCrit assigns particular importance to the experiential knowledge of people of color and other 'outsider' groups (including those marginalized by assumptions around class, gender, sexuality, and dis/ability) and seeks to foreground their insights, knowledge, and understandings to inform research, analysis, and critique (p. 173)
5. Using numbers for social justice	QuantCrit rejects false and self-serving notions of statistical research as value-free and politically neutral. CRT scholarship is oriented to support social justice goals and work to achieve equity.

The second (numbers are not neutral) and third tenets (categories are neither 'natural' nor given) bring attention to the assumptions of sampling and measurement tools and how they may serve in the interest of White supremacy. Similar to Weiland (2017), these tenets suggest that it is important for statisticians and data scientists to carefully attend to the sociopolitical context of data collection, analysis, and conclusions. For example, Du Bois (1899) brought attention to the sociopolitical contexts (e.g., differentiated access to health care) that led to racial health disparities. This places an emphasis on the structures that led to the phenomenon rather than placing responsibility on the individual. Similarly, Gutiérrez (2008) notes that a gap-gazing lens provides a limited view of educational equity that may focus on measurable outcomes (e.g., teacher knowledge, standardized assessments) but fails to account for the larger structural social factors (Ladson-Billings, 2006b).

The fourth tenet (voice and insight: data cannot 'speak' for itself) is about foregrounding experiential knowledge of stakeholders (e.g., Black Indigenous, and People of Color, students, families) to make sense of multiple, sometimes conflicting interpretations. This is especially important when trying to humanize data and situate it within real-life experiences. Finally, the last tenet (using numbers for social justice) challenges the assumption that data is objective, apolitical, or authoritative (Gillborn et al., 2018). This assumption may be attributed to people feeling intimidated by data (Crawford et al., 2018) or because machine learning algorithms are often portrayed as black box methods (models that are not straightforwardly interpretable) . As a result, it is important for statisticians and data scientists to foreground the role of race and racism in data science (Philip et al., 2016), take an anti-racist stance to data analysis, critique deficit data-based arguments, and use data to advance social justice (Castillo & Gillborn, 2022; Gillborn et al., 2018).

Conceptual Framework: Mathematical Habits of Mind

Research about statistical practices provides illustrations of what it means to engage with data in a way that is aligned with the QuantCrit tenets. This research is rooted in Cuoco et al.'s (1996) notion of mathematical habits of mind and communities of practice (Brown et al., 1989;

Lave & Wenger, 1991). Particularly, Cuoco et al. (1996, p. 375) stated that "for generations, high school students have studied something in school that has been called mathematics, but has very little to do with the way mathematics is created or applied outside of school." Cuoco and colleagues argued for a shift in how we view mathematics (and the learning of mathematics) from one that is about applying mathematical properties and memorizing objective facts to a view that includes the ways in which we practice mathematics. This parallels research that suggests that learning happens in situated authentic environments and encourages students to do the practices that are reflective of what experts in that field would do when working with a particular problem (Brown et al., 1989). These practices, or *mathematical habits of mind* (Cuoco et al., 1996), are: (a) reflective of what mathematicians do (Chance, 2002; Levasseur & Cuoco, 2003), (b) interconnected and build off each other (Lee & Tran, 2015), (c) eventually become automatic processes when engaging with mathematical tasks (Goldenberg, 1996), and (d) are developed throughout the mathematical problem-solving process (Levasseur & Cuoco, 2003). Furthermore, it is possible that content and technology might evolve over time, but habits of mind will remain transferable and relevant.

Research about mathematical practices and habits of mind has shaped mathematics educational policy and practices. For example, the National Council for Teachers of Mathematics (NCTM) *Principles and Standards* (2000) included Process Standards that describe ways of *doing* mathematics. These processes cross all grade levels and disciplines of mathematics, and include processes like using and adapting a variety of appropriate strategies, making mathematical arguments, using language to communicate mathematical ideas precisely, and recognizing and using connections between mathematical ideas when problem-solving. Around the same time, the National Research Council's (NRC) *Adding it Up* (2001) provided five strands of mathematical proficiency, which included representing mathematical situations in different ways (strategic competence), making connections between mathematical ideas (conceptual understanding), and flexibly, accurately, efficiently, and appropriately executing procedures (procedural fluency). More recently, the Common Core State Standards for Mathematics (National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010) published the eight Standards for Mathematical Practice (SMPs). The eight SMPs include a combination of the NCTM's process standards and the NRCs strands of mathematical proficiency.

Regarding statistical practices and teacher education, Burrill and Biehler (2011) asked: "What are the statistical habits of mind teachers and students should develop as they grow in their understanding of the fundamental concepts in statistics?" (p. 66). Similar to mathematical practices and habits of mind, I broadly define statistical practices as the ways of engaging with statistics that are reflective of how statistics and data scientists engage with statistical investigations. For my dissertation, I draw on two bodies of research to define and illustrate statistical practices: (a) the general statistics education research, and (b) critical perspectives to statistics and data science (Castillo and Gillborn, 2022; Crawford et al., 2018; Garcia et al., 2018; Gillborn et al., 2018; Philip et al., 2016; Weiland, 2017).

First, statistics education researchers have built on research about mathematical habits of mind and practices to describe statistical practices. It is worth noting that "thinking," "reasoning," and "literacy" are often used interchangeably (Chance, 2002) to refer to statistical practices, important statistical concepts, and implications for instruction. Nonetheless, research about statistical concepts (e.g., variation) and implications for instruction (e.g., having students create multiple visualizations for distributions) provide implications for practices (e.g., looking for and visualization variation in data). In one of the few statistical education research articles that specifically references the term “habits of mind”, Lee and Tran (2015) define *statistical habits of mind habits* as practices that are “productive for engaging in while doing statistics” (p. 1). In terms of data science, Gould (2017) posits that data literacies are statistical literacies, but adds that data science should also account for the modern ways in which data is used (e.g., how data is stored, data privacy and ethics). Combined, statistics and data science education researchers have focused on statistical practices that differentiate statistics from mathematics. For example, some practices include relating data and conclusions back to the problem context (Burrill & Biehler, 2011; Chance, 2002; Franklin et al., 2007; Lee & Tran, 2015; Visnovska & Cobb, 2019), paying attention to sources of data and how they are related to the attribute of interest (Burrill & Biehler, 2011; Gould, 2017; Lee & Tran, 2015), using and comparing visualizations to communicate data patterns (Burrill & Biehler, 2011; Gould, 2017; Lee & Tran, 2015; Pfannkuch & Rubick, 2002; Pfannkuch & Wild, 2004), anticipating and accounting for variation (Franklin et al., 2007; Lee & Tran, 2015), remaining skeptical (Chance, 2002; Lee & Tran, 2015), and asking questions (Gould et al., 2017) as we problem solve with data.

Researchers have also highlighted the importance of developing critical practices to identify, analyze, and challenge social injustices. For example, drawing from critical education research (Freire & Macedo, 1987; Giroux, 2001; Gutstein, 2006) and in alignment with the QuantCrit tenets (Crawford et al., 2018; Garcia et al. 2018, 2018; Gillborn et al., 2018), Weiland (2017) presented a framework for critical statistical literacy that foregrounds *reading* (identifying and learning about social injustices) and *writing the world with statistics* (engaging in actions that challenge social injustices). By statistical literacy, Weiland (2017) refers to functional mathematical literacy (Frankenstein, 1994), or ways of teaching statistics that prepare students for a workforce and do not have an explicit critique of social or racial injustices. Statistical literacy may also be associated with the statistical pedagogical goals of teaching statistics and data science for social justice. Weiland expands statical literacy to include critical statistical literacy, or ways of using data reading (or reflection from praxis) and writing the world (or action from praxis) with statistics and data science.

A helpful characteristic of Weiland's (2017) framework is that it illustrates the similarities between statistical literacy, critical literacy, and critical statistical literacy, showing how statistical and social justice pedagogical goals may be interweaved. Examples are shown in Table 2.3. One example focuses on reading the world with statistics by using statistics to identify and learn about social structures and the other example focuses on writing the world with statistics by discussing, communicating, and reshaping social structures. For this dissertation, Weiland's framework provides an illustration of the intersectionality between statistics or data science and social justice, where critical statistical literacies may also be interpreted as critical statistical consciousness.

Table 2.3: Example from Weiland's (2017, p. 41) framework for critical statistical literacy

	Reading	Writing
Statistical Literacy	Make sense of and critiquing statistical and quantitative data-based arguments encountered in diverse contexts	Discussing or communicating the meaning of statistical information
Critical Literacy	Identifying and interrogating social structures in our world	Actively influencing and shaping structures in society
Critical Statistical Literacy	Identifying and interrogating social structures which shape and are reinforced by data-based arguments	Using statistical investigations to communicate statistical information and arguments in an effort to destabilize and reshape structures of injustice for a more just society

In what follows, I synthesize and organize the implications of research about statistical practices. The practices are categorized into four groups: (a) transnumeration, (b) variability, (c) interpreting data, and (d) implications of data analysis. I will define each practice using the related literature and illustrate the practices shown the task shown in Figure 2.1. In this task, students are provided data about 102 middle schools at Dolores Unified School District (DHUSD) and asked to address the following main question: Which three schools should we visit and why?

This task is intentionally designed to use traditional achievement data as measured by standardized tests for four reasons. First, unlike common data sets (e.g., the Iris flower data set), this task affords opportunities to explore social injustices. Second, it is common for parents and families, school leaders, policymakers, educational researchers, and other stakeholders to use school achievement data as a proxy for "high quality" education. Third, and in response to the second reason, the task highlights how someone may engage in gap-gazing and reify deficit narratives about schooling (e.g., which types of schools are "better") without a careful and intentional focus on how race and racism (along with socioeconomic status) shape schooling experiences. Finally, combining the first two points relates to Gutiérrez's (2016) idea of *creative insubordination* by repositioning and challenging deficit narratives, for example, by decentering the achievement gap. This includes participants recognizing the tensions of using standardized assessments to measure learning.

Transnumeration

Wild and Pfannkuch (1999) first described transnumeration as "forming and changing data representations of aspects of a system to arrive at a better understanding of that system" (p. 227). Forming data includes transforming raw data into graphical representations and summaries

Background: The superintendent of the Dolores Huerta Unified School District (DHUSD) recently read the same article ((article provided in the beginning of the interview)). They are interested in learning more about the great mathematics teaching and learning and their middle schools. To highlight some of this awesome work, the superintendent is interested in visiting three schools. Since DHUSD knows that you are familiar with statistical investigations, they are hoping that you can help them design, explore, and analyze data for this project.

Task: DHUSD would like you to make one or two recommendations that would address the following questions. The only requirement is that you use at least one linear regression model.

Question 1: Which three schools should we visit and why?

Question 2: What are predictors of a school's mathematics learning?

Question 3: What evidence do you have to support your questions? Include any necessary plots and analysis.

District Information: DHUSD is one of the largest districts in the state. The rising bio and tech industry, proximity to an international border, and racial and ethnic diversity has shaped the history of the city.

Data:
The data, variable names, and type of data are summarized below:

Variable Name	Description	Data Type
CDSCode	School code assigned from the government	Text
School	Name of the middle school	Text
Charter	Marks if the schools is charter or not charter	Binary: 0 - Traditional, 1 - Charter
PercentFRPM	Proportion of students who qualify for free or reduced priced meals (FRPM)	Numeric
BIPOC	Proportion of students classified as Black, Indigenous, or People of Color (BIPOC)	Numeric
TotalEnrollment	Total enrollment at the middle school	
MeanScaleEng6 / 7 / 8 MeanScaleMath6 / 7 / 8	Average score of all students on a 6th, 7th, or 8th grade English and Math standardized assessment	Numeric
PercentMetAboveEng6/ 7 / 8 PercentMetAboveMath6/ 7 / 8	Percent of students that met or exceeding the 6th, 7th, or 8th grade English / Math standards on a standardized assessment	Numeric

Figure 2.1: Sample statistical task used to illustrate statistical practices

to make meaning of data (Pfannkuch & Rubick, 2002; Pfannkuch & Wild, 2004). Enhancing data includes comparing and augmenting data to enhance the meaning (Lee et al., 2014; Lee & Tran, 2015; Pfannkuch & Wild, 2004). Forming data typically precedes changing data, where forming data may involve exploratory visualizations that are later enhanced by changing data.

Forming Data

An important part of analyzing data is creating meaningful representations of the data (Burrill & Biehler, 2011; Lee & Tran, 2015). This often includes using data to visualize statistical measures (e.g., measures of spread, variation), identify subsets of data (e.g., outliers), and examine data structures (e.g., missing variables). In doing so, students may begin to engage in an exploratory data analysis (EDA), or the beginning steps of analyzing data that involve summarizing and describing data. It may be the case that students begin to visualize data without a clear purpose in mind (Gould, 2017). Nonetheless, forming data may lead to identifying potential research questions (Burrill & Biehler, 2011) and identifying patterns that may set up future data analysis.

For example, in the task shown in Figure 2.1, students may choose to create a boxplot to show the distribution of the percent of students that met or exceeded the eighth-grade mathematics standards (MathPercentMetAndAbove8), shown in Figure 2.2. This may help students identify the spread of the data (e.g., is it skewed or not, range), any measures of center (e.g., mean or median), and help identify any violations of assumptions (e.g., if it is normally spread and can be used in a regression model).

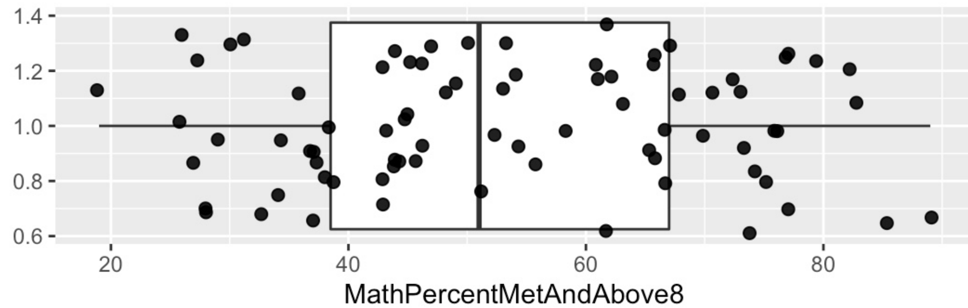


Figure 2.2: Boxplot of the percent of students that met or exceeded the eighth-grade mathematics standards (MathPercentMetAndAbove8)

As another example, students may begin the task from Figure 2.1 by graphing relationships between two variables as an informal analysis, as shown in Figure 2.3. In the left graph, we see that there may be some outliers (circled in red) when graphing the percent of students that met or exceeded sixth grade mathematics standards (MathPercentMetAndAbove6) related to the school's total enrollment (TotalEnrollment). This may prompt students to look further into why those outliers exist, if they should be removed, and, if removed, how removing the outliers may affect the inferences.

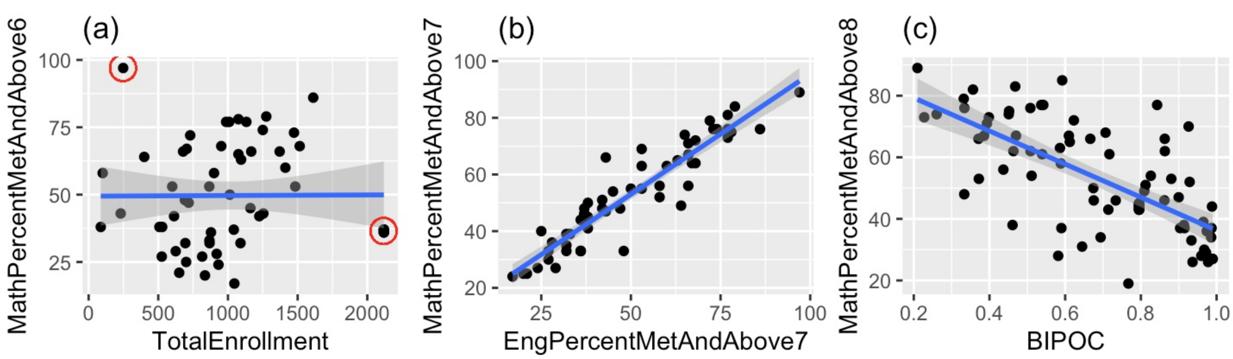


Figure 2.3: Percent of students that met or exceeded mathematics standards related to: (a) a school's total enrollment, (b) the percent of students that met or exceeded English standards, (c) and percent of students identified as BIPOC

Similarly, in the middle graph, students may see that the percent of students that met or exceeded the seventh-grade English standards (EngPercentMetAndAbove7) has a strong linear relationship with the percent of students that met or exceeded the seventh-grade mathematics standards (MathPercentMetAndAbove8). The outliers and strength of the relationships may be later confirmed with a formal statistical analysis (e.g., statistical measures to determine if a point is an outlier, regression analysis).

Enhancing Data

Complementing and extending forming data is changing data to enhance meaning (Lee et al., 2014; Pfannkuch & Rubick, 2002; Pfannkuch & Wild, 2004). This often includes graphical augmentations (e.g., shading points to show multivariate relationships, adding squares to a regression line to visualize residuals), highlighting certain patterns and structures of the data (Lee et al., 2014), and comparing visualizations to communicate different meanings (Lee & Tran, 2015). For example, Lee et al. (2014) documented how 62 teachers engaged in transnumerative practices while working on statistical tasks. They found that 77% of all teachers create at least one visual representation, with 27% using at least three types of representations. In terms of changing data, they found that 63% of teachers added some statistical measure to their graph and 53% also added augmented features. Altogether, changing data allows students to analyze data during the exploratory data analysis, augment visualizations to support their claims, and enhance their ability to communicate data (Lee et al., 2014).

Turning to the sample task, students may also choose to add a gradient scale to the points to show how the concentration of students who receive free or reduced priced meals (FRPM) varies across the data, shown in the bottom graph of Figure 2.4. Adding the FRPM layer shows that there might be a relationship between the proportion of students that met or exceed the eighth-grade standards and the proportion of students that received FRPM at a school, allowing students to engage in informal data analysis before using formal statistical models (e.g., t-tests, regression).

Variability

A defining characteristic of statistics is the role of variation (Burrill & Biehler, 2011; Cobb & Moore, 1997; Franklin et al., 2005; Groth, 2013; Pfannkuch & Rubick, 2002). Thus, students should be able to anticipate, look for, and control for variation as they are engaging with a statistics task (Franklin et al., 2007; Lee & Tran, 2015; Wild & Pfannkuch, 1999). This includes creating and recognizing that statistical questions are about describing noise in data (e.g., "what is the average SAT score across all schools in a district") instead of deterministic properties that are more mathematical (e.g., "what is the SAT score of one student"), using random sampling to model the variability from a target population in a sample, and using distributions to model and account for variability (Franklin et al., 2007).

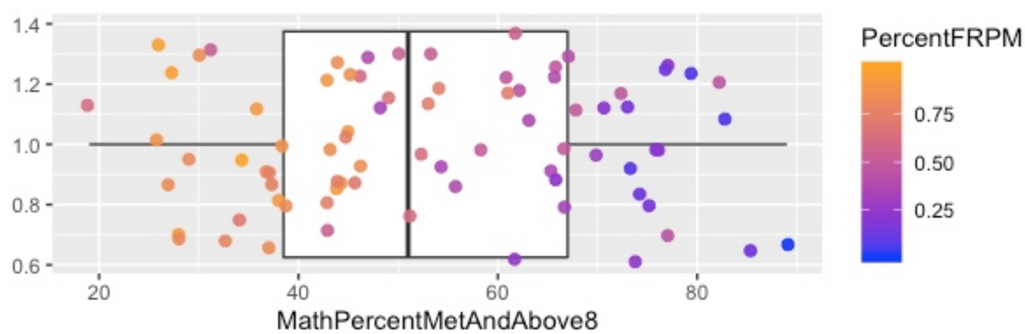


Figure 2.4: Transnumeration of Figure 2.2 by shading points (graphical augmentation) according to the percent of students who received FRPM

Anticipating and Looking for Variation

Some examples of anticipating and looking for variation include describing variability within a group, variability within and across groups, covariance, and variability in model fitting (Franklin et al., 2007). Variability within a group may include calculating the standard deviation of the standardized assessments scores in DHUSD to understand the spread of a distribution. Variability within and across groups includes comparisons across groups, such as the average difference in the averages of the scores between two schools. Covariability includes a relationship between two groups, such as how the percent of students that met or exceeded the English standards may be used to find the percent of students that met or exceeded the English standards (Figure 2.3). Finally, variation in model fitting may include how well the regression lines in Figure 2.3 predict a variable using measures of fit (e.g., residual plots, R^2 values, mean squared error). Variation in model fitting is perhaps the most common form of variability in data science, particularly since model fitting and validation is important in predictive modeling.

Generalizability

Another practice related to sampling and variability is considering the generalizability of statistical studies (Franklin et al., 2007; Lee & Tran, 2015). This includes considering how measurement tools may reproduce similar results in other contexts as well as specifying under what constraints the inferences are generalizable. If data is not generalizable to a target population, students may change or edit their research questions to be generalizable to the appropriate population or specify a population to which their study may be generalizable to. This is especially important in statistics and data analysis since a goal is to provide inferences and implications from a sample to a larger population.

Table 2.4: Schools with the highest percent of students that met or exceed the eighth grade-mathematics standards

School	% BIPOC	% FRPM	% Met or above 6th grade	% Met or above 8th grade	% Growth	% Change
RRM	21	<1	97	89	-8	-8
MVM	59	12	79	85	+6	8
JJH	36	45	58	82	+24	41

In the sample task, students may notice that the top three schools in Table 2.4 are not representative of the school district. Particularly, the percentages of BIPOC (21%, 59%, 36%) and students that qualify for FRPM (<1%, 12%, 45%) are much lower than the district's demographics (65% BIPOC and 60% FRPM). As an alternative, students may choose to consider ways to split the data or different sampling methods, such as by using stratified sampling to ensure that at least two of the schools are majority BIPOC.

Interpreting Data

Another defining characteristic of statistics is that "data are not just numbers, they are numbers with a context" (Cobb & Moore, 1997, p. 801). Thus, the context provides meanings to the data. This suggests that students should attend to sources of data and how they are related to the attribute of interest as students are engaging in statistical tasks (Burrill & Biehler, 2011; Lee & Tran, 2015).

Relevance of Data

Reading data includes considering how well data measures an attribute of interest in a statistical task (Burrill & Biehler, 2011; Chance, 2002; Lee & Tran, 2015; Visnovska & Cobb, 2019). This may include drawing on experiential knowledge to understand the problem context and working with an expert in the field to ensure that the data is well suited to answer research questions and measure the attribute of interest. If data is not relevant or cannot appropriately be used in the task, then students may consider collecting other data (Chance, 2002; Burrill & Biehler, 2011) or developing other statistical questions that are responsive to the data that has already been collected (Lee & Tran, 2015). This is especially important to ensure that the data analysis and conclusions are not based on inappropriate data.

For example, in the task shown in Figure 2.1, students may recognize that standardized assessments are a common (and dominant) measure of educational equity. In fact, organizations like GreatSchools.org, the California School Dashboard, and its predecessor the Academic Performance Index assign schools ratings that are often a function of these standardized assessments, where higher scores lead to higher ratings. Thus, to address the first question in the task (which three schools should the superintendent visit), a student may be interested in identifying the top three schools with the highest eight-grade assessments. This may lead to the schools shown in Table 2.4. This way of identifying the top three schools is not necessarily critical because it does not consider the social, cultural, historical, and political context. Interestingly, the RRM school showed a decrease from sixth to eighth grade. Nonetheless, RRM still had the highest percent of students that met or exceeded the eight-grade standards.

Alternatively, students may consider creating new data. Such data may include looking at the growth (difference between the eighth-grade and sixth grade assessments) or the percent change (difference divided by the initial value). In this case, the rationale for using growth and percent change may not be explicitly motivated by the social, cultural, historical, and political context. The top three schools using these methods are shown in Table 2.5 and 2.6.

Table 2.5: Schools with the highest percent growth

School	% BIPOC	% FRPM	% Met or above 6th grade	% Met or above 8th grade	% Growth	% Change
JJH	36	45	58	82	+24	41
OGM	54	47	43	61	+18	42
JM	91	88	21	38	+17	81

Table 2.6: Schools with the highest percent change

School	% BIPOC	% FRPM	% Met or above 6th grade	% Met or above 8th grade	% Growth	% Change
JM	91	88	21	38	+17	81
CM	99	98	17	27	+10	59
GM	59	83	25	37	+12	48

To determine the relevance of data, students may also draw on their personal experiences and familiarity with the problem context as well as the knowledge of others to examine the relevance of data. For example, to address the second question (relationship between collected variables and performance on a standardized test), students may recognize that variables like the percent of students that receive FRPM (proxy for socioeconomic status), BIPOC, and students classified as English Learner are often treated as a measure of diversity and, consequently, may be of interest when considering educational equity. Similarly, students might be interested in seeing if there is a relationship between the percent of students that met or exceeded the English standards and mathematics standards. Conversely, variables like a school's total enrollment may not be of interest. These relationships are shown in Figure 2.3. As shown, the number of students at a school does not have a strong relationship with the percent of students that met or exceeded the math standards. However, the percent of students that met or exceeded the English standards and the proportion of students identified at BIPOC show a stronger relationship.

Sociopolitical Nature of Data

Building on the previous practices and similar to reading the world with statistics, another practice that students may develop is considering the social, cultural, historical, and political nature of data (Castillo & Gillborn, 2022; Gillborn et al., 2018; Weiland et al., 2017), specifically the racialized context of the data (Philip et al., 2016). This is related to the second, third, and fourth tenets of QuantCrit (see Table 2.2). Particularly, the third tenet states that numbers and categories are not neutral, thus we should consider how data is situated in a sociopolitical context. The fourth tenet aims to "foreground their [Black, Indigenous, and People of Color's] insights, knowledge, and understandings to inform research, analysis, and critique" statistics (Gillborn et al., 2013, p. 173), such as by drawing on their experiential knowledge. Similarly, Weiland (2017) suggests that students should evaluate the source, collecting, and reporting of data, and consider how they are shaped by a sociopolitical context. Thus, students should consider who the data serves (individuals or structure), assumptions about the data, and how the data may reify or serve for social injustices.

For example, students may question the relevance of using standardized tests to measure high quality schooling. Particularly, students may recognize that using standardized test scores to measure school performance is in line with Gutiérrez (2017) dominant axis of equity that fails to account for a students' identity and social transformation. Turning to the sample task, ranking the schools by the percentage of students that met or exceeded the mathematics standards (Table 2.4) does not paint a full picture of the type of learning that occurred in the classrooms if the racialized context is not foregrounded in the interpretation.

At the same time, students may come to view the achievement gap as an important step in investigating education equity (Gutiérrez, 2017). Thus, in an effort to decenter gap-gazing and address the task in Figure 2.1, students may be guided to considering measures of achievement that go beyond the percent of students that met or exceeded mathematics standards. This may require that students consider other quantitative measures, such as growth (Table 2.5) or percent change (Table 2.5) that may move away from defining learning as what students know at the time of an assessment to learning as percent growth. As a result, students may consider showing the relationship between demographic variables (e.g., percent BIPOC) and learning as growth or percent change. Graphs showing these relationships are shown in Figure 2.5 as well as a colored layer of the percent of students who qualified for FRPM. Here, we see that not only was there an increase in the percent of students that met the mathematics standards from sixth to eighth grade in the majority of schools (71%), but schools that were predominantly BIPOC and low-income showed larger increases than their counterparts. Countering the approach in the previous category that does not foreground the sociopolitical context, this approach highlights how learning does occur in schools with high proportions of BIPOC and lower-income students and, more importantly, that we can learn about the success stories at these schools.

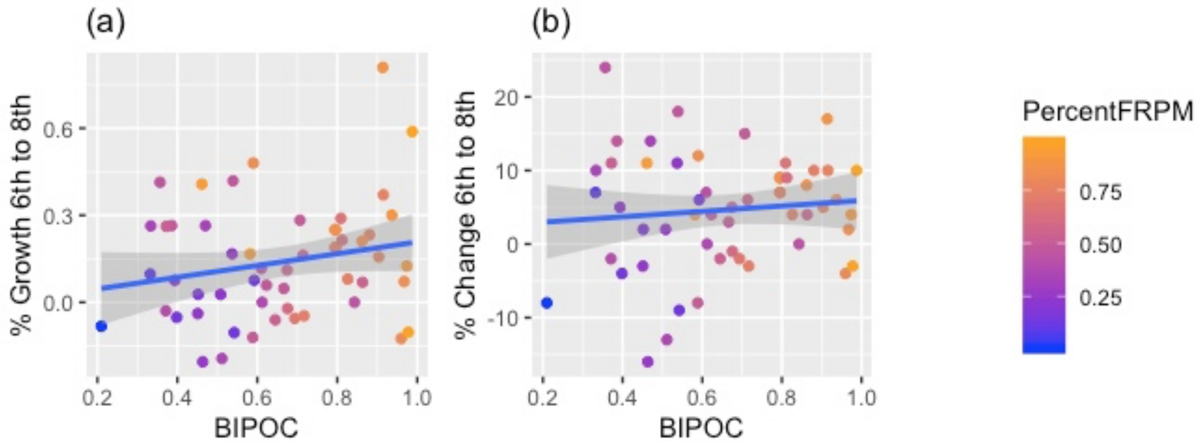


Figure 2.5: Percent growth and change related to the percent of students identified as BIPOC layered with the percent of students that received FRPM

There are a couple of characteristics in the illustration of this practice to unpack. First, it is worth noting that the task in Figure 2.1 is intended to raise some of the tensions with data science and social justice, particularly around gap-gazing and reclaiming data to advance social justice. Particularly, the nature of the task is to predict student performance and use student performance to measure learning, but the task is open so it allows analysts to engage in forms of creative insubordination (Gutiérrez, 2016) that add a nuance to what constitutes achievement. In this example, this may be enacted by considering growth across grade levels instead of one grade level or thinking about ways in which the data is representative of a BIPOC and low-income students in the district. This balance between attending to the dominant axis of data while simultaneously considering a critical lens is related to what Gutiérrez refers to as *Nepantla*, or the tensions between playing the game and changing the game. These tensions are especially amplified in educational equity contexts where, on one hand, quantifying learning may provide an insight into the types of learning that occur. On the other hand, quantifying learning may reduce learning to a number that does not paint a full picture of what and how students learn, as is the case with standardized assessments. Nonetheless, Weiland (2017) recommends that we should acknowledge and state those tensions. Thus, this practice is about recognizing that numbers and categories are not neutral as it is about acknowledging, stating, and *embracing Nepantla* when analyzing data.

Second, this practice is not only about including traditional diversity markers in data analysis. In fact, Table 2.5 and 2.6 could both be created without attending to diversity markers. Rather, this practice is about how students consider the sociopolitical nature of measurement tools and their implications for data analysis and inferences. For example, one student may consider the relevance of the data and choose to use percent change since it is a measure that considers growth over time. This approach is more related to the relevance of a variable and its mathematical implications (percent change instead of using one point) regardless of the implications or assumptions in the data. Conversely, a student may think about the sociopolitical nature of the data by asking if standardized assessments are reflective of individuals (e.g., student learning) or systems (e.g., teaching to the test or access to resources like tutoring), who is and is not included in data (Lesser, 2007), or question how race and racism underpins the data collection and analysis. This subtle but important difference is what separates students considering the relevance of data and students considering the sociopolitical and racialized nature of data.

Implications of Data Analysis

Complementing interpreting data is considering the implications of data analysis. This practice is about relating conclusions back to the problem context (Burrill & Biehler, 2011; Chance, 2002; Franklin et al., 200; Lee & Tran, 2015; Visnovska & Cobb, 2019).

Implications of Data

Simply reporting a *p*-value or effect size to determine a statistical significance is not enough. Rather, students should consider the problem context when providing data-based conclusions (Chance, 2002; Franklin et al., 2005; Lee & Tran, 2015; Pfannkuch & Wild, 2004; Visnovska & Cobb, 2019). A common example is contextualizing the difference between correlation and causation. For example, Rossman (1994) and Rossman and Chance (2001) use data where the life expectancy and the number of people per television in 40 countries have a moderate negative correlation ($r = -0.606$). A causal and decontextualized interpretation of this may imply that decreasing the number of people per television (less people share a television) increases the average life expectancy for people in that country. However, considering the context may shed some light into other factors that describe that association, such as countries with a lower person per television ratio probably have more access to healthcare.

A similar approach can be taken to the graph showing the relationship between the percent of BIPOC students and the percent of students that met or exceeded the eighth-grade mathematics score (graph c in Figure 2.3). Here, students might consider that the implication is not that increasing the percent of BIPOC students decreases the percent of students that met or exceeded the eight- grade mathematics scores. Rather, there may be other factors that explain that situation, such as socioeconomic status, the average experience of teachers at a school, or racial biases in the tests.

Sociopolitical Implications of Data

Extending the previous practice, Weiland (2017) emphasizes the importance of foregrounding the sociopolitical context when providing data-based arguments. Philip et al. (2016) and QuantCrit extend this by foregrounding the role of race and racism in data-based arguments. Particularly, data analysis and implications should be situated within the sociopolitical and racialized context to avoid reifying or reinforcing inequities. For example, in a literature review about the representation and generalizability of physics education research, Kanim and Cid (2020) analyzed 417 papers about physics education research from 1970 to 2015. They found that the majority of the research tended to focus on affluent, higher-tracked, predominantly White settings. While this emphasizes the practice of considering generalizability stated above, Kanim and Cid (2020) also note concerns about the implications of the physics education research. Particularly, they stated how there is an "implicit expectation that the results obtained by a researcher in one physics classroom should be replicable in another physics classroom" (p. 16), but this may homogenize student populations or exacerbate educational inequities. Furthermore, they stated that while the effects of generalizing from non-representative research is difficult to assess, Kanim and Cid (2020) bring awareness to how data-based arguments may be weaponized for the oppressor.

For example, using the percent of students that are identified as BIPOC to predict a school's performance (Figure 2.3) may send the message that learning does not happen in schools with predominantly BIPOC and lower-income populations, and, conversely, "high quality" instruction is associated with whiteness. In other words, without carefully attending to issues of race and racism, data analysis may reinforce deficit narratives about the schooling of BIPOC and lower-income students.

What separates this practice (Sociopolitical Implications of Data) from the previous practice (Implications of Data) is the explicit focus on avoiding reifying or reinforcing inequities or deficit narratives. Particularly, it is considered good practice to recognize that correlation does not imply causation. However, considering the sociopolitical implications of data entails attending to how data may be weaponized to serve the interest of the oppressor, explicitly naming those tensions (Weiland, 2017), and working to avoid weaponizing data analysis.

Summary of Statistical Practices

The statistical practices are reflective of what statisticians and data scientists may do as they collect, analyze, and make data-based conclusions. Although the practices are presented and illustrated apart from each other, they are all interconnected and build off each other (Lee & Tran, 2015). For example, visualizing distributions may help describe variability within and across the data. Further, the sampling methods, generalizability, and relevance of the data may inform constraints of data implications. More importantly, statistical practices may be praxis themselves if they are considering the sociopolitical and racialized context of data and being used to advance social justice.

Statistical Investigation Cycle

Similar to Levasseur and Cuoco (2003) who state that mathematical practices are developed throughout the mathematical problem-solving process, Lee and Tran (2015) suggest that statistical practices are developed throughout the statistical investigation process. One model for the statistical investigation process is Wild and Pfannkuch's (1999) statistical investigation cycle. Adapted from MacKay and Oldford's (1994) statistics course, Wild and Pfannkuch include five major stages in the process: Problem, Plan, Data, Analysis, and Conclusion (PPDAC). An overview is shown in Figure 2.6.

First, the *problem* phase is about gaining familiarity with the problem context and its importance. Since statistics is an interdisciplinary field, this part of the phase encourages teachers and students to bring in related knowledge that may be covered outside of the classroom (Cobb & Moore, 1997), such as popular media, news or other experiential knowledge about the problem context. Second, the *planning* phase includes collecting the data. This includes drafting statistical questions about the investigation, developing questions about the variables that may be collected and how they could be analyzed to address a problem, and considering how data will be collected and where the data will be stored. Third, the *data* phase consists of collecting, managing, and cleaning the data. In school settings, students are typically given clean, pseudo-real data or visuals (Makar & Fielding-Wells, 2011). Given the push for data science, it is important that students are given first hand experiences with big messy data (Gould et al., 2017). By “messy” data, I mean data that is missing entries, has duplicates, has incorrect or inaccurate data, has data is or may be incorrectly misformatted (e.g., categorical data as numbers), contains other forms of erroneous or misleading information.

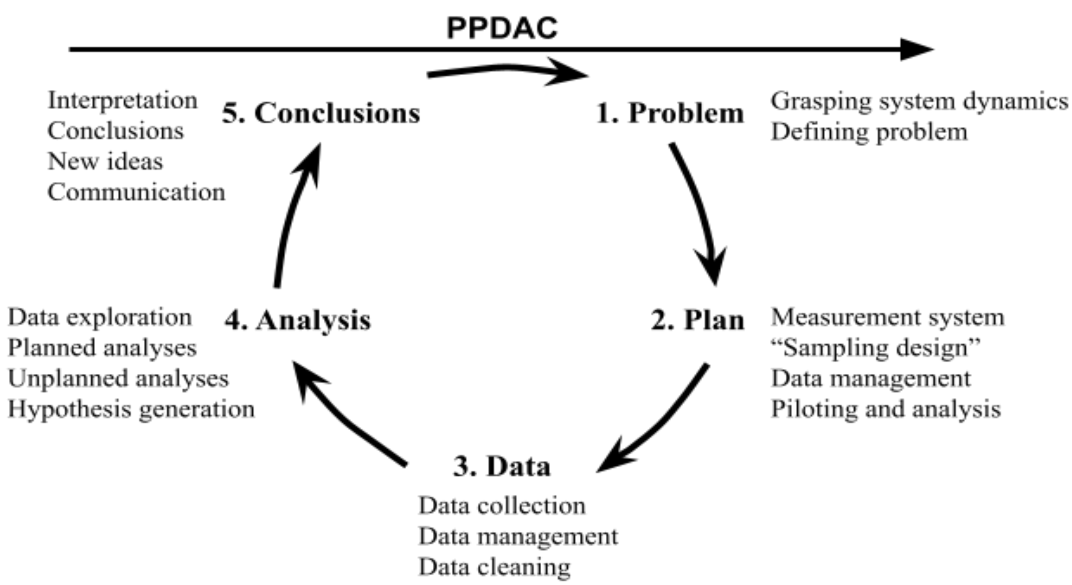


Figure 2.6: Statistical investigation cycle from Wild and Pfannkuch (1999, p. 226)

These first three phases (problem, plan, data) are what Visnovska and Cobb (2019) call the *data-generation discussion*. Visnovska and Cobb (2019) state that mainstream secondary statistics courses often do not include aspects of the data generation discussion, providing a students with limited view of what it is to engage in authentic statistical investigation (Chance, 2002). It is worth noting that statistics courses may not include the data-generation process because data is often provided to students. Nonetheless, even if the data sets are provided, students can engage in thought experiments about the problem context and how they would plan and collect data.

The fourth phase is the *analysis*. This phase primarily includes describing patterns in the data, for example through the exploratory data analysis, null significance hypothesis testing, and other planned or unplanned analysis that emerge during the process. Of the five phases, this is often considered the most mathematical part of the statistical investigation because it is deterministic (Visnovska & Cobb, 2019). Finally, Wild and Pfannkuch's (1999) statistical investigation model ends with the *conclusion* phase. This phase includes reflecting on the evidence from the analysis phase and, more importantly, relating the evidence back to the problem context.

Other statistical investigation models have emerged from or are similar to Wild and Pfannkuch's (1999) characterization of the PPDAC model. For example, GAISE I (2007) and more recently GAISE II (2020) use a four-step statistical problem-solving model: formulate statistical questions, collect data, analyze data, interpret data. Similarly, Gould et al. (2016) uses a data cycle model that includes: asking questions, considering data, analyzing data, and interpreting data. Nonetheless, all models place an emphasis on the statistical investigation cycle as an iterative, often non-linear problem-solving process that students may engage in as they are engaging in collecting and analyzing data (Chance, 2002; Franklin et al., 2007; Gould et al., 2017).

Furthermore, each phase may lend itself well to certain statistical practices. For example, the problem phase may prompt students to ask questions about the data, whereas the planning and data phase may prompt students to explore the relevance and sociopolitical nature of data. However, practices may also occur across all the phases.

Assessing Statistical Practices via the Statistical Investigation Cycle

In a study including nine mathematics teachers who were in their first year of teaching a high school Introduction to Data Science (IDS) course, Gould et al. (2017) documented how two groups of high school teachers moved through the data cycle process. Using a Markov chain, Gould and colleagues (2017) were able to track the sequential movement throughout the data cycle in a model-eliciting activity (MEA). They found that Group 1 (labeled as the more successful group) had less transitions between the phases, began by asking statistical questions then considering data, and spent more time interpreting the data within the context. On the other hand, Group 2 moved more transitions throughout the cycle and began by considering data before asking statistical questions. Both groups spent the majority of their time analyzing the data.

Although Gould et al. (2017) did not have an explicit focus on statistical practices, they did present data about teachers engaging in different forms of practices. For example, they note that an important part of assessing statistical questions was determining if teachers were considering the relevance of the data as it pertains to answering the task. Similarly, they note how teachers began by making histograms to look at distributions and use visualizations to identify characteristics in data, such as averages.

In another project with students who completed a second statistics course, Woodard and Lee (2021) documented how students used the R programming language to move through problem-solving phases during task-based interviews. They also documented the types of statistical computing actions that were taken within each phase. For example, as part of an implementing phase (carrying out a planned strategy), one of the students, Allison, made boxplots to compare multiple distributions. She stated "So looking at this, you can definitely tell that there is a difference in the average tip given between those (moves the cursor between the first and second group) and so that, I think it's worth looking at" (p. S151), then used an ANOVA test to confirm her hypothesis. One interpretation of this is that Allison was in the initial stages of the analysis phase, where she was generating hypotheses (asking questions) and using visuals to explore the data (transnumeration: forming data).

Summary of Literature Review

The TMSJ framework provides a model for designing courses with both statistical and critical literacies as the goal of the course. In doing so, students engage with praxis (a cyclical and continuous relationship between reflection and action) in their journey of statistical and data scientific critical consciousness. QuantCrit extends this framework by foregrounding the role of race and racism in the statistical investigation cycle for statistical and data scientific racial consciousness.

The research on habits of mind provides a framework to document how students engage with data, focusing on the intersection of statistical or data scientific and critical practices. Such practices include considering the relevance of data as it relates to the problem context (Burrill & Biehler, 2011; Lee & Tran, 2015), transnumeration (Burrill & Biehler, 2011; Lee et al., 2014; Pfannkuch & Rubick, 2002; Pfannkuch and Wild, 2004), accounting for and considering variability (Lee & Tran, 2015; Franklin et al., 2015) and its implications for generalizations (Franklin et al., 2007; Lee & Tran, 2015), remaining skeptical (Chance, 2002; Lee & Tran, 2015), and asking questions (Gould et al., 2017). Furthermore, critical perspectives also foreground the importance of interpreting and situating data-based arguments in a sociopolitical context (Gillborn et al., 2018; Weiland et al., 2017), especially to avoid reifying deficit narratives. These practices are developed and interwoven throughout the statistical investigation process (Lee & Tran, 2015).

What is needed in the literature is more research about how a social justice-oriented approach to teaching in content (mathematics and statistics) classes influences PSMTs' learning of both the statistical or data scientific and critical issues. Particularly, Thanheiser, Harper et al. (2020) recently stated that

> [l]ittle attention has been paid to the potential for TMfSJ [teaching mathematics for social justice] in mathematics content and methods courses impact PTs' [pre-service mathematics teachers'] mathematics learning, understanding of social issues, and mathematics identities as well as their teaching practices (p. 186).

That is, we need more evidence of the types of learning that occur in classrooms that use a social justice-oriented approach to teaching, including learning about the content, practices, and social issues. Furthermore, there are few empirical studies that document critical consciousness in mathematics classrooms, and fewer that center race and racism in statistics or data science classrooms (Stephan et al., 2021). Thus, in this dissertation, I use a teaching statistics for social justice framework to design a data science and statistics content course for PSMTs with the goals of developing PSMTs statistical learning and understanding of social issues in education on their journey of statistical and data scientific critical and racial consciousness.

Research Questions

I will be guided by the following research questions:

Research Question 1: Design Features

a. What design features support students' understandings of race and racism in the context of statistics and data science?

b. How were the design features enacted in the curriculum?

Research Question 2: Statistical and Data Scientific Content Knowledge

a. What was the effect of the teaching experiment (TE) on statistical content knowledge as measured by the student response patterns on curriculum-aligned assessments?
b. How did the response patterns by question type (e.g., conceptual or procedural, study design and regression) vary across the TE?

Research Question 3: Statistical and Data Scientific Practices

a. How do students' engagement with the statistical investigation cycle evolve through the course of a TE that uses a social justice-oriented approach to teaching statistics?
b. How do students' statistical practices evolve through the course of a TE that uses a social justice-oriented approach to teaching statistics?

Research Question 4: Focusing Phenomenon

a. How do elements of the TE contribute to the students' understanding of race and racism in the context of statistics and data science?

RQ1 focuses on the design features that were used to design a course during the Summer 2021 term. The course focuses on using statistics and data science for social justice. Topics included statistical questions, study design, and regression analysis. A goal of the course was to identify ways to engage with statistical and data scientific critical (and racial) consciousness. RQ2 focuses on the traditional statistical content knowledge as measured by a pre- and post-assessment. While acknowledging the biases in how standardized tests may be used or misused, this RQ is in alignment with the TSSJ framework's goals of "succeeding in a traditional sense" (Gutstein, 2006, p. 41). RQ3 focuses on statistical practices that emerged in the pre- and post-task-based interviews. Special attention will be given to practices that are rooted in QuantCrit's (Crawford et al., 2018; Garcia et al., 2018; Gillborn et al., 2018) goal of centering, examining, and transforming how data undergirds race, racism, and power. Finally, RQ4 focuses using a focusing phenomenon framework (Lobato et al., 2013) to coordinate how aspects of the classroom environment (e.g., design features, tasks, tools, and the teacher) directed students' attention towards understandings of race and racism in the context of data science.

Chapter 3: Methods

The goal of this project is to create a course that aims to develop pre-service mathematics teachers' statistical, data scientific, and critical literacies through a social justice-oriented statistics and data science curriculum. The main data source for this dissertation came from a teaching experiment (TE; Prediger et al., 2015) taught during the summer 2021 term. I designed and taught the TE. The TE included four main units: (a) introduction, (b) study design, (c) regression, and (d) course summary and project. Furthermore, summary of the research questions with the respective data and analysis is shown in Table 3.1.

The data collection sources included in- and out-of-classroom data, including assessments, interviews, and videos. Research followed the request of SDSU's IRB process. All data was stored in an encrypted external hard drive locked in a file cabinet at a home office. Participation in the research did not affect the students' course grade. I core component to this research was having another graduate researcher as a co-researcher during the data collection process. In particularly, the graduate student helped take field notes during most of the class sessions and met with me weekly to debrief classroom observations.

The remainder of this chapter describes the methods employed in this study. I begin with presenting details about design-based research and its implications for this study. I then discuss the setting and participants in this study. After, I discuss how the TE was developed, including the statistical content, social justice content, and lesson sequencing. I then discuss the data collection and analysis for each research question. I end by discussing my research experience and researcher positionality in this study.

Table 3.1: Research questions, data, and analysis

Research Question	Data	Analysis
Research Question 1: Design Features		
a. What design features support students' understandings of race racism in the context of statistics and data science?	**Design Features** a. Literature review about teaching statistics, mathematics for social justice, and critical consciousness in math	a and b. Implications for teaching data science for social justice with examples of how the design features were enacted
b. How were the design features enacted in the curriculum?	b. Classroom examples of how the design was implemented	
Research Question 2: Statistical and Data Scientific Content Knowledge		
a. What was the effect of the teaching experiment (TE) on statistical content knowledge as measured by the student response patterns on curriculum-aligned assessments? b. How did the response patterns by question type (e.g., conceptual or procedural, study design and regression) vary across the TE?	**Assessments** Pre- and post-assessments. Assessment will include sample tasks from LOCUS questions and performance tasks	a. Mean difference (e.g., t-test, effect size, Wilcoxon) and statistical matching within and across the course b. Separate data by groups: unit, procedural or conceptual. Similar analysis as part a
Research Question 3: Statistical and Data Scientific Practices		
a. How do students' engagement with the statistical investigation cycle evolve through the course of a TE that uses a social justice-oriented approach to teaching statistics? b. How do students' statistical practices evolve through the course of a TE that uses a social justice-oriented approach to teaching statistics?	**Task-Based Interviews** a and b. Pre- and post-task-based interviews.	a. Qualitative coded the phases of the PPDAC cycle and practices, allowing for prior research as long as it is not too restricting b. Compare the pre- and post-interviews, highlighting practices that emerged in the post-interview
Research Question 4: Focusing Phenomenon		
a. How do elements of the TE contribute to the students' understanding of race and racism in the context of statistics and data science?	**Classroom Data** a. Recordings of whole class conversations and homework assignments	a. Focusing phenomenon framework (Lobato et al., 2013) to coordinate how aspects of the classroom environment (e.g., design features, tasks, tools, and the teacher) directed students' attention towards understandings of race and racism in the context of data science.

Design-Based Research

To create the teaching experiment, I leveraged features of design-based research (DBR) with a focus on developing curriculum products (Prediger et al., 2015). DBR is an iterative research methodology where researchers purposefully design a learning environment, explore phenomena that emerge as a result of the design, and refine the design for future iterations (Prediger et al., 2015; Design-Based Research Collective, 2003). A benefit of using DBR is that DBR aims to bridge theory and practice by exploring authentic classroom ecologies and, in doing so, developing both theoretical and concrete products that may enhance learning experiences. Although the implementation of DBR varies, there are five main features of DBR: (a) it is interventionist, (b) theory informs practice, (c) it takes place in authentic learning environments, (d) it is iterative, and (e) it generates theory and concrete products (Prediger et al., 2015). Below, I unpack the features and make connections with the purpose of this dissertation and frameworks used in this dissertation.

Interventionist

A core feature of DBR is that the research employs new forms of instruction that are interventionist. For this dissertation, the intervention was the teaching experiment that used a social justice-oriented approach to teaching data science to PSMTs. Similar to DBR, TEs aim to approximate traditional classroom environments, but with intentional modifications to explore phenomena that emerge as a result of the instructional design (Prediger, 2015). The topics in the TE were selected to cover topics that are generally covered in first or second semester statistics courses (study design and simple linear regression). However, as mentioned above, the TE includes a social justice-oriented approach with a focus on the role of race and racism in the statistical and data scientific process.

Theory Informs Practice

Another feature of DBR is that theory prospectively informs the design of the intervention (Prediger et al., 2015). Furthermore, I was particularly guided by the implications of teaching mathematics for social justice (TMSJ; Gutstein, 2006), QuantCrit (Crawford et al., 2018; Garcia et al. 2018, 2018; Gillborn et al., 2018), and statistical and data scientific practices (Burill & Biehler, 2011; Chance, 2002; Gould et al., 2017; Lee & Tran, 2015; Lee et al., 2014; Pfannkuch & Rubick, 2002; Pfannkuch & Wild, 2004; Weiland et al., 2017) to develop students' statistical and critical literacies, described in Chapter 2.

In particular, the literature on teaching mathematics for social justice and related critical pedagogies guide the learning goals for the course (developing critical statistical and data scientific consciousness). QuantCrit extends TMSJ by foregrounding the role of race and racism in the statistical investigation and provides implications for how data can be used to advance social and racial justice. Finally, the research on habits of mind provides implications for the activities and lessons, where a learning goal was to enculturate students into a community where we develop and implement critical statistical and data scientific practices.

Iterative

A defining feature of DBR is that it relies on an iterative, cyclical process to refine a design (Cobb, Confrey, et al., 2003; Design-Based Research Collective, 2003; Prediger et al., 2015). This process was motivated by engineering design (Prediger et al., 2015) and suggests that researchers begin with a prototype of a design, experiment and test the design, analyze the impact, and refine the design for future iterations. Furthermore, Prediger et al. (2015) noted that small teaching experiments or interviews may be useful in the initial stages of the project or between iterations to explore student understandings. However, they are not required nor are they considered iterations by themselves.

As shown in Figure 3.1 (figure and phases adapted from the Zahner et al., 2021, Figure 2, p. 4), there are three phases in this dissertation. The first phase data was a pilot of the surveys and interviews. The primary goal of this phase was to rehearse conducting and analyzing the survey and interview data. The results from the first iteration informed survey and interview modifications for the second phase. The second phase included a four-lesson pilot of the TE as well as the pre- and post-surveys and interviews. The primary goal of this phase was to inform lesson features (e.g., preparing students to use technology and software used in the TE, modifying lessons), teaching practices (e.g., familiarity with teaching over Zoom, displaying student work), and research practices (e.g., recording and analyzing Zoom classroom data, analyzing and comparing pre- and post-surveys and interviews). Finally, the third phase was the main dissertation study, including the survey, interviews, and lessons in the TE. Details of the participants are presented later in the chapter.

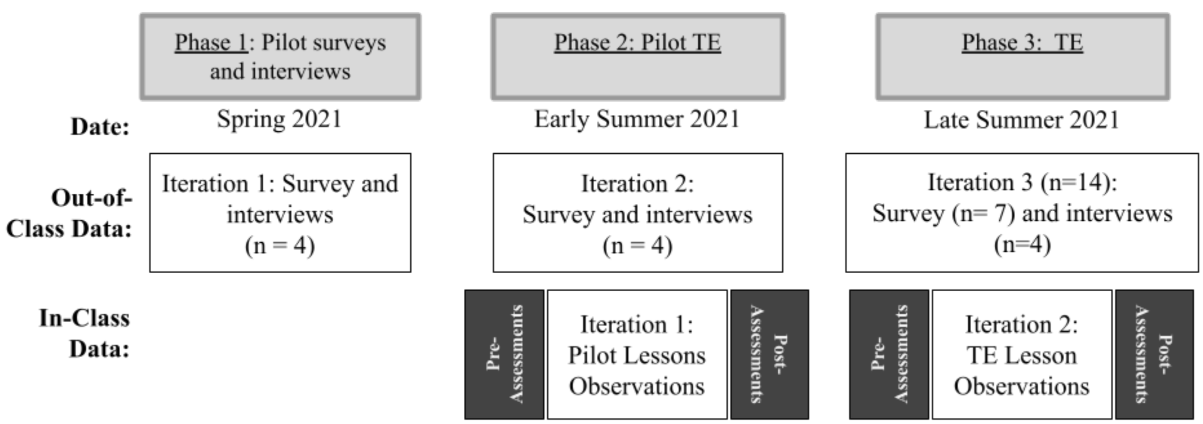


Figure 3.1: Iterations of the survey, interview, and TE lessons

Authentic Learning Environments

A fourth feature of DBR is that the research is situated in authentic learning environments or settings that approximate these environments (Design-Based Research Collective, 2003;

Prediger et al., 2015). This feature is related to the ecological validity of the intervention so that the design is reflective of and generalizable to authentic learning environments. There were two learning environments in this dissertation. The first is the Pilot TE that is part of Phase 2. Here, the environment was authentic in that it will be structured as three lessons of a larger class. Phase 3 extends this and situated the TE as part of a larger credit-bearing special topics mathematics course that is more reflective of an authentic learning environment.

Generates Theory and Products

A final feature of DBR is that it generates theory and products. In terms of theory, the goal is not to test theories but rather to develop and refine local theory by retrospectively reflecting on the implementation of the design (Prediger et al., 2015). The primary theory that will be generated in this dissertation was developing design principles for developing students' critical statistical and data scientific consciousness using a social justice-oriented approach to teaching (Research Question 1). In terms of products, the primary contribution from this dissertation included the TE lessons on study design and simple linear regression. Sample lesson plans, activities, and other course materials are shown throughout the results chapters.

Setting and Participants

This study took place at a four-year public Hispanic Serving Institution in the US-Mexico borderlands of Southern California. Pre-service mathematics teachers at this institution were required to take two one-semester statistics courses as part of their coursework for a major in mathematics in preparation for teaching. Data came from three settings: (a) pilot surveys and interviews from Phase 1, (b) pilot surveys, interviews, and pilot lessons from Phase 2, (c) the TE that was part of a special topics course from Phase 3. As mentioned above, a graduate student in mathematics education helped collect field notes and met with me weekly during Phase 3 to discuss classroom observations.

Phase 1: Pilot Surveys and Interviews

The first phase began with piloting the survey and interview protocol. As mentioned before, the goal of this phase was to rehearse collecting and analyzing data from the survey and task-based interviews. Six participants were recruited using a combination of Patton's (1990) convenience and maximum variation sampling. The sample was convenient in that all participants were in my social network. However, I also purposefully recruited participants with a wide range of statistics and professional experience, shown in Table 3.2. Four participants were available to complete the survey, and two participated in the interview (Yajaira and Susan). All pseudonyms were selected by the participant. Participants were not compensated for their participation.

Table 3.2: Phase 1 participant information

Pseudonym	Comfort with statistics	How many statistics classes have you taken?	Current Occupation
Jane	Fairly confident	4+	Data scientist at a large technology company
Yajaira	Not at all confident	1	Middle school math teacher
Alex	Slightly confident	2	Pre-service math teacher
Valentina	Slightly confident	1	Vice principal, former middle and high school math teacher

Phase 2: Pilot TE, Surveys, and Interviews

The second phase included four of the TE lessons that were piloted in addition to the pre- and post-surveys and interviews. The goal of this phase was to rehearse teaching the TE and analyzing data. The four lessons were taught online in the beginning of summer 2021. All lessons took place over Zoom, following the COVID-19 regulations in place at the time of the study.

Participants for the Pilot TE were recruited through snowball sampling. In particular, I asked participants from the Phase I study to recommend participants for the Phase II study. There were four participants in total. All four participants took the pre- and post-assessments, pre- and post-surveys, and attended all lessons (about eight hours in total). All four participants were pre- or in-service mathematics teachers in their first five years of teaching. Participants were compensated $100 each for their time.

Phase 3: Teaching Experiment

The TE included lessons from a special topics credit-bearing course taught in the 2021 summer term. A special study course was optimal for this study since it allowed for the use of experimental curriculum materials. Additionally, students who enrolled in the course were aware that the course is a "special topics" course, which may have helped foster buy-in from students. Following COVID-19 regulations, the special topics course was taught online from July 6th to August 13th.

The opportunity to enroll in the special topics course was shared with all students enrolled in the mathematics education major via email, but students outside of the mathematics education major (e.g., applied mathematics or mathematics majors) were also allowed to enroll. Following Patton's (1990) criterion sampling method, the selection criteria, from most important to least important, were students who: (a) have taken a first semester statistics course (prerequisite for the TE course), (b) were pre-service mathematics teachers, (c) were available for the TE over the 2021 summer term, (d) were interested in learning about teaching statistics for social justice, and (e) had not taken a second semester statistics course. This criterion was similar to the Phase 2 criteria for the exception of two differences. First, I did not make it mandatory that students attend all sessions. This allowed for students to miss class due to personal reasons (especially during the COVID pandemic). Second, I specifically recruited pre-service mathematics teachers instead of both pre- and in-service mathematics teachers since the goal of the study was to develop pre-service teachers' statistical and critical literacies.

There were 14 students enrolled in the course, 13 of which were pre-service mathematics teachers. I use the term "students" to refer to the 14 students in the class (not their students). The number of participants was ideal because it allowed for variety in group interactions and while keeping the data collection and analysis at a reasonable amount. Students were expected to attend all of the TE sessions, but they were not penalized for missing class. Students were not compensated for participating in the TE since the TE was part of a credit-bearing course. Students were also not compensated for the surveys since they were part of the course. Students were also asked if they wanted to select their pseudonym or have me select one for them.

All students were invited to participate in the pre- and post-interview. Students were compensated for the interviews since they occurred outside the TE. Students were compensated $50 for completing the pre- and post-task-based interviews ($25 per interview). Participation in the task-based interviews did not affect the students' course grades or their relationship with the university. Five students initially took the pre-interview, and four of those students also completed the post-interview. The four students that completed the pre- and post-interview were included in the analysis since it allowed for comparison across both interviews.

Instruction in the Teaching Experiment

TE Content

The TE was taught during the summer 2021 term. The design-based research model was used in designing lesson plans that aim to (a) foster statistical and data scientific knowledge about study design and regression, and (b) critical knowledge about the role of race and racism in the statistical investigation cycle. In what follows, I provide the rationale and details for the statistical content, social justice problem context, sequencing, and other considerations for designing the TE lessons.

Statistical and Data Scientific Content

There were two core units in the TE: (a) study design, and (b) regression. A unit on study design was chosen because study design, sampling, and biases are foundational topics in statistics and data science (e.g., designing A/B experiments). This topic affords opportunities to talk about the role of race and racism in statistical designs (surveys, experimental, observational) and how we can design observational or experimental studies to minimize those biases. This topic has direct ties to the Problem Plan Data Analysis Conclusion (PPDAC) statistical investigation cycle (Wild & Pfannkuch, 1999). For example, statisticians should consider the context when deciding which statistical design to use, data is situated within a larger sociopolitical context, over and undersampling affects the study's generalizability, and design biases may inform limitations and areas of future research for a study or populations.

Regression was chosen as the second unit of focus since it is a predictive model that lays the foundation for many predictive machine learning tools that are common to practicing data science (e.g., decision trees and random forests) and introduces the need for model diagnostics and comparison methods that are common in machine learning (e.g., cross validation, mean-squared error). This topic also has direct ties to the PPDAC cycle. For example, regression is one method of analysis, conclusions from the analysis should be situated within the problem context, and researchers must develop an understanding of the variables used and their relevance in a model.

Across the two units, the core statistical concepts are: variation, prediction, and statistical inference. Although all parts of the PPDAC were discussed across the entire unit, both of these topics also allow for focused conversations about the PPDAC cycle. For example, the study design unit may focus on the planning and data phases and the regression unit may focus on the analysis and conclusion. The problem context is involved across all phases (Chance, 2002; Franklin et al., 2005; Lee & Tran, 2015; Pfannkuch & Rubick, 2002; Pfannkuch & Wild, 2004; Visnovska & Cobb, 2019). Computer programming in R was also included across both units. For example, students learned how to code probabilistic sampling methods, upload and clean data, and use programming to derive the ordinary least squares regression line and perform regression analysis. The R programming language was selected because I am the most familiar with this language and because (from my experiences) it is one of the most common programming languages for practicing data scientists and statisticians (the other being Python). Finally, we discussed the role of race and racism within each statistical concept and phase of the PPDAC cycle. For example, we discussed how facial recognition data (used for filters on social media as well as predictive policing) are racially biased and produce biases in the algorithms.

Social Justice Problem Context

Central to Freire's work is *problem-posing pedagogy* in which a community poses a real-life problem that the community can collectively explore and transform through problem-solving (Berry et al., 2020; Freire, 1988; Gutstein, 2012). In this study, these *generative themes* were: (a) problem contexts that were of interest and related to the students' experiences with their world, (b) may have *generated* class discussions about the theme, and (c) may have helped *generate* discussions about the sociopolitical contexts of the theme to help advance social and racial justice.

In this study, I chose to focus on the "achievement gap" in educational outcomes as a theme. The theme was referenced across the entire course (including all four units). This term is often used to refer to the quantitative disparities between different groups of students (e.g., predicting a school-wide average SAT score given the proportion of Students of Color at a school). Other terms used to define similar themes include the "opportunity gap" which refers to the distribution of resources given to students as well as the "learning gap" which refers to disparities between what students were expected to learn (e.g., as stated by national standards) and what students learned (e.g., measured on standardized tests aligned to the national standards).

This theme was chosen for two main reasons. First, this is a topic that most students, especially pre-service teachers, may have some experience with, either personally or through exposure to media, news, or courses that discuss the topic. Second, analyzing achievement gap data provides a step into identifying the dominant axis of equity (achievement and access), but also generates opportunities to discuss the limitations of not accounting for critical perspectives of equity (Gutiérrez, 2017) and how data may reify deficit narratives if the sociopolitical context is not considered (Gillborn et al., 2018; Milner, 2010; Weiland, 2017). As Gutiérrez (2017) states,

> They [teachers] can see the benefits of using achievement data as a first step to identify who is not being served well by the school system, but they recognize the limitations of defining equity around such things as "closing the achievement gap." They understand that, more than just getting all kids to perform better or the

> same on tests of achievement, we should be invested in helping students become the kinds of people they want to be, fulfilling goals they have defined for themselves, which can mean different, not same outcomes (p. 21)

Thus, the generative theme around the achievement gap directly was used to guide conversations about the dominant axis of equity (achievement and success) as well as the critical axis of equity (identity and power). Furthermore, it was used as an entry point to guide conversations about how we can use data to identify and understand social injustices as well as how the data may guide action.

Design Sequencing

Table 3.3 provides an overview of the lesson sequencing and activities for the entire course. There are four main phases. The introduction unit was about laying the foundation for the course, including setting classroom norms with students, discussing the role of statistical practices in learning, and introducing critical perspectives. The second unit focused on statistical designs. The third unit was on regression and regression analysis. Finally, the last unit included project presentations. The second and third units were the core components of the TE.

Unit 1: Introduction

The first unit included the introduction to core ideas in the course: statistical investigation cycle, foundations in R, and critical perspectives to statistics and data science. The topics in this intro unit will be referenced throughout the unit. For example, the statistical investigation lesson (Class 3) was explicitly referenced throughout the course (e.g., how statistical designs relate to the planning and data collection phases of the PPDAC cycle, how regression relates to the data collection and conclusion phases of the PPDAC cycle). Furthermore, after reading about and discussing statistical practices (Class 3), students developed a class set of statistical practices. In the following lesson (Class 4), students read about QuantCrit (Crawford et al., 2018; Garcia et

Table 3.3: Overview of lesson sequencing and activities

Monday (3 hours)	Wednesday (3 hours)	Friday (3 hours)
	Class 1 **Unit**: Introduction **Part 1**: Introduction to the class, project, Google Colab **Part 2**: Introduction to statistics and data Science, survey, schedule interviews **Project**: Choose topic and groups	**Class 2** **Unit**: Introduction **Part 1**: Introduction to R programming, Google Colab **Part 2**: Representing data in R: Descriptive statistics, variable types, and data structures **Project**: Choose topic and groups
Class 3 **Unit**: Introduction **Part 1**: Introduce statistical investigation cycle **Part 2**: Introduce critical race theory (CRT) and quantitative critical race theory (QuantCrit) **Project**: Choose topic and groups, start drafting research questions	**Class 4** **Unit**: Study Design **Part 1**: QuantCrit and implications for data science and the statistical investigation cycle **Part 2**: Statistical questions and anti-deficit questions **Project**: Edit research questions, if needed	**Class 5** **Unit**: Study Design **Part 1**: Debrief reading, Surveys **Part 2**: Visualizations and cleaning data in R **Project**: Draft survey questions
Class 6 **Unit**: Study Design **Part 1**: Sampling and sampling bias **Part 2**: Randomization and sampling in R **Project**: Draft introduction. Identify appropriate sampling technique and sample population	**Class 7** **Unit**: Study Design **Part 1**: Types of Studies (Experimental studies and Observational studies) **Part 2**: Build connections between QuantCrit and Sampling **Project**: Review survey question feedback from professor. Make changes if needed. Positionality statement. Mentor text is provided	**Class 8** **Unit**: Study Design **Part 1**: Peer feedback and group check ins **Part 2**: Take home test **Project**: Review survey question feedback from peers. Make changes if needed. Identify appropriate survey tool

Table 3.3: Overview of lesson sequencing and activities, Continued

Monday (3 hours)	Wednesday (3 hours)	Friday (3 hours)
Class 9 **Unit**: Regression **Part 1**: Equity and Mathematics Education **Part 2**: Introduction to Regression **Project**: Draft methods. Start collecting data. Revisit rubric.	**Class 10** **Unit**: Regression **Part 1**: Introduction to regression and ordinary least squares (OLS) **Part 2**: OLS Part 2 **Project**: Collecting data	**Class 11** **Unit**: Regression **Part 1**: Comparing regression models in R part 1: R^2, correlation, mean squared error **Part 2**: Comparing regression models in R part 2: Cross validation **Project**: Collecting data
Class 12 **Unit**: Regression **Part 1**: Multiple Linear Regression and derivation **Part 2**: Multiple linear regression in R **Project**: Start preparing and analyzing data	**Class 13** **Unit**: Regression **Part 1**: Classification and regression trees (CART) **Part 2**: Connections with QuantCrit and the statistical investigation process **Project**: Analyzing data. Revisit rubric.	**Class 14** **Unit:** Project **Project:** Office Hours Check In (Sign up for time)
Class 15 **Unit**: Project **Project:** Office Hours Check In (Sign up for time)	**Class 16** **Unit**: Project Presentation	**Class 17** **Unit**: Project Presentation **Project:** Submit final paper

al., 2018; Gillborn et al., 2018) and coordinate the QuantCrit tenets and implications with the practices that the class developed (e.g., how using variables that analyze systemic structures may help combat deficit narratives as part of the data phase, how qualitative data and anecdotes can be used to support quantitative findings as part of the conclusion phase). These statistical practices served as a classroom-created artifact and will be referenced throughout the course.

Unit 2: Study Design

The second unit included ten lessons on study design. Three statistical design methods were included: surveys, experimental, and observational design. Across the unit, we also discussed biases that may arise, including sampling biases (Class 6), experimental or observational studies (Class 7), and in survey questions (Class 8). In terms of statistical programming, students learned about sample randomization techniques, reading data, cleaning data, and engaging in the exploratory data analysis process using the R programming language (Class 5 and 6). In terms of critical perspectives, students also had opportunities to build connections with QuantCrit (Class 7) and related them back to the classroom artifact of practices that was developed in previous lessons.

Unit 3: Regression

The third unit included lessons on ordinary least squares regression. This unit had a larger emphasis on computer programming than the previous two partially because I anticipated that students would feel more comfortable with the R programming language after a few lessons and activities. Students learned about regression, ordinary least squares (Class 10), and diagnostics by describing what occurs in the code and the output before being formally introduced to topics like correlation, R^2, and mean squared error (Class 11). This is similar to the *Predict-Check-Explain* (PCE) model that has been used with dynamically-linked representations (e.g., Schorr &

Goldin, 2008, Vahey et al., 2013) as well as the *Predict-Run-Investigate-Modify-Make* (PRIMM) that is used to teach computer science (e.g., Sentance & Waite, 2017). Central to PCE and PRIMM models is allowing students to learn by exploring the content and using code snippets before formalizing the content.

Furthermore, students used cross validation to assess and compare models. Cross validation is a model validation technique in which a portion of the data (training data) is used to build the model and the remainder of the data (testing data) is used to assess the predictive accuracy of the model (Class 11). Although cross validation is not typically used for regression, this approach was selected to introduce students to a method that is common in data science and machine learning. In terms of critical perspectives, the regression unit provided an opportunity to talk about the tensions between the dominant (access and achievement) and critical (identity and power) axes of equity described by Gutiérrez (2009). Particularly, students had the opportunity to talk about the importance of using achievement data to analyze educational inequities while also embracing a critical perspective to identify structures that lead and can mitigate inequities (Class 9 and 13).

Unit 4: Course Project

One way to help students engage with the entire statistical investigation process is through a course project (Burrill & Biehler, 2011; Chance, 1997, 2002). Statistics projects are especially useful since they allow students to collect, clean, and analyze their own data of their interest (Chance, 2002), instead of using a prescribed well-structured data. Thus, a core component of the course was a term project. The course project included an individual written report and a group presentation that were given on the last day of the course. Students worked in groups of two to three. Students were able to select their group or, otherwise, be assigned a group. Since there was an emphasis on racial justice throughout the course, students were encouraged to discuss the role of race and racism in their social justice project.

The project served two goals. First, the project was an opportunity for students to apply their knowledge to a topic of their choosing. Second, course projects or artifacts are a great way to raise awareness about social injustices (e.g., Kokka, 2020; Tate, 1995; Thanheiser, Rosencrans et al., 2020; Turner & Strawhun, 2005). This is in alignment with Gutstein's (2006) goal of *writing the world* that includes engaging in actions that combat social injustices, such as by raising awareness. Particularly, as part of the report and presentation, students were asked to raise awareness about a social justice inequity and discuss ways to challenge and dismantle those inequities.

Chance (1997) describes features of designing effective statistics projects. Three features and their implications for project and course design are discussed here: (a) integrating the project and course, (b) providing students with timely and constructive feedback, and (c) providing students with clear guidelines and expectations.

Integrate the Project and Course

First, the course project was well integrated into the course. Particularly, all main components of a project (e.g., motivation, methods, descriptive statistics, analysis, conclusion) were addressed as part of the class. The project then became an opportunity for students to apply what they learned in class to their project.

In the course project, planning, data collection, analysis, and conclusion (PPDAC) phases of the statistical investigation cycle were scaffolded throughout the TE. For example, students selected their group and a social justice topic of their choosing (Class 1 and 2). Students also created a survey to collect data (Class 5) and received feedback from their peers and instructor

(Classes 7 and 8). Similarly, students identified the appropriate sampling technique (Class 13). Students then learned how they could use regression to analyze the data, including how this data can be analyzed using a critical lens (Class 10). In terms of modeling, each step of the project was modeled using educational data.

Timely and Constructive Feedback

Providing timely and constructive aimed to promote dialogue between the instructor and students, as well as between students. This had two goals. First, students engaged in a form of Stronger and Clearer (Zwiers et al., 2017) and Jansen's (2020) rough draft mathematics in which students refined their ideas and revised their original thoughts and. For example, students had opportunities to work with and learn from others when they peer reviewed their first draft of the surveys. Instructor feedback was also given at multiple points of the term that target different parts of the PPDAC phases. For example, I gave feedback on their survey questions before they were shared with other students, their sampling process, and their planned analysis. Second, this may have helped students see that the statistical investigation process is an iterative and collaborative process.

Guidelines and Expectations

Finally, Chance (1997) recommended that students should be provided with clear expectations and guidelines of what is expected from the project. Similar to Chance (1997), this included a checklist and timeline in the syllabus with what is expected to be included in the final project report. I also provided sample reports that students could use as mentor text.

Sample Lessons

All lessons in the TE followed a Launch-Explore-Summarize teaching model (Lampert, 2001). The Launch portion of the lesson was designed to access prior knowledge and introduce a core concept that will be discussed in the class. The launch sometimes also included a *What's Going on In This Graph* activity that is inspired by the New York Times (New York Times, n.d.) where students were provided a data visual and asked to react to and explain the visual. The Explore portion of the lesson introduces students to the concept of the day. After a short introduction to the purpose of the lesson, students then engaged in a structured activity (e.g., computer programming in pairs, identifying sampling biases in reports) in breakout rooms before being formally introduced to the concept during a whole-class interaction. Finally, the summarize section included a whole-class discussion summarizing the learning goals of the day. This included a combination of student share-outs and lectures.

Furthermore, all instruction occurred through Google Collaboratory (Colab). Colab is a Jupyter notebook environment that allows students to create, edit, and share documents containing text and R code. Colab is free, cloud-based, and is similar to Google Docs in that it allows for multiple students to edit the document and add comments. This allowed students to collaborate across devices (MacBook Pro, Windows PC, or Google Chromebook). Furthermore, Colab allows for markdown text, HTML, and coding environments. Thus, rather than cycling through different windows (e.g., PowerPoint, R Studio, class handouts), all course instruction and materials for a particular lesson were embedded in the Colab document.

The lessons in the TE draw from four main resources: (a) lessons I created when I was teaching high school, (b) the MODULE[S2] curriculum (Casey et al., 2019), (c) the College Preparatory Mathematics (CPM) statistics curriculum (Griswold et al., 2018), and (d) the Introduction to Data Science Curriculum (IDS; Gould et al., 2016). The MODULE[S2] curriculum was especially useful since it has an explicit focus on developing students' statistical knowledge for teaching. Similar to Groth (2014), the MODULE[S2] provides opportunities for students to analyze and react to student sample responses. Although the CPM curriculum is intended for high school students, the CPM curriculum provided the foundation for the sequencing, lesson notes, and some classroom activities. Finally, the IDS curriculum was also intended for high school students, but has an explicit focus on data science. The IDS curriculum was especially useful for providing a foundation of how to introduce and use the R programming language in introductory courses.

Below, I illustrate and describe the rationale for the initial design of two lessons. The first lesson (Sampling Bias) provided opportunities for students to learn about different sampling biases (convenience, voluntary, under coverage, and nonresponse) by reading sample reports. This lesson provides a more explicit social justice focus. The second lesson (Sampling Methods) covers non-probabilistic and probabilistic sampling methods and introduces the R programming language. Combined, these two lessons introduced students to different sampling methods they can use in their course project and may mitigate sampling biases that may emerge in their study. It is worth noting that there were changes to these lessons, described in the results chapters.

Lesson 5: Sampling Bias

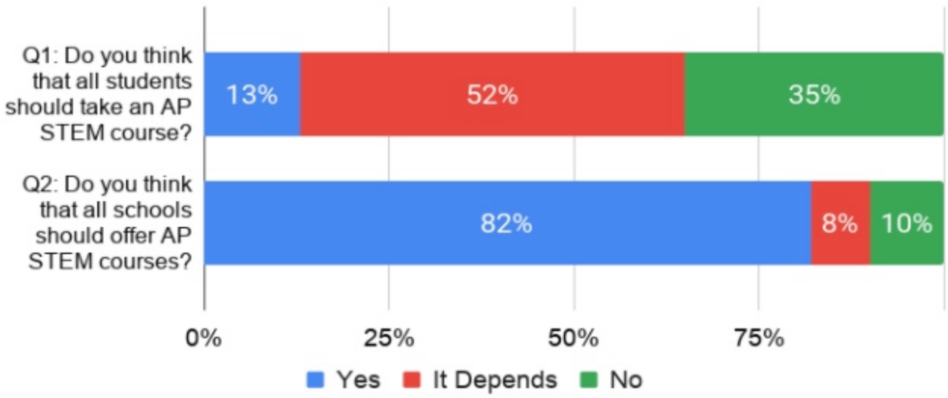
The Sampling Bias lesson begins with the Launch activity shown in Figure 3.2. Students are expected to recognize how a question about a similar topic can lead to two different conclusions depending on the wording. Further, students should recognize that one student is about all students *taking* Advanced Placement (AP) classes whereas the second is about *giving access* to AP classes.

What's going on in this graph?

Ms. Torres is chair of the math department at Corazón High School. She is interested in offering more AP courses. She asks students, parents, and local community members the following questions:

1. Do you think that all students should take an AP STEM course?
2. Do you think that all schools should offer AP STEM courses?

After collecting data, she made the visual below.



Press play below to submit your answers via a Google Form.

Launch Questions:

1. Who is the population of interest?
2. How are the implications of each question different?
3. How does each question analyze systemic structures?
4. Write a catchy headline that captures the graph's main idea.

Figure 3.2: Launch activity in the sampling bias lesson

The four prompts for the launch activity are shown at the bottom under "Launch Questions." The second is related to the previous lesson on survey design, questions, and biases (Class 4) and the third question is related to the lesson about QuantCrit (Class 3). The fourth question is a low-stakes question that allows students to use their creativity to create a name for the graph's main idea. The first question relates to the topic of this lesson: sampling. All questions were answered in a Google Form that is embedded in the Colab document. This Google Form document were then be referenced to *Collect and Display* (Zwiers et al., 2017) student responses.

In the Explore phase of the lesson, students were introduced to the topic of the lesson. There was an explicit connection to the PPDAC statistical investigation cycle, particularly with how sampling relates to the planning and data phases. From there, students discussed reports that they were asked to read for homework. Particularly, each participant was asked to read at least one of the following reports: (a) industrial pollution in Barrio Logan (environmental justice), (b) gentrification in San Diego (civil justice), (c) ethnic studies in the K12 curriculum (educational justice), or (d) effects of racial and gender stereotype threat (racial and gender justice). These reports were motivated by real social justice studies, but were intentionally modified to insert sampling bias in their survey designs. Across the four studies, the types of biases were convenience sampling bias, voluntary response bias, undercoverage bias, and nonresponse bias. Three of the studies used a survey design (environmental, civil, and educational justice) and one study used an experimental design (racial and gender justice). Students were placed in groups with other students who read the same report as them, then asked to create a Google Slide addressing the questions in Figure 3.3.

For the last homework, you selected and read a sample project. Today, you will be in a group with other people who read the same project. In groups, you are going to idenitfy the following for this project:

1. What is the project about?
2. Population of interest
3. Population sample
4. Sampling technique
5. Possible sampling biases
6. Connections with QuantCrit

Prepare a 2-3 minute presentation for the class using Google Slides.

Figure 3.3: Explore activity in the sampling bias lesson

The main purpose of this activity is for students to identify different types of sampling bias before being formally introduced to sources of sampling bias. However, students will also learn about different social justice topics, preview sampling methods that will be referenced in the next lesson, and build connections with QuantCrit discussed earlier in the course.

The Summary portion at the end of the lesson begins with student presentations and responses to the questions in Figure 3.3. Throughout the presentation, I made explicit connections to the four sources of sampling bias and how the biases may be avoided or reduced. This will be followed by class discussion about connections with QuantCrit. Anticipated responses include the implications of representation in data sampling, the relevance of the problem context and community when designing statistical studies, and generalizations of implications given the sample. These student generated connections will then be added to the class set of statistical practices that was created in Lesson 2.

Lesson 6: Sampling Methods

Lesson 6 introduces the R programming language, expands on the non-probabilistic sampling techniques (convenience, voluntary response, purposeful, and snowball) and introduces four probabilistic (simple random, systematic, stratified, and cluster) sampling methods. This is the first time that we will use R programming as a class, but students will have previewed R as homework. To allow for time to review R, the lesson skips the Launch activity and begins by reviewing the basics of R: coding R in Colab, arithmetic functions, creating and manipulating variables, reading data, finding descriptive statistics of data and data frames, and troubleshooting. This portion of the lesson will be guided by the instructor, but students will be able to follow along on their own devices.

The Explore begins with a short introduction to four non-probabilistic sampling methods. These methods were informally introduced in the reports that were discussed in the previous lesson, but formally identified as sampling methods in this lesson. Similar to the beginning of the lesson, students will then use R to learn about and program four probabilistic methods. Throughout this portion of the Explore phase, students will be asked to predict what they think will happen in a code, then run the code, and then evaluate their prediction (Predict-Check-Evaluate) to learn about probabilistic sampling methods. For example, students are presented with the code in Figure 3.4.

```r
%%R
set.seed(1) #reproducibility

####Data with students who prefer converse
converse = studentData[studentData$Shoes == "converse",] #Only rows with converse
converseRows = sample(1:nrow(converse), size = 3) #Randomly sample three students
converse[converseRows,] #Return randomly sampled rows from line 6

####Data with students who prefer converse. Similar as above
vans = studentData[studentData$Shoes == "vans",]
vansRows = sample(1:nrow(vans), size = 3)
vans[vansRows,]

###Combine Data
rbind(converse[converseRows,], vans[vansRows,])

```

Figure 3.4: Explore activity in Lesson 6

It is noted that the code is commented (in green) for the purpose of this proposal, but the comments were not be shown to students since their task was to predict the code. This code shows an example of stratified sampling, where three students who prefer Converse and three students who prefer Vans are selected from pseudo data, returning a stratified sample with six participants. In reading the code and predicting the output, students are able to learn about how stratified sampling works before formally being introduced to a function in R that accomplished the same task (called stratified) and before being formally introduced to stratified sampling. Similar Predict-Check-Evaluate activities are embedded throughout the entire TE.

The lesson ends with an application of the topics from the last two lessons (sampling biases and sampling methods). An example is shown in Figure 3.5. This question was adapted from Casey et al. (2019).

Let's say you are teaching about sampling methods in your class. Your class wants to determine how the students at their school feel about the cafeteria food. The class decides to take a sample to find out. Your students start offering suggestions for acquiring the sample. Some of their responses are below.

Student	Suggestion
Javier	I think we should just stand at the front of the lunch line at the start of lunch and ask the first students in line.
Demaria	I can make a poster to hang in the hallway and ask for students to sign up to give us their opinion.
Josh	Why can't we just use our class as our sample?

a. How would you respond to Demaria's suggestion?
b. How would you respond to Josh's suggestion?
c. What are some things you would ask the class to consider when determining a sampling method?
d. What are some potential drawbacks to using a convenience sample?

Figure 3.5: Hypothetical teaching scenarios Lesson 6

Analysis

RQ1 (Design Features) Data Collection and Analysis

The first research set of questions presents and illustrates the design features used to design a data science for social justice course. The research questions are:

Research Question 1: Design Features

a. What design features support students' understandings of race racism in the context of statistics and data science?
b. How were the design features enacted in the curriculum?

This research question is in alignment with DBRs feature that theory informs practice. I particularly draw on literature about teaching mathematics for social justice (TMSJ; Gutstein, 2006), Freire's (1988) critical consciousness and praxis, QuantCrit (Crawford et al., 2018; Garcia et al. 2018, 2018; Gillborn et al., 2018), and statistical practices (Burill & Biehler, 2011; Chance, 2002; Gould et al., 2017; Lee & Tran, 2015; Lee et al., 2014; Pfannkuch & Rubick, 2002;

Pfannkuch & Wild, 2004; Weiland et al., 2017) to create design features about considering the role of race and racism in data science courses. As a result, this research question is framed more as a literature review than a traditional qualitative analysis (e.g., qualitative coding). This approach is similar to the design principles presented by Zahner, Calleros, and Pelaez (2021).

I draw on various lessons to define and illustrate how the design features were enacted in the TE. For each lesson, I present the lesson goals, activities, and relevant student work. I also provide potential modifications for future iterations of the lessons.

RQ2 (Statistical and Data Scientific Content Knowledge) Data Collection and Analysis

The second research question aims to measure traditional statistical content knowledge in the TE courses. Particularly, the research questions are:

> **Research Question 2:** Statistical and Data Scientific Content Knowledge
> > a. What was the effect of the teaching experiment (TE) on statistical content knowledge as measured by the student response patterns on curriculum-aligned assessments?
> > b. How did the response patterns by question type (e.g., conceptual or procedural, study design and regression) vary across the TE?

Data includes pre- and post-assessments that were administered in the beginning and end of the TE course. While acknowledging the biases of standardized tests, this RQ is in alignment with the TSSJ framework's goals of "succeeding in a traditional sense" (Gutstein, 2006, p. 41) and aims to provide an insight into the development of students' subject matter knowledge (Ball et al. 2008; Hill et al., 2008).

Data Collection

All students took a pre- and post-assessment about the topics of interest at the beginning and end of the respective units. Questions came from the Comprehensive Assessment of Outcomes in Statistics (CAOS) and the Levels of Conceptual Understanding in Statistics (LOCUS) assessment. The CAOS assessment is the nation's first standardized assessment of student understanding in statistics courses (Delmas et al., 2007). The LOCUS assessment explicitly focuses on conceptually-aligned problems that are embedded within the processes of statistical problem solving (Jacobbe et al., 2014). Combined, the CAOS and LOCUS assessments provide a measure to compare learning in the TE to the STAT. Samples are shown in Appendix 1 and 2.

In terms of reliability, Del Mas et al. (2007) conducted a reliability assessment using CAOS data from 2005 and 2006. The inclusion criteria included students who completed the entire assessment between ten and 60 minutes and students that did not take an Advanced Placement Statistics course. This resulted in 1470 students across 36 statistics courses and 33 institutions. Fifty-seven percent of the students were female and 74% of which were identified as White. Overall, the CAOS assessment has an acceptable level of reliability as suggested by an analysis of internal consistency that produced a Cronbach's alpha coefficient of 0.82 (DelMas et al., 2007), where acceptance lower limits range from 0.5 to 0.7 (Pedhazur & Schmelkin, 1991).

Whitaker et al. (2015) conducted a reliability assessment in the LOCUS pilot from 2013. The inclusion criteria included secondary schools that were close to a LOCUS contact person, schools that were identified as high-performance as measured by standardized assessments, and schools that offered statistics courses prior to the Common Core State Standards. This resulted in 3324 students, 2075 of which were majority middle school students (97%) that took a beginner / intermediate level assessment and 1249 that were all high school students and took an intermediate / advanced level assessment. Forty-six percent of the total sample were female. Overall, the LOCUS assessment had an acceptable level of reliability as suggested by a classical test theory (Crocker & Algina, 1986) that produced a stratified alpha between 0.70 and 0.72 for each of the beginning / intermediate assessments and 0.87 for intermediate / advanced assessments. It is important to note that the inclusion criteria may result in a bias reliability measure that is based on a sample of high-performing secondary schools. Nonetheless, they provide an insight into the reliability of using the LOCUS assessment.

It is also important to note that the LOCUS assessment was designed for high school students' first statistics course. However, the content is similar to what may typically be presented in first- and second-semester statistics courses.

Data Analysis

There were two parts to this analysis: (a) overall assessment gains and patterns within the course, (b) assessment gains by question type (e.g., conceptual or procedural, study design and regression). For both the free-response questions, I also include examples of the students' responses. All questions were graded using the provided rubrics. Furthermore, only students with a pre- and post-assessment will be included in the analysis. All quantitative analyses will be conducted in R (R Core Team, 2017).

It is worth noting that findings from this research question should be interpreted with caution since there is a small sample size. Future research could consider different study designs (e.g., comparing treatment and control groups), larger and random national samples, or other modifications and analyses (e.g., propensity score matching). Furthermore, it is important to recognize that there may exist some grading bias since I graded the assessments at the end of the term and expected higher results. Further research (i.e., double coding) may help strengthen this analysis.

The first analysis begins by comparing pairwise pre-post assessments gains within the course. In an effort to provide various measures of growth, I will use three pairwise statistical tests: (a) paired sample t-test, (b) paired Wilcoxon signed-rank test, (c) paired effect sizes

(Cohen, 2013; Hedges, 1981), and (d) single-student normalized gain (Chance et al., 2016; Hake, 2002). The purpose of providing multiple analyses that aim to measure a similar phenomenon is to provide multiple sources of evidence. Appropriate statistical corrections will be taken to account for the small sample size and model assumptions. Data will also be parsed into different subsets. This included parsing data by unit, by conceptual or procedural focus, and individually by question.

I used pairwise statistical tests since I will have pre- and post-assessment scores for students. Students who do not have both a pre- and post-assessment were excluded from the data. The paired sample t-test is a common parametric hypothesis test that determines whether the average difference in repeated measures is statistically different from zero. Paired t-tests require that the pre- and post-assessments are normally distributed. A paired Wilcoxon signed-rank test is the nonparametric counterpart to the paired t-test. The paired Wilcoxon signed-rank test determines whether the median difference in repeated measures is statistically different from zero and requires that the pre- and post-assessments are ordinal. Cohen's (2013) effect size is similar to the paired t-test and Wilcoxon signed-rank test, but aims to measure the magnitude of the difference. Hedges (1981) effect size is similar, but uses a pooled weighted standard deviation and performs better when the sample size is under 20. Both Cohen and Hedges' effect size can be calculated using the 'effsize' package in R (Torchiano, 2020). Finally, the single-student normalized gain (Hake, 1998, Chance et al., 2016) measures the percent gain from pre- to post-assessment, but accounts for the potential gain. For example, there may be a ceiling effect where students who had a 70% in the pre-assessment have more room for gains than a student who scored a 90%. All statistical assumptions were interrogated before the analysis.

RQ3 (Statistical and Data Scientific Practices) Data Collection and Analysis

The third question focuses on teachers' practice as evident in pre- and post-task-based interviews. Particularly, the research question is:

> **Research Question 3:** Statistical and Data Scientific Practices
>
> a. How do students' engagement with the statistical investigation cycle evolve through the course of a TE that uses a social justice-oriented approach to teaching statistics?
>
> b. How do students' statistical practices evolve through the course of a TE that uses a social justice-oriented approach to teaching statistics?

The goal of this analysis is to highlight the intersection between critical and statistical practices. That is, how are students putting QuantCrit tenets into practice? What evidence is there of students situating conclusions in a larger sociopolitical context? How do students consider the role of race and racism in the sampling process?

Data Collection

Task-Based Interviews. Students from the course were asked to participate in 60-minute task-based interviews about the selected topics before and after the course. An email was sent to students a week before the course began inviting them to participate in a pre-interview. Pre-interviews took place on or before the first day of the course. The same students were invited to participate in a post-interview at the end of the course. As mentioned previously in this chapter, participation in the interviews did not affect the students' grade or relationship with the university.

Similar to Gould et al. (2017), each interview contained four main parts: (a) read an article, (b) answer follow-up questions and thought experiment, (c) do a task, and (d) respond to follow-up questions. The first part asks participants to read an article about the problem context. Both contexts will be related to measuring student learning and educational equity. Then, participants are asked follow-up questions about the problem context as well as a thought experiment that asks participants to design a hypothetical study. This is similar to how Visnovska and Cobb (2019) used thought experiments to ask teachers to discuss how they would design a study. In doing so, the thought experiment provides an insight into the problem and planning phases of the PPDAC cycle. The third part contains a task where participants are asked to analyze data. This part of the phase will provide an insight into how participants engage with the data, analysis, and conclusion phases of the PPDAC cycle. The interview ends with the interviewer asking follow-up questions about the participants activities during the task (e.g., asking for clarification on what they chose to do). Samples of the interview protocol shown in Appendix 3.

Data processing for the task-based interviews included contact summary forms, descriptive accounts, and transcribing prior to coding. All steps after the task-based interview are forms of analysis and data reduction. First, after each interview, I wrote a contact summary form (Miles et al., 2020) describing important moments in the interview. Contact summary forms were especially useful since I could not simultaneously interview and take detailed notes during the interview. The contact summary forms were also used in the analysis, mainly to plan for the post-interview, suggest new or modify qualitative codes, and guided the analysis of video recordings. Second, after all interviews were collected and the TE ended, I watched the interviews and wrote descriptive accounts that include more detailed descriptions and comments by episode. Since the data was analyzed after the course was over (as stated in the IRB), contact summary forms informed which particular events were detailed in the descriptive accounts. Select excerpts were transcribed using the InqScribe software.

Data Analysis

Since all interviews and lessons were taught over Zoom, data was partially transcribed using the Zoom software and edited as needed. This significantly reduced the data processing time. Data was coded using MaxQDA after the data collection process was done and after the course grades were submitted. I created a contact summary form (Appendix 4; Miles & Huberman, 1994) after each interview describing the practices that emerged during the interview. These contact summary forms were the beginning of the analysis and served as a reminder of what occurred in the class and helped identify key moments to look at when beginning to transcribe, code, and analyze data.

Qualitative Coding. The interviews were analyzed using elements of Miles et al.'s (2020) tactics for generating meaning in qualitative data and their implications for coding. A summary of the coding analysis is shown in Figure 3.6.

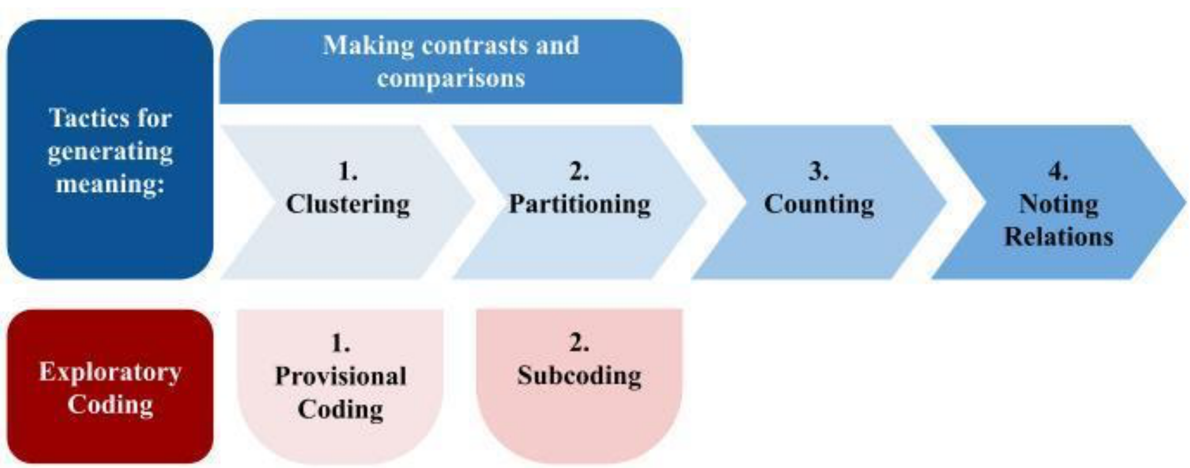


Figure 3.6: Tactics for generating meaning from qualitative data and exploratory coding analysis used in this dissertation (Miles et al., 2020)

The process began by clustering data. Miles et al. (2020) describes clustering as "trying to understand a phenomenon better by *grouping* and then *conceptualizing* objects that have similar patterns and characteristics" (p. 276, emphasis in the original text). This process does not have to be self-invented (Miles et al., 2020). In this dissertation, this is implemented using *provisional coding* (Miles et al., 2020) to create cluster codes. Provisional coding began with a priori codes, or deductive codes that are guided by a theoretical framework, existing literature, or prior research and experiences (Miles & Huberman, 1994). Provisional coding was ideal for two reasons. First, provisional codes allowed me to build on previous research (e.g., statistical practices). In doing so, a priori codes provided a foundation for clustering the phenomena of interest into pre-identified categories that were motivated by the literature. This is in alignment with DBR's core feature that theory prospectively informs practice. Second, given the novelty of this study, provisional codes may be revised, modified, or expanded to include new codes (Miles et al., 2020). Thus, the use of codes is not restricted to a priori codes.

Next, I partitioned the clusters into smaller, more detailed codes. Miles et al. (2020) describes partitioning as the process of subdividing data to avoid data blurring. The goal was thus to identify smaller clusters of data that are all related under a broad theme but have differentiated features. In this dissertation, this was enacted by *subcoding*. Subcodes were assigned after an initial primary code (the provisional codes). Similar to the cluster codes, I also allowed for a priori codes that were stated in or deduced from the literature. Thus, after all the cluster codes are created, I created subcodes that add more detail to the types of phenomena in each cluster.

Central to clustering and partitioning data was making contrasts and comparisons. To enact this tactic, I drew on the *constant comparison* process of grounded theory (Corbin & Strauss, 1990; Strauss & Corbin, 1994). The constant comparison is an iterative and reflective process where we create conjectures about the data, compare them to other parts of the data, and modify categories as needed (Corbin & Strauss, 1990; Strauss & Corbin, 1994). This helped further identify categories and codes, their characteristics, and relationships with each other.

Further, the constant comparison process brought awareness to some of the emerging themes and may have helped guard against biases that may emerge from being restricted by fixed or given categories.

The final two phases included counting data and noting relations. For this dissertation, this included creating descriptive statistics and visualizations of the cluster and subcodes. Miles et al. (2020) notes that counting data (e.g., by descriptive statistics or visualizations) is useful in qualitative analysis since it easily views distributions, which may in turn help noting relations (e.g., differences in distributions).

Coding Practices. I leverage the coding analysis model shown in Figure 3.6 to code students' statistical practices in pre- and post-task-based interviews: (a) applied cluster codes to describe statistical practices, (b) applied subcodes to further define and differentiate practices within each cluster code, (c) applied cluster codes describing different phases of the statistical investigation cycle, and (d) counted and noted relations in the data.

First, I began with the cluster codes shown in Table 3.4. These codes were provisional in that they were created using the literature described in Chapter 2. This was applied to all pre-task-based interviews. Drawing on the constant comparison process (Corbin & Strauss, 1990; Strauss & Corbin, 1994) I used the codes in Table 3.4 as a starting point but remained open to coding new practices that emerged from the data. This process was similar to how Woodard and Lee's (2021) approach to coding interviews of students engaging with statistical computing tasks where they began with a set of codes describing how students engaged with computational tasks, coded what the participant was doing, and allowed for new codes as needed. Next, I subcoded each of the cluster codes, also allowing for new codes to emerge.

Table 3.4: A priori codes for statistical practices with definitions and examples

Cluster	Code	Definition	Examples
Transnumeration			
	Forming data	Creating meaningful representations of the data (Burrill & Biehler, 2011; Lee & Tran, 2015; Pfannkuch & Rubick, 2002; Pfannkuch and Wild, 2004)	Visualizing statistical measures, identifying subsets of data, and examining data structures
	Changing data	Altering visuals to enhance meaning (Lee et al., 2014; Pfannkuch & Rubick, 2002; Pfannkuch and Wild, 2004)	Adding augmentations, highlighting certain patterns and structures data, and using different visualizations to communicate different meanings
Variability			
	Anticipating and Looking for Variation	Describing variability within a group, variability within and across groups, covariability, and variability in model fitting (Franklin et al., 2007)	Calculating the standard deviations, difference in the averages between two groups, linear relationships, measures of fit
	Generalizability	Considering the generalizability of data (Franklin et al., 2007; Lee & Tran, 2015)	Identifying constraints of study, considering how measurement tells will reproduce similar results,
Interpretations			
	Relevance of Data	Considering how well data measures an attribute of interest in a statistical task (Burill & Biehler, 2011; Chance, 2002; Lee & Tran, 2015; Visnovska & Cobb, 2019)	Determining if data is well suited to answer research questions, collecting new data to address the research question
	Sociopolitical Nature of Data	Considering the sociopolitical nature of data (Gillborn et al., 2018; Weiland et al., 2017)	Evaluating the source, collecting, and reporting of data, consider how they are shaped by a sociopolitical context
Implications			
	Implications of Data	Considering the problem context when providing data-based conclusions (Chance, 2002; Franklin et al., 2005; Lee & Tran, 2015; Pfannkuch and Wild; 2002; Visnovska & Cobb, 2019)	Referencing the problem context when describing associations
	Sociopolitical Implications of Data	Considering the sociopolitical context when providing data-based arguments (Gillborn et al., 2018; Weiland, 2017)	Explicitly focusing on avoiding reifying or reinforcing inequities, stating tensions

Table 3.5: Code definitions and examples for the phases of the PPDAC cycle

Code	Definition	Examples
Problem	Identifying statistical questions and the context	Identifying the statistical question (i.e. question about variation) to be addressed, identifying the systems or structures at hand
Plan / Data	Considering how data is collected, defined, stored, and cleaned	Identifying what data will be needed to address the question of interest, what tool (e.g., survey) or procedures (e.g., sampling, randomization) to be used, addressing missing data, data formatting and storage, planning the analysis
Analysis	Identifying patterns related to the question, related to the research question at hand	Identify patterns in the data that are directly tied to the question of interest, enacting the analysis, generating hypotheses. May include interpreting findings that are not directly related back to the problem context (e.g., interpreting the correlation but not identifying three schools)
Conclusion	Summarizing, communicating, and relating findings back to the problem context	Relating the findings back to the original question (identifying three schools to visit). Preparing conclusions and presenting information to others. May include follow-up questions

Coding the PPDAC phases. After all practices were coded, I also applied cluster codes describing episodes in students' interviews using Wild and Pfannkuch's (1999) problem, planning, data, analysis, and conclusion (PPDAC) phases of the statistical investigation cycle. The codes, definitions, and examples are shown in Table 3.5. As mentioned before, these codes were provisional in that I will be guided by Wild and Pfannkuch's (1999) PPDAC model. However, I remained open to the possibility of adding or merging different phases. This is similar to how Woodard and Lee (2021) organized students' work in terms of their identified problem-solving phases and how Gould et al. (2017) coded teachers' interactions with the data cycle model.

Counting and Noting. I counted and noted relations after all the practices and phases of the PPDAC cycle are coded in the transcript. In particular, I created a visual showing the proportion of time that students engaged with each phase of the PPDAC cycle with an additional layer describing the statistical practices that students engaged with within phases of the PPDAC cycle. This was similar to Gould et al.'s (2017) visualization of the proportion of time that students spend in each of the data cycle phases and Woodard and Lee's (2021) visualizations of the proportion of time that students spend in each of the problem-solving phases with statistical computing actions overlayed, shown in Figure 3.7. These visualizations helped "shed light on larger-scale patterns of behavior and help understand how the investigation evolves over time" (Gould et al., 2017, p. 327). All visuals had accompanying tables displaying frequency and relative frequency counts.

The analysis for this research question ended by comparing the practices from the pre- and post-interviews, paying special attention to how students engage with critical statistical practices. Similar to Gould et al. (2017) and Woodard and Lee (2021), I compared quantitative and qualitative features of the interviews. Quantitative features included the amount of time that students spend on different phases of the PPDAC cycle, frequency counts of the practices that emerged in the interviews, and identifying any shifts in the practices from the pre- to the post-interviews. This part of the analysis also helped guide which interactions to look for in the fourth research question (How do elements of the TE contribute to the students' understanding of race and racism in the context of statistics and data science?).

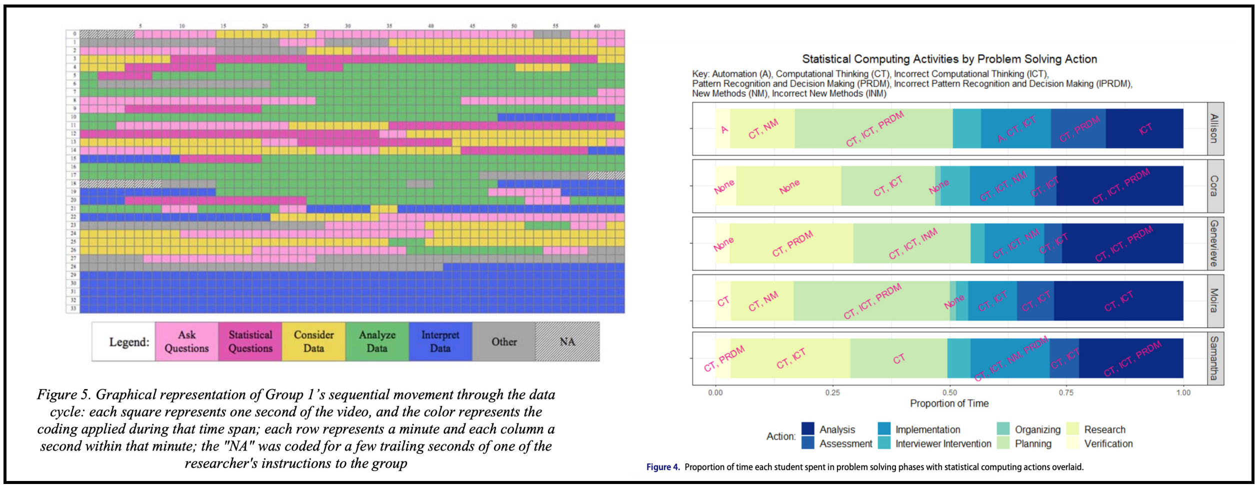


Figure 3.7: Sample visualizations for student activities from Gould et al. (left; 2017, p. 317) and Woodard and Lee (right; 2021, p. S150)

Coder Agreement

Double-coding in qualitative research is helpful because multiple codes may be able to contribute different interpretations to the data and expand our understanding of the data. However, I was not able to double-code the data presented in this analysis. Instead, I present how I could double-code in a follow-up analysis. Given there may be up to eight hours of interviews (four one-hour pre-interviews and four one-hour post-interviews), I could double code all the interviews. I will be the primary coder for all interviews. The second coder could be a graduate student in a Mathematics and Science Education doctoral program. I would aim to have the second code be the same across all double-coded videos, but will allow for up to two double coders (up to two hours each) given time and availability constraints.

RQ4 (Focusing Phenomenon) Data Collection and Analysis

The fourth and final question used elements of the focusing phenomenon framework (Lobato et al., 2013) to coordinate how aspects of the classroom environment (e.g., design features, tasks, tools, and the teacher) directed students' attention towards understandings of race and racism in the context of data science. Particularly, the research question is:

> **Research Question 3:** Focusing Phenomenon
> > a. How do elements of the TE contribute to the students' understanding of race and racism in the context of statistics and data science?

I began by looking at student responses to questions about race neutrality in statistics and data science from class assignments, noting properties of features of race neutrality that students mention in their responses (Centers of Focus). Then, I identified discourse practices that may have directed students' attention to the Centers of Focus (Focusing Interactions), describe the features of the tasks surrounding that focusing interaction, and discuss the participation dynamics and class-established norms that may have influenced what students noticed (nature of mathematical activity).

Data Collection

Homework Responses. This analysis was grounded in students' responses to the four prompts about race and data neutrality shown in Table 3.6. All questions were collected as homework (journal reflections for that day). These four prompts were selected for two main reasons: (a) they focused on race and data neutrality and (b) the timing when they were asked in the course. First, although there are slight variations across the questions, all four prompts were designed to ask students about data neutrality. This allowed me to identify any emerging or evolving conceptions about students' views on data neutrality and, during whole class discussions, the role of race and racism in the PPDAC cycle.

Second, all four prompts were assigned at specific points in the course. Particularly, although race and racism were discussed throughout the entire course, the four prompts from Table 3.6 were asked in preparation for lessons where we were designed to summarize the role of race and racism in the PPDAC cycle. The timing and placement of these prompts was designed intentionally so that students would be prepared for the class discussion and so that I could analyze their individual responses to data neutrality as well as the class discussion that followed about race and racism in the PPDAC cycle.

Table 3.6: Questions about data neutrality used to identify the centers of focus in the TE data with short descriptions of the class when the question was assigned

Class	Class 1: Pre-Survey	Class 3: Homework	Class 6: Homework	Class 12: Post-Survey
Question	**1**. A student says, "numbers speak for themselves." Do you agree or disagree? Please explain.	**2**. A friend tells you that "data can't be racist, numbers don't lie." Do you agree or disagree with this statement? Explain	**3**. One of your colleagues states that "technology is politically neutral, therefore data is politically neutral." Do you agree or disagree? Explain	**4**. A student says ,"numbers speak for themselves." Do you agree or disagree? Please explain.

The first question was assigned after Class 1 when students were introduced to the class and the goal of the class, but before any class activities about race and racism in the context of statistics and data science. The second question was assigned after students had read about Critical Race Theory and talked about it as a class and was assigned alongside a reading about Quantitative Critical Race Theory (QuantCrit). At this point, we had also discussed the role of the problem context in statistics and data science (the Problem phase of the PPDAC). The third question was assigned for the Class 6 homework. By this point, students had learned about sampling, randomization, study designs (the Planning and Data phases of the PPDAC cycle), and also watched a video on algorithmic bias (Fong, 2021) as part of their homework. The last question was assigned during Class 12 after students had learned about different regression models (the Analysis phase of the PPDAC cycle). This last question was not debriefed as a class. The Conclusion phase of the PPDAC cycle was discussed throughout the class (e.g., role of researcher positionality when writing reports, impact of sampling on generalizability, how to interpret statistical models).

Classroom Data. The second source of data was classroom lessons, activities, and recordings of the lessons. In Particular, I used the contact summary forms and field notes to identify interactions of interest. Field notes were taken by another graduate student that served as a co-researcher while I taught the course. The interactions included in this dissertation all occurred in whole class formats. Although there were key moments in small group activities that occurred in breakout rooms, I was not able to record breakout room activities and discussions because it was not an option over Zoom. In some cases, I present data from breakout room interactions based on the field notes or contact summary forms, but future studies (and in-person studies) should find ways to record small group interactions.

Data Analysis

Drawing on elements of Walters' (2017) and Lobato et al.'s (2003, 2013) four analytical passes for focusing phenomenon, the first analytical pass entailed inferring Centers of Focus (CoFs). Particularly, the goal of the first analytical pass was to identify the Centers of Focus in the four homework questions about biases in data neutrality, shown in Table 3.6. Since the first homework was assigned in the beginning of the course before talking about the role of race and racism in data science, the first homework question was used as a benchmark to help identify Centers of Focus that emerged during the class. Thus, for this dissertation, Centers of Focus were defined as properties or features of data neutrality that at least one student stated in their responses to the questions from Table 3.6. It is important to note that the absence of a CoF in a student response does not necessarily imply that the student did not notice that property of feature of data neutrality but, rather, that they did not explicitly note a property of feature of data neutrality in their homework. Furthermore, one student stating a property, feature, regularity, or conceptual object related to data neutrality may have not necessarily been a Center of Focus for the entire group or class. However, given the limitations of teaching and conducting research over Zoom (e.g., not being able to record breakout out room conversations that may provide further evidence for Centers of Focus), the responses to the four prompts were used to help identify moments during the whole class discussions related to data neutrality. Avenues for future research and modifications for in-person research are described in the conclusion.

The second analytical pass (identifying focusing interactions) entailed identifying class contributions and discourse practices (from teachers and students) that were related to each of the CoFs. For this dissertation, I considered CoFs that emerged throughout the course, then looked for focusing interactions during class meetings prior to the first assignment in which that CoF emerged in the homework questions. Drawing from cognitive science and applied linguistics anthropology, Lobato et al. (2013) used three codes to identify focusing interactions in the classroom videotaped data, shown below in Table 3.7 (excerpt from Table 3, p. 824). Instances of the three codes were noted in all of the field notes.

For this dissertation I specifically focus on highlighting because there were few instances of quantitative dialogue and renaming as evident in the field notes. It is worth noting that quantitative dialogue may have not been as evident because it may be different from statistical or data scientific dialogue. For example, statistical and data scientific dialogue may instead be verbal communication that focuses attention on variation, related to the context, data wrangling (i.e., cleaning data), or other attributes that are statistical but not mathematical. Future research could define and look for statistical dialogue as well look through the classroom videotaped data to confirm that there were little to no instances of statistical dialogue and renaming.

Table 3.7: Codes that Lobato et al. (2013) used to identify focusing interactions in the classroom videotaped data (excerpt from Table 3, p. 824)

Focusing Interaction Codes	Description of Codes
Highlighting	Operating visibly on external phenomena, including the acts of labeling, marking, annotating, and gesturing.
Quantitative Dialogue	Verbal communication that focuses attention on quantities as measurable attributes of objects.
Renaming	Changing the name of a construct that has been previously defined, using a category of meaning from mathematical practice.

Furthermore, I extended highlighting to look for Mathematical Language Routines (MLRs) that may have focused students' attention to the particular CoF by operating visibly on external phenomena (e.g., labeling, marking, annotating, gesturing, or displaying information to the entire class). I extended highlighting to look at MLRs for two reasons. First, MLRs were part of the design features (DF3: Communicate), which could provide further illustrations of how the design features were enacted and relate them to student learning. Second, the MLRs could provide specific examples of the activities that led to highlighting a specific CoF. For example, the Collect and Display routine may *highlight* student contributions by noting, annotating, labeling, or adding on to student contributions that are displayed to the whole class.

The third analytical pass aimed to identify features of the mathematical task that may be related to the particular CoF. For this dissertation, this included providing a description of the task as it was designed and identifying any possible affordances and constraints of the task. I also present possible modifications for the task for future iterations. These possible modifications are guided by my reflection in the contact summary form that was written immediately after I taught the lesson, many of which included specific recommendations for future iterations based on how the lesson was enacted that day.

The fourth analytical pass entailed describing the nature of the mathematical activity. The nature of the mathematical activity are the classroom norms that may have influenced participation dynamics and, consequently, what students notice in the activity. There were two steps to this process: (a) identifying general classroom norms, and (b) describing how the general classroom norms were related to students' noticing. First, I identified any "general classroom obligations" (Cobb et al., 2009, p. 52, as cited by Lobato et al., 2013, p. 823) from the lesson plans and classroom videos of whole-class conversations. For example, students may be expected to provide examples to strengthen their claims, share their screen to show their computer programming code, or ask other students follow-up questions. These general classroom norms may differ across lessons and activities. Since all lessons were taught online over Zoom, I do not have video recordings of breakout room conversations. In a future study, it may be worth examining the norms within the small group or breakout room conversations. Second, I described how the general classroom norms may be related to the students' roles, the teacher's role, and the students' noticing in the activity.

I end by connecting the CoFs, focusing interactions, features of the task, and nature of the mathematical activity to the design features. The six design features in the TE, which are described in Chapter 4 were about incorporating opportunities for students to: (a) reflect on the structures of social injustices, (b) deepen and revise thinking, (c) communicate, (d) engage with relevant contexts, (e) engage with all phases of the PPDAC investigation cycle, and (f) design and implement a statistical study throughout the course. Although the third design feature (dialogue) includes journals, they were not included in this analytical pass because they were used in the first analytical pass to help identify the Centers of Focus in student response to the homework question.

Coder Agreement

Similar to the qualitative coding of practices, I was not able to double-code this data. Future iterations of this study should consider double coding the student responses to the four prompts. Additionally, it may be worth having an additional researcher help identify focusing interactions as well (mainly, helping identify instances in the classroom interactions that may have guided students attention to a particular Center of Focus).

Related Researcher Experience

I leveraged my experiential knowledge as a student, educator, and research using both quantitative and qualitative methods. In particular, I have experience teaching high school and college-level math courses, including mathematics for pre-service mathematics teachers and attempting to teach mathematics for social justice. I also have a graduate degree in statistics and experience using the statistical skills and methods that were used in Research Question 2.

In terms of qualitative research experience, I am currently collaborating with my advisor, Dr. William Zahner, on a project that explores how high school mathematics classrooms can be designed to create opportunities for students, including English Learners, to participate and contribute to whole class discussions. Through this project, I developed experience designing and analyzing TEs (Zahner et al., 2021a) and creating design principles (Zahner et al., 2021b) similar to the analysis in Research Question 1. I have also developed familiarity with the MaxQDA qualitative coding software used in this study and transcribing interviews and classroom interactions. Additionally, part of an independent project, I worked with six teachers to explore their knowledge about correlation and regression. Across these experiences, I developed experience with creating tasks for interviews and using elements of grounded theory to analyze data, similar to the qualitative coding used in Research Question 3.

Researcher as Teacher Positionality

It is also important to acknowledge the biases that may arise from being the researcher and teacher in this study which may be classified as practitioner research, teacher inquiry, and action research (Anderson, 2002; Brantlinger, 2013; McKernan, 1991). Particularly, I am an “insider” (Anderson & Herr, 1999) in that I am a member of the classroom, have teaching experience, have strong relationships with other educators, am interested in teaching mathematics for social justice, and plan on pursuing a career in education after this doctoral program. At the same time, I am an “outsider” in that I am taking the role of a researcher with a research agenda. More importantly, my role as an outsider puts me at risk of using “data to tell a deception as easily as a truth” (Connelly & Clandinin, 1990, as cited by Anderson & Herr, 1999).

In an effort to mitigate these potential biases, I drew on Anderson and Herr’s (1999) validity criteria for action research. This criterion includes: (a) outcome, (b) process, (c) democratic, (d) catalytic, and (e) dialogic validity. Outcome validity includes the extent to which the desired outcome occurs. This requires that researchers ask if the action (TE intervention) was successful or not and under what conditions. In some cases, this may require that researchers are flexible in reframing research designs and questions. The outcome validity of this study will be the central goal of the conclusion of this study.

Process validity is about the research design, including the cyclical and iterative nature of research and using appropriate methods to ensure outcome validity. This was especially important in this study as it helped determine whether findings emerge from the data or my personal biases. One method I used to ensure process validation is triangulation. *Triangulation* is a qualitative method in which multiple sources of data are used to develop an understanding of a phenomenon (Miles et al., 2020; Patton, 1999). For example, in this study, the descriptive accounts helped me identify key moments in the lessons that supplement findings from the interviews, allowing for repeated verification (Miles et al., 2020).

Democratic validity is about the extent to which stakeholders are included in the research. This is similar to Gutiérrez's (2017) notion of knowledge with students and Milner's (2007) notion of researching the self in relation to others. Of special interest is who the research is benefiting and at the expense of who: outsiders or insiders. Outsiders include researchers, policymakers, or other individuals and organizations that are not practitioners. Insiders include students and practitioners. A primary goal of this dissertation was to prepare the next generation of data science educators. Thus, the PSMTs are the center of this study. To ensure this, I asked for continuous feedback from the students via both anonymous and identifiable surveys throughout the course. The surveys were administered using a Google Survey. This form included an informal assessment to measure student learning of that topic as well as a check-in with students where they provided anonymous feedback about the lessons and course.

Catalytic validity is about deepening the researcher and participants' knowledge about the action. This often includes tracking the learning growth of all stakeholders involved. As a researcher, this includes keeping a diary (e.g., Brantlinger, 2013) or reflexive memos (e.g., Kokka, 2020). In my study, I kept analytical memos that were created after I teach each class. These memos were added to the field notes or related contact summary forms. Furthermore, these memos helped keep track of wonderings, breakthroughs, challenges, and other notable moments that guided my analysis and, in some cases, presented in the results or discussion.

Similarly, students' coursework served as reflection opportunities for them to track their learning. For example, students were asked to react to their initial definitions of equity throughout the course (once in the middle and once in the end).

Finally, dialogic validity is about the collaborative inquiry within research fields, typically through peer review. One application of this is including a reflective partner. In fact, Prediger et al. (2015) also recommends collaborating with other knowledgeable teachers or researchers throughout the design process. Thus, I invited a mathematics education researcher to partake in the lesson planning process, analysis, and act as a co-researcher. The co-researcher was another graduate student who was working on their master's thesis. They attended almost every class session, took field notes during the class, and met with me once a week to review the field notes and discuss any important moments from the class.

Chapter 4: RQ1 (Design Features)

The first set of research questions were about the design features that guided the design of the TE. In particular, the research questions were:

> **Research Question 1:** Design Features
> a. What design features support students' understandings of race racism in the context of statistics and data science?
> b. How were the design features enacted in the curriculum?

The goal of the design features was to provide students opportunities to develop critical statistical and data scientific consciousness. Drawing from Freire (1988), *critical consciousness* refers to "learning to perceive social, political, and economic [oppressor-oppressed] contradictions, and to take action against the oppressive elements of reality" (p. 17). One way of developing critical consciousness is through engaging in *praxis*, or a cyclical and complementary relationship between *reflection* (critiquing the social, political, and economic oppression) and *action* (individual or collective action taken to challenge oppression).

Reflection and action are cyclical because we are in a permanent state of discovery (Freire, 1988), which entails forming knowledge about, reflecting on, healing from, and resisting oppression as well as reflecting on that growth. Reflection and action are complementary because action is embedded in reflection and reflection is embedded in action. Particularly, action that is not critiquing social oppression may not necessarily lead to transformational change (Solórzano & Delgado Bernal, 2001) if it is not targeting the larger social, political, cultural, or historical structures that lead to that oppression. Similarly, reflection without action may not transform or challenge oppression. Thus, action and reflection do not occur separately but, rather, are complementary.

Although developing critical consciousness and praxis are a focus of critical pedagogy (e.g., teaching mathematics for social justice), they are not well researched in mathematics education research (Martinez, 2020). For this dissertation, I posit that statistics and data science may help develop *critical statistical and data scientific consciousness* if it is used as a sociopolitical tool used to identify and learn about social and racial injustices (reading the world with statistics, or reflection) and challenging the oppressive conditions that foster those injustices (writing the world with data, or action). Since Freire does not foreground the role of race and racism in social justice, I also adopted QuantCrit's focus on the centrality of race and racism in statistics and data science as a defining source of racial injustice in the USA context. For example, how do over- and under-representation biases replicate racism in facial recognition algorithms (Buolamwini, 2017)? How do artificial intelligence recruiting tools reinforce sexist hiring practices? How can racially conscious computing be used to advance racial justice in the USA?

The remainder of this chapter presents the design features I used in the TE and illustrate how they were enacted in the TE curriculum. There are six design features that were used in the course design, shown in Figure 4.1. The six design features are about opportunities for students that were considered when designing the course lessons and activities. Inspired by Freire's conceptualization of praxis, three design features were categorized under reflection (blue) and three were categorized under action (green). The design features are presented separately and across the two categories for simplicity, but occurred alongside and supported each other throughout the design of the course. This is reflected in the visual in Figure 4.1 by the overlap and connectedness with between all the features. I also present the design features as part of a larger sequence (noted by the three dots on the left and right) to emphasize that consciousness is an ongoing and cyclical process. In doing so, I recognize the different types of knowledges that students are bringing into the classroom as well as the new types of knowledges that they will create on their path towards critical statistical and data scientific consciousness.

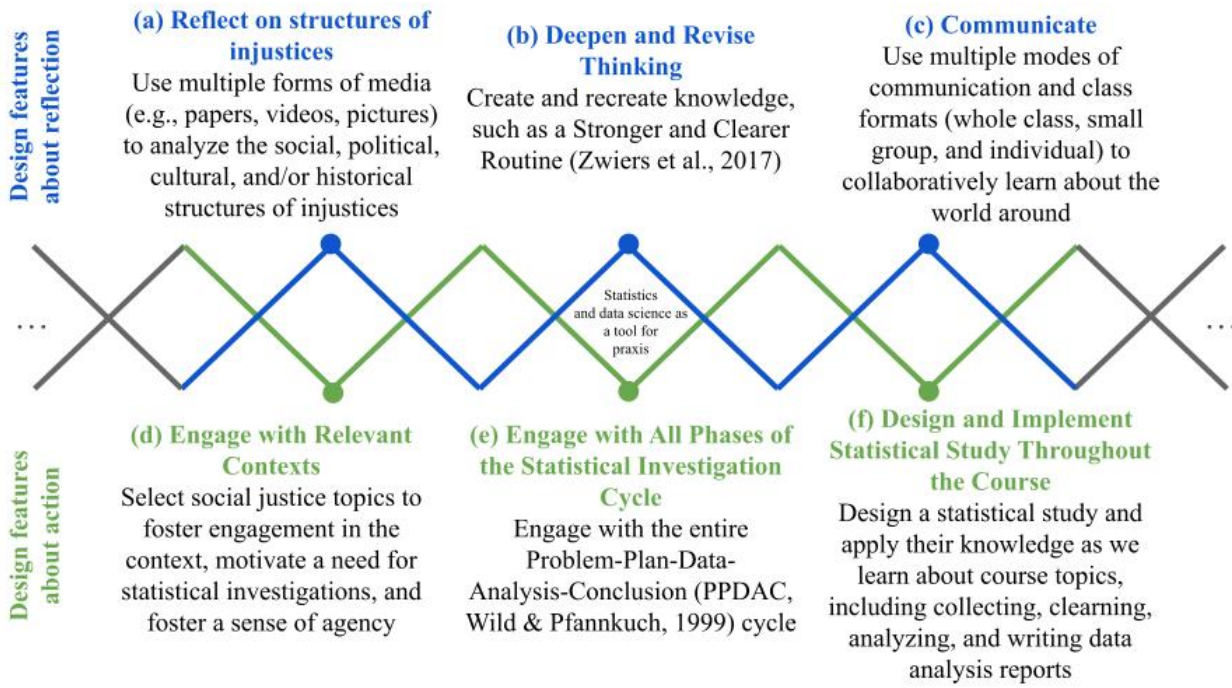


Figure 4.1: Design features about the opportunities for students that were incorporated into the curriculum

In what follows, I define each design feature, describe the motivation for the design feature, and illustrate how the design features were enacted in the curriculum. It is important to note that the examples used to illustrate how the design features were enacted in the curriculum were chosen to highlight the potential of the intended curriculum. Therefore, the examples may not be representative of all classroom interactions or students' experiences. I end the chapter by discussing the overlap of the six design features and how they support each other.

Design Feature 1: Reflect on Structures of Social Injustices

The first design feature is about incorporating opportunities for students to reflect on the social, political, cultural, and historical contexts of oppression in the context of statistics and data science. An important consideration when designing a course using TMSJ is moving beyond what Giroux (2001a) identifies as pedagogies of despair, or a cynical and pessimistic view of social justice that is not contextualized in a critique of oppression. Giroux states that "[c]ritical pedagogy locates discursive practices in a broader set of interrelations, but it analyzes and gives meaning to such relations by defining them within particular contexts constructed through the operations of power" (p. 19). That is, a goal of critical pedagogy is to contextualize the injustices within the larger social, political, cultural, and historical landscape (operations of power) rather than placing responsibility on individuals (emphasizing the notion that action is guided by reflection).

Carefully attending to those operations of power and action is important for at least three reasons: (a) avoiding reinforcing privilege (Esmonde, 2014; Kokka, 2020), especially with students from privileged experiences, (b) avoiding reinforcing deficit narratives, especially for BIPOC students (Brantlinger 2013; Rubel et al. 2016), and (c) situating statistics and data science within a sociopolitical context. For example, a lesson on the "achievement gap" may be intended to analyze some of the educational outcomes of students. However, gap-gazing may provide a limited view of educational equity that may only focus on measurable outcomes (e.g., teacher knowledge, student scores on standardized assessments) that lead to deficit interpretations that place blame on students for not performing well or teachers for not preparing students. These deficit interpretations fail to account for the larger structural racism and social factors (Ladson-Billings, 2006b) and may lead to stereotype threat (Brantlinger 2013; Rubel et al. 2016). Finally, since "data are not just numbers, they are numbers with a context ... [the] context provides meanings" (G. W. Cobb & Moore, 1997, p. 801), it is important to situate data science and statistics in the context from which they are socially created. From a QuantCrit perspective and in this TE, situating data in its context entails discussing the racialization of data

<table>
<tr><th colspan="2">Class 4: Statistical Questions. Reviewing QuantCrit and its implications for this class</th></tr>
<tr><th colspan="2">Lesson Summary</th></tr>
<tr><td colspan="2">In the first part of today’s lesson, we talk about three types of variation using R: (a) variability within a group, (b) variability within and across groups as well as covariation, and (c) variability in model fitting. Students will then look at different questions (some statistical, some not although not called “statistical questions yet”) and describe the different types of variation that each question is considering. After showing which ones are statistical questions, we will come up with a class definition of what statistical questions are.
The second part of today’s lesson will build on Class 3 and our understanding of statistical questions. We will talk about deficit and anti-deficit framing of questions by looking at some examples by Harper (2010) and then talk about Harper’s (2010) and QuantCrit’s implication for the course project. We will end by creating a class set of guiding questions to consider for this project related to: 1. What is the goal of this research project?, 2. What is your research question? 3. What is the context? What do we need to know about this context?, and 4. Why is this context important?. Note: These are the same questions they had for homework.</td></tr>
<tr><th>Statistical Pedagogical Goals</th><th>Social Justice Pedagogical Goals</th></tr>
<tr><td>Understand how to read data, find descriptive statistics of data, and make basic visualizations of data using the R programming language

I.a. Formulate statistical investigative questions
Formulate multivariable statistical investigative questions and determine how data can be collected and analyzed to provide an answer

IV.f. Interpret results
Use multivariate thinking to understand how variables impact one another</td><td>Justice 13: JU.9-12.13
I can explain the short and long-term impact of biased words and behaviors and unjust practices, laws and institutions that limit the rights and freedoms of people based on their identity groups

Action 20 AC.9-12.20
I will join with diverse people to plan and carry out collective action against exclusion, prejudice and discrimination, and we will be thoughtful and creative in our actions in order to achieve our goals.</td></tr>
<tr><th colspan="2">Homework</th></tr>
<tr><td colspan="2">1. Reading
Harper, S. R. (2010). An anti-deficit achievement framework for research on students of color in STEM. New Directions for Institutional Research, 2010(148), 63-74. Link provided

This paper discusses how we can write anti-deficit research questions. Answer the following questions:
a. The authors imply that the way we ask questions has an important role in research. Do you agree? Why or why not?:
b. Your student is worried that their research question might carry some deficit-oriented framing. What recommendations would you provide for students to rewrite their question to use an anti-deficit framing?:
c. What, if any, is the relationship between race, racism, and statistics?:

2. Pre-Assessment
Due date</td></tr>
</table>

Figure 4.2: Lesson summary, pedagogical goals, and homework of Class 4

during the PPDAC cycle to fully understand the oppressive structure of social injustices. Thus, to avoid reinforcing privilege or reifying deficit narratives, give data context, and fully understand the oppressive structures of social injustices, in this data science for teachers course, I incorporated multiple opportunities (e.g., papers, videos, pictures) for students to reflect on the different social, political, cultural, and historical contexts of the data discussed in the TE.

How Design Feature 1 Was Enacted in the TE

The guiding question for this design feature was: How can we attend to the social, political, cultural, and historical contexts of racial oppression in the context of statistics and data science? Below I present examples of how Design Feature 1 was enacted in the TE.

Class 4: Statistical Questions, Framing, and QuantCrit

The first example comes from Class 4 where we discussed statistical questions and QuantCrit. An overview of the lesson, pedagogical goals, and homework assigned at the end of the class are shown in Figure 4.2. Prior to Class 4, we had discussed Critical Race Theory, read about QuantCrit for homework, and students were asked to answer the questions shown below in Figure 4.3 in preparation for the project. The purpose of this assignment is to help students draft their introduction and motivation. These questions were also revisited later in Class 4.

1. What is the goal of this research project?
2. What is your research question?
3. What is the context? What do we need to know about this context?, and
4. Why is this context important?

Figure 4.3: Homework for Class 3

In the first half of Class 4, we used the R programming language to review different types of variation. We also were first introduced to statistical questions, then viewed sample statistical questions and what type of variation they aim to explore. The first part of Class 4 did not have an explicit critical focus, but it laid the foundation for some of the critical applications discussed in the second half of Class 4. In the second half of the class, we began by debriefing the homework readings on Critical Race Theory (Ladson-Billings, & Tate, 1995) and Quantitative Critical Race Theory (Crawford et al., 2018), and their implications for statistical questions. We did this by reviewing sample deficit and anti-deficit reframed questions presented by Harper (2010), shown in Figure 4.4.

Table 6.4. Sample Reframed Research Questions for Students of Color in STEM

Deficit-Oriented Questions	*Anti-Deficit Reframing*
Why do so few pursue STEM majors?	What stimulates and sustains students' interest in attaining degrees in STEM fields?
Why are they so underprepared for college-level mathematics and science courses?	How do STEM achievers from low-resource high schools transcend academic underpreparedness and previous educational disadvantage?
Why are their grades and other indicators of academic achievement disproportionately lower than those of their White and Asian American counterparts?	What enables students of color in STEM to make the dean's list, compete for prestigious fellowships and research opportunities, and earn high GPAs?
Why do so many change their majors to non-STEM fields?	What compels students of color to persist in STEM fields, despite academic challenge and the underrepresentation of same-race peers and faculty?
Why do so few continue on to graduate degree programs in STEM?	What are common aspects of students' pathways from high school completion through doctoral degree attainment in STEM fields?

Figure 4.4: Reframed research questions from Harper (2010, p. 69, Table 6.4)

It is important to note that while Harper is not explicitly referencing CRT or QuantCrit and that the focus is of the paper is research questions (not necessarily statistical questions), this activity was designed to help students see examples of deficit questions and how they can be reframed to anti-deficit questions. Furthermore, statistical questions address a type of variation whereas Harper mainly discusses research questions as questions that guide a research investigation (that may not necessarily be statistical). Nonetheless, Harper (2010) helped guide conversations about how we can write anti-deficit framed statistical questions for the course project.

After comparing and contrasting the sample deficit and anti-deficit reframed questions, students were asked to respond to the following questions in groups and be prepared to share to the class:

1. What do the deficit-oriented questions have in common?
2. What do the anti-deficit reframings have in common?
3. If a student asked you how you could transform a deficit-oriented question to an anti-deficit question, what would you say?

In response to the first question, students noted that the deficit questions were "more about a person than the structure" and that some of these questions "perpetuate the model-minority idea" during the whole-class debrief. The student that mentioned model-minorities had learned about this in a previous class, then provided a short description for the rest of the class. The students also noted that the anti-deficit reframing "focus on the surroundings, structures" and that there is an "emphasis on social structures." Finally, for the third question, multiple groups noted that one way to transform a deficit-oriented question was by "asking questions about surroundings, structures, systems, dot dot dot" instead of asking questions that "blame individuals" In other words, students in the TE noticed that the deficit-questions placed responsibility on individuals, either implicitly or explicitly, whereas the anti-deficit reframed questions were about the same topic but aimed to explore the surroundings, structures, and systems at play.

We ended the class by discussing the QuantCrit implications for the course project. We focused on expanding on the fifth (Social justice/equity orientation) and first tenet (Centrality of Race and Racism) of QuantCrit and how they can be used to think about the four prompts assigned for homework (Figure 4.3) that were used to draft the introduction of the final course project. The goal of this activity was to build explicit connections between QuantCrit, statistical investigation, and the course project. As a class, we added the sub questions (text in black shows the original questions assigned for homework) shown in Figure 4.5.

1. What is the goal of this research project? (Tenet 5: Social justice/equity orientation)
    a. Convince us that this topic is important and why
    b. How are you using anti-deficit framing and avoiding deficit framing?
    c. How are you making an intentional effort to address biases?
2. What is your research question? (Tenet 5: Social justice/equity orientation)
    a. How are you using anti-deficit framing and avoiding deficit framing?
        i. Is this question about individuals or structures?
    b. Why is this question a statistical question?
    c. What type of variation is this question addressing?
3. What is the context? What do we need to know about this context?, and (Tenet 1: Centrality of Race and Racism)
    a. What background information do I need to know to understand this context?
    b. Is the problem taking into consideration race and racism?
        i. Or other forms of discrimination?
4. Why is this context important? (Tenet 1: Centrality of Race and Racism)
    a. How will your project bring awareness to issues in this topic?
        i. Race and racism?
    b. What change might come out of raising awareness for this topic?
    c. How will your project help advance social justice in this context?
        i. At the individual level? Structures? Other

Figure 4.5: Subquestions for the original three questions from Class 4

Central to this activity was thinking about the larger power structures at play. For example, students considered how their research questions were anti-deficit and addressed larger structures instead of individuals. Similarly, we discussed how the project should forefront race and racism (and/or other forms of discrimination). Finally, we discussed how the project may lead to some social change that must consider changes at the individual level but also implications for possible changes at the structural level that go beyond raising awareness. In other words, students were engaging with praxis as they were considering how they could draw on their reflections about their larger power structures to inform action that targets different levels of social oppression.

Class 6: Sampling and QuantCrit

The second example of Design Feature 1comes from Class 6 where we discussed sampling methods and sampling bias. An overview of the lesson, pedagogical goals, and homework assigned at the end of the class are shown in Figure 4.6. Class 6 was the beginning of a unit on statistical design, which included sampling and types of studies. For homework, students were asked to watch a video on automating racism (Fong, 2021). This video builds on the work about algorithmic bias described by Buolamwini (2017) and Benjamin (2019). The full assignment is shown in Figure 4.6. Notably, students ranked this assignment as one of the activities that best helped illustrate how race and racism are interwoven into data science in the post-survey for the class, one overall one of their favorite activities of the class. Building on the ideas of QuantCrit, the goal of this activity was for students to begin to see data science as a social activity, including how structures of social injustice are encoded through data.

For example, in response to Question 1c about the role of humans in machine learning, a student wrote that “the machine follows human patterns, so whatever it is that said group is interested, machine learning caters to those interests.” Similarly, in response to Question 1d about reinforcing social biases, a student stated:

> It has been brought to our attention that computer algorithms and code are being taught our social biases. Which means that some forms of technological prediction or personalization can have racist effects. Technology is used to supposedly make our lives easier, but who created the technology and what actually went into the process of its design? We do not see all the human decisions that go into the design of technology. Machine learning algorithms are taught by people, their examples come from people. This means that the decisions are not separate from us or our biases or our history.

<table>
<tr><th colspan="2">Class 6: Sampling Methods and Sampling Bias</th></tr>
<tr><th colspan="2">Lesson Summary</th></tr>
<tr><td colspan="2">In the first part, we revisit the PPDAC cycle to talk about the planning/data phases, mainly sampling methods and sampling bias. Students are then shown four sampling scenarios and asked to describe the sampling method and react to the sampling method (e.g., is it an appropriate method). In the second half of the class, we build on the sampling scenarios by discussing four non-probabilistic (convenience, voluntary, purposeful, snowball) and four probabilistic (simple random, systematic, stratified, cluster) methods. For each probabilistic method, students are either presented code and asked to describe it or asked to write their code for that method. We ended the class by discussing affordances and constraints of each sampling method (which will be part of their homework).</td></tr>
<tr><th>Statistical Pedagogical Goals</th><th>Social Justice Pedagogical Goals</th></tr>
<tr><td>II.d. Collect/consider data
Understand the role of random selection in sample surveys and the effect of sample size on the variability of estimates

II.e. Collect/consider data
Understand the role of random assignment in experiments and its implications for cause-and- effect interpretations

II.i. Collect/consider data
Understand that in some circumstances, the data collected or considered may not generalize to the desired population, or this data may be the entire population</td><td>Justice 13 JU.9-12.13
I can explain the short and long-term impact of biased words and behaviors and unjust practices, laws and institutions that limit the rights and freedoms of people based on their identity groups.

Action 20 AC.9-12.20
I will join with diverse people to plan and carry out collective action against exclusion, prejudice and discrimination, and we will be thoughtful and creative in our actions in order to achieve our goals.</td></tr>
<tr><th colspan="2">Homework</th></tr>
<tr><td colspan="2">1. Video
Watch <u>this</u> YouTube video and answer the following questions. When possible, provide examples from the video to support your claim.
a. Around 6:45, the video introduces “how machine learning works and what can go wrong.” Assume that you are teaching a high school data science course and one student asks you “how does machine learning work?” How would you respond?
b. In the video, they mentioned “algorithmic bias.” What do they mean by this term and how might this be related to sampling? How might it be related to QuantCrit?
c. What is the role of humans in machine learning? Provide at least two to three examples to support your claim.
d. In her book Race After Technology, Ruha Benjamin states that “Ultimately the danger of the New Jim Code positioning is that existing social biases are reinforced – yes. But new methods of social control are produced as well. Does this mean that every form of technological prediction or personalization has racist effects? Not necessarily. It means that, whenever we hear the promises of tech being extolled, our antennae should pop up to question what all that hype of “better, faster, fairer” might be hiding and making us ignore. And, when bias and inequity come to light, “lack of intention” to harm is not a viable alibi. One cannot reap the reward when things go right but downplay responsibility when they go wrong”. Assume that you are teaching a high school data science course and want to discuss this with your students. How would you explain this to them?
e. What can we do to help mitigate algorithmic bias?
f. One of your colleagues states that “technology is politically neutral, therefore data is politically neutral.” Do you agree or disagree? Explain
2. Sampling Review
Complete the Google Doc below: (<u>link</u>)
Be prepared to share with your group on Wednesday</td></tr>
</table>

Figure 4.6: Lesson summary, pedagogical goals, and homework of Class 6

In both of these responses, students highlight how machine learning is being trained to follow human patterns. Sometimes, these patterns may appear to help make our lives easier (e.g., when Amazon suggests that I need batteries for a toy that does not include batteries). However, the second student highlights how these patterns may replicate social biases which may lead to algorithmic biases that encode racism.

Design Feature 2: Deepen and Revise Thinking

The second design feature was incorporating opportunities for students to deepen their learning by creating, reflecting, and recreating knowledge across different time scales (e.g., within a class, across two or three classes, across the entire course term). This is inspired by the Stronger and Clearer MLR routine (Zwiers et al., 2017) that aims to strengthen student responses, including a focus on reflecting on what they learned over a given period of time (e.g., one time in a short classroom activity, across an entire lesson or unit, in the beginning and end of the course). This is also similar to Jansen's (2020) rough draft math where students may be hesitant to share their answer if they are not sure if they have the correct answer. This may be amplified when discussing social and racial justice in the context of mathematics where there may be a perception of a high social cost to having one misstep. Yet, learning often entails talking and writing about in-process or unfinished ideas. Turning to Freire (1988), this design feature is also motivated by the cyclical and ongoing relationship between reflection and action, where we are in a permanent state of discovery. This includes forming knowledge about, reflecting on, healing from, and resisting oppression as well as reflecting on that growth. Thus, I tried to reinforce a learning model that is open to growth and involves changing ideas.

For example, similar to Gutiérrez (2002), I used to define educational equity in terms of no longer being able to predict mathematics achievement based on race, ethnicity, class, or other identity markers. This was perhaps influenced by my educational background (mathematics and statistics) that may have placed an emphasis on quantifying educational equity and using standardized measures as a "fair way" to compare students. However, my own understanding of educational equity has evolved and grown over time to include critical axis of identity and power. Turning to the broader social justice landscape, this growth is similar to the evolution of terminology like BIPOC that shift away from POC, marginalized, minoritized, or minority, where BIPOC is person-first and aims to illuminate the injustices affecting Black and Indigenous communities. This terminology will likely change and evolve in the future. In fact, this permanent state of discovery and reflecting on that growth highlights new perspectives, why those new perspectives are important, and may help normalize that growth process. Thus, for this TE, I included opportunities for students to reflect on instances where their perspectives have changed and why their perspectives changed.

How Design Feature 2 Was Enacted in the TE

The guiding questions for this design feature were: what opportunities are there for students to respond to a prompt, engage in some course activity, be asked to view their previous response, clarify their initial response if needed, and reflect on any changes? How can this reflection be supported and what are some anticipated changes? Below, I provide three examples of how this was enacted at the individual, group, and whole-class level.

Deepen and Revise Thinking at the Individual Level

At the individual level, students were often asked repeated questions or variations of the same questions that targeted the same concept and asked to reflect on their learning. For example, students were asked the four prompts show in Figure 4.7. All questions were collected

Figure 4.7: Sample responses about data neutrality. Emphasis added

Student	**Class 1: Pre-Survey** A student says, "numbers speak for themselves." Do you agree or disagree? Please explain.	**Class 3: Homework** A friend tells you that "data can't be racist, numbers don't lie." Do you agree or disagree with this statement? Explain	**Class 6: Homework** One of your colleagues states that " technology is politically neutral, therefore data is politically neutral." Do you agree or disagree? Explain	**Class 12: Post Survey** A student says ,"numbers speak for themselves." Do you agree or disagree? Please explain.
Nicole	Numbers are objective, they are definitive.	**At first**, I did agree because numbers are objective and are not sueded by opinions. But I failed to look at the bigger picture, the **historical and political context that may be surrounding the data.** So in a way, I would agree that data cannot be racist but the **data interpreters** can be for neglecting to look at the bigger	**Before I would have agreed** but lately, with the exposure to the subjectivity of data interpretation, I would have to disagree. It' s true that 10 means 10 but then we forget how **behind these sets of data is a human being interacting with it and coming to conclusions**.	No. Numbers are just numbers but in order to interpret, **we must look at the bigger picture. This might show if there is anything wrong with the data or how to interpret the data.**
Jaime	I would have to disagree with this student as **numbers can be manipulated for whatever purpose the person needs**. A recent example of this would be the COVID 19 death in New York when some of the deaths were put into different columns and not the actual reason of COVID. The numbers showed a slightly different story than what the actual truth was at the time.	I would have to disagree with this statement. This is due to the fact the **numbers can be manipulated** to say pretty much whatever they want. Along with that if you survey people about a topics they do not know about then that would might show how the population might not be interested in that topic yet the people that might know the topic would show it is important	I disagree with it fully… **machine(s) learns from the humans**' writing so i**f there is a bias that the humans have written it then it will leak into the code**. Like the University that had the machine-learned what was, it had a bias to spot White faces better and that was due to the lack of other faces that the machine was shown as examples…. so we should be able to teach the machine better if we give other examples.	I would disagree as numbers can be manipulated and do not always show the real data

as homework. Although there is slight variation across the questions, all four prompts are generally about data neutrality. For the exception of the first question (from Class 1), students were able to view their responses to the previous question before responding. For example, students saw their response to the question from the Class 1 survey, then asked to react to the question in Class 3.

The first student, Nicole, is a student who initially stated that she agreed that numbers speak for themselves because they are "objective" and "definitive." However, after the QuantCrit discussion in Class 3, she added that we need to also consider the "historical and political context that may be surrounding the data" that may lead to biased interpretations (from a data consumer perspective). Similarly, for the Class 6 homework on sampling and sampling bias, she stated that "before I would have agreed," but began to talk about some of the social construction of data and how "behind these sets of data is a human being interacting with it and coming to a conclusion." Thus, for Nicole, this activity may have afforded opportunities for her to understand new perspectives on how data is not always neutral.

In Class 1, Jaime, stated that numbers do not speak for themselves because "numbers can be manipulated for whatever purpose the person needs." One interpretation of this is that numbers may be manipulated to advance the agenda of the person conducting the analysis (from a data producer perspective). After Class 6, Jaime also added that machine learning algorithms "learns from the humans' writing so if there is a bias that the humans have written it then it will leak into the code." In other words, machine learning algorithms run the risk of encoding social biases. While Nicole is a student whose perspective on data neutrality changed, for Jaime, this activity may have afforded opportunities to extend his understanding of how data is not neutral by providing examples of how data is a social construct.

Deepen and Revise Thinking at the Group Level

A similar cycle of revision would happen when students worked in groups, especially with their group project members. For example, before Class 4, students were asked to respond to four prompts related to their project, including statistical questions, shown in Figure 4.3. This prompt was submitted individually but students were encouraged to work with other members in their group.

Students drafted their research questions before learning about the characteristics of statistical questions or anti-deficit framings of research questions, partially so that they could track their learning. After learning about statistical questions, anti-deficit framings in Class 4, and adding the guiding questions from Figure 4.3 as a class, students worked in groups to rewrite their statistical questions so that there is an explicit anti-deficit framing. Part of this activity included clarifying how their research question used an anti-deficit framing, if the question was about individuals, structures, or both, why the question was a statistical question, and what type of variation the question was addressing (variability within a group, variability within and across groups as well as covariation, or variability in model fitting).

Deepen and Revise Thinking at the Whole Class Level

Finally, at the classroom level, we often referred to classroom-developed artifacts when reviewing previous responses, clarifying or adding on to initial responses, and reflecting on any changes. For example, after introducing the PPDAC cycle and CRT in Class 3, we had a class discussion about the implications of CRT for the PPDAC cycle. The purpose of this activity was for students to come up with a set of "guiding questions and tips" as they engage with the PPDAC cycle during their project using a critical perspective. The student collected responses are shown in black in Figure 4.8.

Figure 4.8: "Guiding questions and tips" as we critically engage with the PPDAC cycle

Problem	**Planning / Data**

• Ask questions about what the data means
• Identify the correct research questions for what you want to look at
• Talk about the role of race and racism in your project and any other types of discrimination
• Ask statistical questions about variation
   ○ Use anti-deficit research questions
• Be clear about your own personal agenda
   ○ Aka state your researcher positionality
Talk about the historical and political context of the topic that you picked
	• Representative sampling
	   ○ Who's included in the data? Who's not included in the data?
	   ○ Who can we generalize the findings to? Not?
	• How are we defining types of data? For example, what is race? Ethnicity? Gender?
	   ○ Use inclusive language in the surveys (e.g., "Do you identify as male or female?" → "What is your gender identity?" and let people fill in the blank
	• Categorical Alignment: For example, how some labels are political or associated with certain things, like race being used to compare groups (Referencing reading)
	If the data is already collected, think of the historical and political context of how and when the data was collected

Analysis	**Conclusion**

• Don't just remove outliers because they might be special cases that you want to know more about
Make sure that categorical data that is inputted as numeric is treated as categorical in R
	• Provide solutions (Include 2-3 action items)
	• Don't intentionally manipulate data or the story to advance your own agenda
	   ○ Especially important since we don't know how people might interpret everything
	• When possible, bring in quotes, stories, or other types of experiential knowledge to contextualize the data
	   ○ Work with participants to give meaning to the data
	   ○ Member Checking
	• Refer to the historical and political context
	• State your researcher positionality
	• Goal: advance social justice
	• Talk about any new questions that you might have
	• Talk about things that weren't perfect in your study (e.g., the sampling) and how you could make it better
	Talk about any tensions that may have come up

Key for when the text was added and question asked that day

Class 3: Based on what we talked about today (PPDAC and CRT), what should you consider during the phases of the PPDAC cycle? Think of this as "guiding questions and tips" as you are engaging with the PPDAC cycle

Class 3 HW, discussed in Class 4: How can QuantCrit be applied to the Problem, Planning, Data, Analysis, and Conclusion (PPDAC) cycle?

Class 7: Based on the last few lessons, how can we consider race and racism throughout the PPDAC cycle? Anything we should add or edit?

Class 13: Now that we're done with the class, is there anything that we want to add about QuantCrit, race and racism, or anything else to consider during the PPDAC cycle, especially as you're finishing up your project?

We revisited this task as a class during the introduction of the second part of Class 4 where we debriefed about QuantCrit (shown in red), at the end of unit on statistical design in Class 7 (shown in blue), and at the end of the regression unit in Class 7 (shown in green). The specific prompt for each class is also shown in Figure 4.8. Before each whole-class conversation, students met in breakout rooms to prepare responses for the whole-class conversation.

In addition to adding new information, there were instances where students clarified or added on to what they had previously said. For example, after reading about statistical questions and anti-deficit-oriented research questions, a student added that we should be "clear about your own personal agenda" during the problem phase so that we are not going into the project with a bias. After talking about positionality statements in Class 7, a student clarified that one way to be clear about our agenda and biases is by adding a positionality statement to or after the introduction of our final report.

Similarly, during the debrief at the end of Class 3, a student mentioned that we need to be careful about how we define types of data, thinking particularly of terms like race, ethnicity, and gender. Near the end of the course in Class 13, students revisited this point and added examples of how we could use inclusive language in surveys. For example, they suggested that instead of providing a multiple-choice question that asks "Do you identify as male or female?," we can ask the open-ended question "What is your gender identity?"

Finally, this may have activity afforded opportunities for students to add critical layers to previous thoughts. For example, in Class 4, students mentioned that we should ask statistical questions that consider at least one of the three types of variation discussed in class (variability within a group, within and across groups or covariation, and variability in model fitting). At the end of the class, another student added that we should use anti-deficit statistical questions. This is similar to Jaime's example from Figure 4.7 where he extended his initial thoughts to include a more explicit and critical perspective.

Design Feature 3: Communicate

The third design feature is about supporting opportunities for students to communicate and engage in dialogue with each other. This design feature is motivated by Freire's (1988) notion of dialogue. Freire (1988) describes *dialogue* as a means through which we engage in inquiry about ourselves, our world, and ourselves in the world around us, which entails mutual respect and love. In fact, Martinez (2020) states that dialogue itself may be considered a form of praxis since it includes thinking of what to communicate or question (reflection) and the act of communicating or questioning the world around us (action). Thus, critical consciousness is developed through dialogue. This moves away from a banking model of education which positions students as empty vessels and teachers as responsible for depositing knowledge into students to an approach where teachers and students are working alongside each other to learn about the world around them. In this sense, dialogue serves two purposes: (a) to deepen understanding about our world, and (b) to build community, respect, and love needed to achieve social justice.

For this dissertation, I draw on sociocultural perspectives to interpret communication and dialogue as the language and discourse practices that mediate learning (Forman, 1996). The goal of this design feature is thus to provide discourse structures that support colearning and community building that operate at different levels and time scales. This included the following strategies: (a) using Mathematical Language Routines (MLRs; Zahner et al., 2021a, 2021b; Zwiers et al., 2017) to help guide whole class conversations, (b) incorporating opportunities for students to anonymously contribute to whole class conversations, and (c) incorporating opportunities for students to engage in dialogue with themselves through reflective journals.

How Design Feature 3 Was Enacted in the TE

The guiding question for this design feature was: What discourse structures can support student dialogue with themselves, others, and me (the instructor) that operate at different levels and time scales? Below, I present illustrations of how these strategies were used to support these design features.

Mathematical Language Routines

The first strategy (Mathematical Language Routines; MLRs) is about students engaging in dialogue with each other. This strategy is motivated by Zahner et al.'s (2021a) design principles for promoting discussions in mathematics in secondary multilingual mathematics classrooms. The MLRs are a set of eight language structures used to support, amplify, and develop students' language. Two MLRs that were repeatedly used in the class were: (a) Information Gap, and (b) Collect and Display.

The purpose of the *Information Gap* routine is to create a need for students to communicate with each other (Gibbons, 2002; Zwiers et al., 2017). For this routine, different students or groups of students are given different pieces of larger tasks (e.g., different parts of a problem, different parts of a definition), then collaborated with others to share the information that was given to them (e.g., through a presentation). In the process, students may also ask for additional information from each other. For example, in Class 3 after reviewing CRT as a class but before formally reading about CRT for homework, students were asked to complete the task in Figure 4.9 in their breakout room groups.

Questions my group had about CRT:

-

CRT Tenet, description, and implications for mathematics teaching and learning

CRT Tenets	Description	Implication for mathematics teaching and learning	Implications for statistics and data science
1. The centrality and intersectionality of race and racism			
2. The challenge of dominant ideology			
3. The centrality of experiential knowledge			
4. The interdisciplinary perspective			
5. The commitment to social justice			

Figure 4.9: Information gap for Critical Race Theory tenets description, implications for mathematics teaching and learning, and implications for statistics and data science

There were five breakout rooms, each with two to three students. Each breakout room completed the row for the tenet that responded to their breakout room number. If groups finished early, they started the next row. After about 10 minutes, we came back as a whole class and each group shared what they wrote for each tenet. As students shared, I shared my screen with a blank template of the activity, took notes of what they were saying, and included any other questions that students had. By the end of the activity, we had a class artifact describing each tenet, implications for mathematics teaching and learning, and implications for statistics and data science. This document was revisited in Class 4 after students had read about CRT and QuantCrit after a class discussion about QuantCrit, and in Class 7 while revisiting the PPDAC cycle.

Perhaps the most used MLR was the collect and display routine. The purpose of this routine is to capture student contributions so that students are able to build on each other's reasoning, notice vocabulary or phrases that they can refer back to later, and make connections with other topics or future discussions. For this routine, the teacher will listen for and scribe student contributions, either as they collect contributions during small group interactions (e.g., in breakout rooms) or during whole class conversations and display them publicly to serve as a classroom artifact. For example, I collected student contributions during the Class 3 activity from Figure 4.9, which was referred to and built on during Class 4 and used in Class 7. Similarly, the "guiding questions and tips" as we critically engage with the PPDAC cycle shown in Figure 4.8 was initially collected in Class 3 and revisited, edited, and added on throughout the course.

Anonymous Contributions

The second strategy (anonymous contributions) was included to provide students opportunities to provide honest and sensitive thoughts that may be used to guide whole class or small group conversations. Students were always told when their responses would be shared with the class and given the option to not respond if the responses would be shared. Students were asked if I could share their response if the contribution came from a small group discussion. While being careful to not censor any contributions, I always vetted the anonymous contributions before sharing them as a whole class to ensure that the contributions were not harmful to students and the classroom climate.

For example, in the beginning of Class 3, students were asked to respond to the question "How can data science be used to advance social justice?" and told that their responses would be shared during the whole class debrief. During the break, I reviewed the responses to ensure that the responses would contribute to a productive and brave conversation. After the break, I displayed the contributions and facilitated a whole class conversation about similarities between the responses and how we might expand on some of these contributions during class.

Later in the term when there was a stronger classroom community, I also asked a student if I could share their response to the Class 6 homework question "One of your colleagues states that ‘technology is politically neutral, therefore data is politically neutral.’ Do you agree or disagree? Explain.” Their response was:

> Before I would have agreed but lately, with the exposure to the subjectivity of data interpretation, I would have to disagree. It’s true that 10 means 10 but then we forget how behind these sets of data is a human being interacting with it and coming to conclusions.

I chose this response for two main reasons: (a) to amplify a response that aligns with the desired anticipated response, and (b) to expand on what we mean by data and technology. First, when asking students if I could share their contributions, I looked for examples that align with the desired anticipated responses (data is not politically neutral) rather than examples that would be categorized as misaligned. The purpose of this selection process was to stay away from publicly shaming or criticizing a specific student which might discourage them from contributing to whole class conversations in the future. This example was also unique in that it highlighted the (re)creation theme in the class where students acknowledged and reflected on any changes in their learning.

The second reason was to have a whole class discussion about common themes in the Class 6 homework. Particularly, eight students stated that they disagreed with the homework statement that technology and data or not politically neutral, one student agreed with the statement that technology and data or politically neutral, and five students stated that technology was politically neutral but data was not. Those five students also differentiated between data as the result of a social process and technological objects (e.g., machines, computers, computer code). I was curious to know more about why students differentiated between technology and data, so I used this example to expand what we mean by “10 means 10”, the role of humans interacting with data and technology, and the neutrality of technological objects like algorithms and machines. The latter was especially important for me because I struggled to articulate how technological objects were perceived as politically neutral (e.g., a computer is perceived as a politically neutral object, computer programming languages may not carry explicit political views). However, I also saw technological objects as extensions of our cognition, so then they also carry our own biases which are shaped by our social, cultural, historical, and political experiences.

Journals

Finally, the third activity (journals) provided students an opportunity to engage in dialogue with themselves. Furthermore, given that the course was developed without the students (most of the lessons were designed before the course started), Frankenstein (1983) notes that journaling provides an opportunity for students to be involved in the lesson planning process by providing feedback as the learning occurs, allowing me to modify lessons and incorporate student feedback as needed. The journals were a combination of writing prompts aimed to target pre-service mathematics teachers' (PSMTs') statistical knowledge for teaching (Groth, 2012) and Critical Reflectivity Journal (CRJ; Fernández & Magaña Gamero, 2018). Groth (2012) included questions where PSMTs were asked to write their own scenarios for a statistics or data science problem, reflecting on teacher-oriented articles, reacting to sample student work, describing activities that may press for student reasoning, or other writing prompts relating to teaching and learning statistics and data science. CRJs aim to "facilitate students' reflexivity, critical social analysis, and engagement with course topics" (p. 19). CRJ prompts include reactions to class discussions and lectures, course materials, or responses to lived experiences. Combined, the writing prompts and CRJs helped facilitate PSMTs identity formation and reflexivity towards critical consciousness through statistics and data science.

The journal prompts used in the course were generally about using data science for social justice. For example, the Class 3 and 6 homework prompts from Figure 4.7 were about data neutrality. There were other prompts that asked students to reflect on the role of race and racism in the PPDAC cycle that supported the classroom artifact in Figure 4.8 (Class 3 and 4 homework), about the implications of QuantCrit to their course project (Class 3 and 6 homework), about the role of humans in data analysis (Class 1 pre survey, Class 4 and 6 homework, Class 13 post survey). Most of the journal prompts either asked students to react to their thoughts on a given topic or quote from a classroom resource or to react to a hypothetical scenario (e.g., "A student said…", "A friend tells you…"). Between lessons, I often referred back to these prompts to identify topics that I may want to expand on or clarify.

Design Feature 4: Engage with Relevant Contexts

Providing opportunities for students to identify generative themes that are engaging and relevant contexts was the first design feature related to the action component of praxis. Generative themes are topics that are (a) engaging and relevant because they are related to and of interest the students' personal experiences, (b) motivate discussions about the social, political, cultural, and historical contexts, and (c) may lead to action (Berry et al., 2020; Freire, 1988; Gutstein, 2012).

The first part is about engaging students by incorporating culturally relevant experiences into the classroom. This approach to teaching calls for bridging students' home experiences and funds of knowledge (González et al. 2005; Moll et al., 2005) to their experiences in the classroom. As a result, a goal is for students to see the value of bringing in their own culture to make sense of what they are learning. One point worth noting is that what is relevant for one student might not be relevant for another student. Furthermore, as instructors, we are often assuming what is relevant and why. Thus, there is a question about who is this culturally relevant for, according to whom, and why?

While culturally relevant experiences foster engagement with the content, the second part is about using generative themes to motivate discussions about the problem context. Although not from a critical perspective, Visnovska and Cobb (2019) note that problem contexts may motivate a need to analyze data and generate discussions about the data investigation process. These discussions, or data generation discussions "encompasses the phases of statistical activity related to problem formulation and planning how the data will be generated" (Visnovska & Cobb, 2019, p. 289), such as how the target phenomenon should be shared, if data can be used to answer a problem, or other questions about generating data. Furthermore, these phases of statistical activity are an important part of the entire statistical investigation cycle, yet they are often removed from statistics classrooms and therefore provide limited opportunities for students to engage with authentic statistical learning (Batanero & Diaz, 2010; Shaughnessy, 2007; Visnovska & Cobb, 2019). In the TE, I extend this by using the problem context to discuss social, political, cultural, and historical underpinnings of the data.

Finally, the third part about generative themes is they may lead to actions that challenge social injustices. As mentioned above, these actions are rooted in an understanding of the social, political, cultural, and historical contexts of the social injustices (reflection). In fact, this emphasis on action, and action rooted in reflection, is what differentiates a generative theme from a general or realistic problem context.

How Design Feature 4 Was Enacted in the TE

The guiding question for this design feature was: how can I select data that are related to and interest to the students' personal experiences, motivated data generation discussions, and generate discussions that may lead to social justice? There were two data sets used in the class: (a) a main data set that would be used throughout the course, and (b) data sets for each group project. Since each student selected their own data sets, I focus on the data set used for the class in this section.

I looked for two criteria when finding data: (a) educational data related to equity, (b) ill-structured data. Given that this course was designed for pre-service mathematics teachers, I was initially interested in education data related to equity. I also wanted to use ill-structured or "messy" data, or data that contains incomplete, incorrectly formatted, or other forms of corrupted data. The purpose of using ill-structured data is to use data that is more reflective of real-world data and to incorporate a lesson on cleaning data using the R programming language (Class 5). Using the two criteria, I used educational data about schools in the local city, including demographic percentages, aggregated scores on standardized assessments, total enrollment, and funds allocated for faculty and staff. The data came from three different resources, but I merged all the data prior to the course. Furthermore, since I downloaded the data from the California Department of Education and Ed-Data, the data was relatively clean (at least relative to data that I have used in my own work). However, to add ill-structured data, I modified some of the data to include some common cleaning examples. These modifications included having percentages that included the "%" character, having some categorical variables coded as numeric (e.g., 1 or 0 to represent if the school was labeled as Title 1 school or not), adding fake schools that did not have any data besides the school name, misspellings in some of the categorical data, and column labels that were unclear.

The goal of using this data was to guide conversations about how data (e.g., standardized assessment scores) are the result of a human process, including how data analysis and conclusions are only as good as the measurement tool, how we can interrogate educational equity from a quantitative perspective without "gap-gazing," and the importance of including experiential knowledge to interpret data analysis and conclusions (Castillo & Gillborn, 2022; Crawford et al., 2018; Covarrubias et al., 2018; Gillborn et al., 2018;). However, after reviewing the Class 6 homework about algorithmic bias and during the debrief in Class 7, I noticed that students were more interested in the facial recognition data that illustrated how facial recognition algorithms more accurately detected White faces than Black faces.

Although I never asked the students to compare the educational data and facial recognition data, I hypothesized that the facial recognition data was more engaging for at least two reasons: (a) the sampling was more transparent and measures that were more straightforward (e.g., pixel color instead of standardized assessments), and (b) the measurement tool and perceived neutrality of facial recognition data. In terms of sampling, the school data included traditional measurement tools that were used to collect data from all the schools in the city. Thus, while we could critique the sampling method and measurement tools (mainly, bias in standardized assessments), there was a sense of agency that was lost when thinking about how we can assess educational inequalities using standardized quantitative data that was already collected. In fact, besides creating a new (and possibly impossible) standardized assessment that accurately measured educational inequity, many of the suggestions for improving educational inequity were about collecting qualitative data (e.g., surveys, interviews) with members of the students' family, teachers, and community to provide a more holistic assessment of educational equity. While this data is necessary, it did not lend itself well for a quantitative analysis and the discussions in this class.

On the other hand, while facial recognition algorithms are also something that most of us have experience with, the facial recognition data example also showed a clear path for social justice: sample a more diverse and representative population. Furthermore, at first, the facial recognition data examples may seem funny or apolitical, but the Class 6 homework illustrated some of the important implications for facial recognition and its use in policing. Thus, the facial recognition data provided a relatable example that may be perceived as apolitical, illustrates how data science can be weaponized if algorithmic biases are not addressed, but also provides a sense of agency by having concrete examples of how we can address algorithmic bias in pursuit of social justice and an entry point into discussing how something that appears so apolitical is in fact very political.

Design Feature 5: Engage with All Phases of the Statistical Investigation Cycle

The fifth design feature was about incorporating opportunities for students to engage with and discuss how race and racism are embedded into the Problem-Planning-Data-Analysis-Conclusion (PPDAC, Wild & Pfannkuch, 1999) cycle. There are two goals with this design feature: providing opportunities for students to develop (a) statistical and data scientific practices and (b) critical statistical and data scientific practices. The first goal is motivated by research that states statistical practices are developed by engaging with the entire PPDAC statistical investigation process that develops statistical practices (Lee & Tran, 2015). This includes the problem and planning phase that are too often removed from introductory statistics courses (Visnovska & Cobb, 2019).

The second goal extends the first goal by foregrounding the role of race and racism in the PPDAC cycle. This is similar to how Weiland (2017) differentiates between statistical literacies and critical literacies. In fact, I believe that critical statistical practices are the same as statistical practices (or literacies), but differentiate them in this dissertation to emphasize the importance of foregrounding the role of race and racism in statistics and data science. That is, all critical statistical practices may be interpreted as statistical practices, but not all statistical practices may be interpreted as critical practices.

How Design Feature 5 Was Enacted in the TE

The guiding questions for this design feature were: how can we talk about and engage with the PPDAC cycle? How can we expand on the PPDAC cycle to consider race and racism in the PPDAC cycle? There were three main ways in which I aimed to enact this design feature: (a) designing the course around the PPDAC cycle, (b) revisiting the “Guiding questions and tips” as we critically engage with the PPDAC cycle (Figure 4.8), and (c) embedding a course project throughout the course that allowed students to apply what we talked about and learned in class to a project of their choosing. Since the course project was a large portion of the course, that is discussed in more detail in the section on Design Feature 6.

The course sequencing was scheduled around the PPDAC cycle. A brief overview of the curriculum map is shown in Table 3.3, including the class topic and project homework. There were four main units: (a) introduction, (b) study design, (c) regression, and (d) course summary and project. The purpose of including these four main units was to provide a unit that focused on one or two phases of the PPDAC cycle at a time and to include opportunities for students to build on their project as they are learning about different parts of the PPDAC cycle.

The introduction unit aimed to set the foundations for the class. This included an introduction to statistics, R programming in Google Colab, and to the PPDAC statistical investigation cycle. At the end of the unit, we focused on the Problem phase in the PPDAC, specifically using CRT and QuantCrit as a lens to talk about the social, political, cultural, and historical contexts of data. This unit is also when the “Guiding questions and tips” for the PPDAC cycle classroom artifact was created (Figure 4.8).

The second unit was about the planning and data phase in the PPDAC cycle. This included lessons on sampling (random and not random) and statistical designs (experimental, observational, and survey). Race and racism were discussed throughout the unit. For example, after talking about sampling, we talked about how the under- and overrepresentation of certain groups of people in training data may influence who we are able to generalize to (e.g., facial recognition data). Furthermore, students talked about how we can write more inclusive survey questions, categorical alignment, and considering the racialized context of historical data or data that has already been collected (e.g., policing data).

The third unit of the class was on the analysis phase of the PPDAC cycle. Specifically, we focused on regression analysis (simple, multiple), model analytics (e.g., prediction accuracy, squared errors), and model comparison (e.g., errors, prediction, cross validation). We also briefly introduced decision trees, random forests, and parametric tests. Race and racism were discussed at the end of the unit. For example, when reflecting on the implications of QuantCrit to data science, students noted that we should be careful to not interpret any causations between demographic markers and outcomes (e.g., saying “the proportion of BIPOC students is correlated to the school’s average standardized assessments) without considering the larger political context.

Notably, as a course designer, the analysis unit was the most challenging to incorporate opportunities to discuss race and racism. I believe that this may be partially attributed to how model building and comparison may be seen as the most apolitical part of the PPDAC cycle if it is interpreted as applying formulas and algorithms to data. That is, building a regression model may be seen as the most neutral activity of the PPDAC cycle, especially when compared to other parts of the PPDAC cycle. This is especially noteworthy when many introduction statistics courses often focus on applying formulas and algorithms to simplified sets of data (Bargagliotti & Franklin, 2015; Franklin, 2013; Garfield & Ben-Zvi, 2008), or the analysis part of the PPDAC cycle. On one hand, the perceived neutrality of the analysis phase is an argument for including other phases in statistics and data science courses, mainly since discussing race and racism was more clearly evident in the other units. On the other hand, this perceived neutrality encouraged me to look further into the politicization of statistical analysis and computer programming to find other ways to incorporate race and racism in the analysis phase of the PPDAC cycle.

The final unit was about the conclusion phase of the PPDAC cycle. Race and racism were discussed throughout the entire unit. For example, when reflecting on the implications for QuantCrit, students noted that the overall goal should always be to advance social justice in whatever topic each group chose. This included considering a member check in to highlight experiential knowledge of the participants in the study, including two to three action items in the conclusion, and stating any possible avenues for future work that may continue to advocate for social justice.

Design Feature 6: Design and Implement Statistical Study Throughout the Course

The sixth and final design feature was including a course project on a social justice topic of the students' choosing. The goal of the course project was for students to apply their knowledge throughout the course by designing their own statistical study, including collecting, cleaning, and analyzing their own data (Chance, 1997, 2002), instead of using a prescribed well-structured data. Furthermore and in alignment with the goal of writing the world with statistics and praxis, it was important that students are given the opportunity to engage in actions that challenge and learn how others challenge social injustices to create transformative change (Bartell, 2013; Gonzalez, 2009). In doing so, course projects may help raise awareness about social injustices (e.g., Kokka, 2020; Tate, 1995; Turner & Strawhun, 2005) and apply knowledge to advance social justice.

How Design Feature 6 Was Enacted in the TE

The guiding question for this design feature was: How can we support course projects in which students investigate a social justice issue using methods from data science? To address this, I drew on Chance's (1997) recommendation for classroom projects in statistics classrooms: (a) integrating the project and course, (b) providing students with timely and constructive feedback, and (c) providing students with guidelines and expectations.

Integrating the Project and Course

The course project was introduced in the first lesson, including the project description, project rubric, and resources for writing the data analysis report. Details of the project were revisited across the course, shown in Table 3.3. For example, students worked on different components of their introduction (motivation, research questions, positionality) in the first four classes. The guiding questions from the Class 3 homework (Figure 4.3) and the implications from QuantCrit on those questions (Figure 4.5) were used to help students edit the draft of their introduction, which was submitted before Class 5.

In the study design unit, students worked on a draft of their methods section. This included identifying and justifying sampling techniques (Class 6), discussing generalizations (Class 6), drafting survey questions (Class 5), responding to instructor (Class 7) and peer (Class 8) feedback on the surveys, and selecting and justifying appropriate survey methods (Class 8) before beginning the data collection process. At the end of Class 7, students were also given a sample report that they were able to use as mentor text. A similar process to the guiding questions from Figure 4.3 and QuantCrit implications from Figure 4.5 was done at the end of Class 7, shown in Figure 4.10.

1. Positionality (Tenet 1: Centrality of Race and Racism, Tenet 5: Voice and insight, data cannot 'speak for itself')
    a. How do my personal, professional and/or intellectual positionalities (identities, contexts, experiences, and perspectives) cohere with or diverge from my research inquiries?
    b. What legacies (personal, communal, societal, national, transnational and/or global) inform the social contractedness of my positionality?
    c. In what ways, or not, am I conscientiously, or not, reifying, resisting, disrupting, and/or changing the constructs of my positionality through this research process?
    d. How has my own positionality changed, or not, over time, and why? In what ways has it remained static, and why? In what ways has it been dynamic, fluid, emerging and/or generative, and why?
    e. How does my positionality recognize, honor, and/or problematize intersectional notions of difference (politics, economics class, race, ethnicity, nationality, citizenship, legality, age, ability, education, sexuality, gender, and/or religion?) as a conceptual praxis of analysis for my research context?
2. Data collection tools (Tenet 2: Numbers are not neutral)
    a. Data collection tools
        i. Where, how, and when (date/year) was the data collected?
        ii. Who collected the data (researchers, community members, etc.)?
    b. Population
        i. What is your population of interest? Why?
        ii. Who is and is not in the sample?
        iii. Identify the number of people who will be in your sample and why this number is appropriate
        iv. Describe how you plan on selecting your sample from the population so that you'll be able to make generalizations about your population of interest
2. What is your research question? (Tenet 3: Categories are not natural)
    a. Data collection tools
        i. What measures/outcome/surveys were collected? How?
        ii. Are the variables defined and justified?
        iii. How does the data relate to your research question / goal?

Figure 4.10: QuantCrit implications for course project from Class 7 (black text was provided, students added the blue text)

I provided the black text, and the blue text was added as students contributed to a whole class discussion. The text added for the positionality bullet was found by a student online and identified as an important resource that could be used for this course project (Weingarten Learning Resources Center, 2017).

In the regression unit, students focused on the analysis portion of their final project. During Class 9, we reviewed the course project goals, rubric, and started a “Running Questions” document that was revisited in Class 11 and Class 13. Some of the questions included:

- What if the data didn’t show what we expected?
- Would ____ work as a connection to QuantCrit?
- What if we wish we would’ve reworded survey questions?
- What is cleaning data? Do we need to write about this too?
- What if there weren’t any “significant” findings?
- What about outliers?
- What if the sample size is small?
- What if we had to use convenience sampling?

All questions were addressed during whole-class discussions. Once most (or all) of the data was collected, students worked in groups to clean data (Class 12) and analyze the data (Class 13 and on).

As part of their analysis, the students were asked to include a brief description of their exploratory data analysis (using R programming tools from Class 2 or Class 5) and to include two forms of analysis, one of which must be one of the regression models learned in class and the other that could be of their choosing (either from the class, from other analysis methods provided during Class 1, or another one of their choosing). Each analysis method also had to have at least one visual (using R programming tools from Class 5).

Finally, the project unit focused on drafting the conclusion chapter for their final data analysis report. At this point, we revisited the implications of QuantCrit for the data analysis report as well as the "Guiding questions and tips" as we critically engage with the PPDAC cycle (Figure 4.8) and the course rubric.

Providing Students with Timely Feedback

A second recommendation by Chance (1997) is to provide students with timely and constructive feedback. The goal is to promote the dialogue between the instructor and students, as well as between students. In doing so, students engaged in a form of a Stronger and Clearer (Zwiers et al., 2017) routine where they are able to revise and resubmit portions of their final data analysis report. For example, students submitted their research questions and justified why the research question was a statistical question and an anti-deficit question in Class 4. Students received feedback by the end of Class 5 so they can include the final research questions in the introduction draft due by Class 6. A similar process was done with the survey questions, where students submitted their initial survey questions, received feedback from the instructor, edited the survey questions, received feedback from their peers, and edited their final survey questions before distributing the survey. Feedback for the entire introduction, methods, and analysis sections of their final report was also provided within two classes of the student submission.

Clear Expectations and Guidelines

A third recommendation by Chance (1997) is to provide clear expectations and guidelines for the final project. In addition to providing the project description, project rubric, and resources for writing the data analysis report during Class 1, we revisited the rubric in Class 9 and Class 13, viewed a sample data analysis report that could be used as a mentor text in Class 7, and started a running list of project questions during Class 9 that students were able to use as they finalized their project. Groups also set times to meet with me during Class 14 and 15 to answer any additional questions and get feedback before their final data analysis report was submitted.

Summary of Design Features

I end this chapter by summarizing the design features and highlighting their interconnectedness. There were six design features that were motivated by Freire's notion of praxis (three related to reflection and three related to action) and QuantCrit's emphasis on the centrality of race and racism in data. The three design features related to reflection were about providing students opportunities to: (a) reflect on structures of injustices, (b) deepen and revise their thinking, and (c) communicate with each other. Combined, these three design features aim to focus students' attention on the social, cultural, historical, and political understandings of social injustices. For example, the first design feature (DF1: reflect on structures of injustices) encourages students to account for larger systemic or structural causes of injustices in an effort to minimize gap-gazing or avoid reifying deficit narratives. The second design feature (DF2: deepen and revise thinking) builds on Freire's assumption that critical consciousness is an ongoing cyclical process, which implies a sense of learning, relearning (learning something again, possibly clarifying or strengthening previous knowledge), and unlearning (modifying or editing knowledge). This is especially important in fields like mathematics or social justice where there is a perceived high social cost for sharing incomplete or "rough draft" ideas (Jansen, 2020). In doing so, my goal was to provide a learning model that normalizes growth and involves changing ideas. The third design feature (DF3: communicate with each other) builds on Freire's notion of dialogue and Vygotsky's notion of language as a mediator for learning and is intended to provide students opportunities to learn about them, others, and their relation to the world around them.

The remaining three design features aim to provide students opportunities to advance social change using statistics and data science. Particularly, the fourth design feature (DF4: engage with relevant contexts) is about providing students opportunities to engage with problem contexts that are relevant to them, motivate a need for data analysis, and provide avenues for social change. The fifth design feature (DF5: engage with all phases of the statistical investigation cycle) aims to provide students opportunities to learn about the problem, planning, data collection, analysis, and conclusion phases of engaging with data (rather than only focusing on the analysis stage or an individual part of the cycle at a time). Finally, the last design feature (DF6: design and implement statistical study throughout the course) provides opportunities for students to concurrently apply what they learn in the course on their own study.

While these design features may not be inclusive of all the features that course designers may consider when designing data science for social justice courses, I argue that they are interconnected and dependent on each other. For example, first design feature (DF1: reflect on structures of injustices) helped guide how we engaged with the statistical investigation cycle (DF5: engage with all phases of the statistical investigation cycle), such as by creating anti-deficit research and statistical questions in the problem phase, asking questions about representation and what data aims to measure in the planning and data phase, and contextualize the findings in the larger social, cultural, political, and historical context in the conclusion phase. Similarly, in a post-interview, a student suggested that having anonymous contributions (DF3: communicate with each other) may have helped create and brave and safe space where they “felt comfortable not knowing…but knowing that I will continue to learn,” related to the design feature about deepening and revising thinking (DF2). Finally, using data that was relevant and meaningful to students (DF4) may have encouraged students to draw on their own experiential knowledge about the context that may have helped situate the social injustice in a larger context (DF1) and motivated a need for using data to advance social and racial justice in their course project (DF 6).

Chapter 5: RQ2 (Statistical and Data Scientific Content Knowledge)

This research question aims to capture the development of students' traditional statistical content knowledge in the TE, focusing on their understanding of study design and regression. Particularly, the research questions are:

> **RQ2a**: What was the effect of the teaching experiment (TE) on statistical content knowledge as measured by the student response patterns on curriculum-aligned assessments?
> **RQ2b**: How did the response patterns by question type (e.g., by question, multiple choice or free response) vary across the TE?

This research question is in alignment with the TMSJ framework's goal of "succeeding in a traditional sense" (Gutstein, 2006, p. 41) -- i.e., teaching mathematics content. While it is important to recognize that not all learning may be captured in a standardized assessment, assessments may help provide an insight into what students learned in the course.

Data for this analysis includes pre- and post-assessments for the study design and regression units. Both assessments were assigned as part of the homework prior to the respective unit and after the respective unit. All questions came from the Comprehensive Assessment of Outcomes in Statistics (CAOS; Delmas et al., 2007) or the Levels of Conceptual Understanding in Statistics (LOCUS; Jacobbe et al., 2014) assessments, two of the leading assessments of student understanding in statistics courses. The free response questions from LOCUS assessment also provide a grading rubric that was used for this dissertation to guard against biases when grading. The grading rubrics are presented in this chapter. All data were anonymized before I analyzed the data. In some cases, there were minor changes from the pre- to post-assessments, mainly changing examples (e.g., tomatoes in the pre-assessment and potatoes in the post-assessment) and the numbers used in the problems. The goal of having similar pre- and post-assessment questions was to allow for comparisons across the individual questions as well.

There were three main parts to the analysis, one corresponding to each part of the research question. The first analysis focuses on overall gains across the pre-post assessments. This includes a paired t-test, paired Wilcoxon signed-rank test, Cohen's D effect size (Cohen, 2013; Hedges, 1981), and single-student normalized gains (Chance et al., 2016; Hake, 2002). The purpose of using multiple statistical measures that aim to measure a similar phenomenon is to provide multiple sources of evidence and, potentially, find similarities or differences across the statistical measures. Similar statistical findings could suggest further evidence for the particular finding, while different statistical findings would provide mixed or weaker evidence for the particular finding. Despite the assessments being created by outside researchers, I anticipated to see growth from the pre- to the post-assessments since the topics generally aligned to the TE curriculum. The second part of the analysis breaks down the analysis by question type (multiple choice or free response) and by question. The goal was to help identify any patterns within the types of questions (e.g., was there more growth on the multiple choice or free response questions?). Similar statistics are used as the first part. Finally, I analyzed sample responses to the free response questions to provide further insight into response patterns. This process is repeated for both the study design and regression units.

Unit on Study Design

The unit on study design included sampling, randomization, and statistical study designs (surveys, experiment, and observational). These topics were chosen because they are foundational topics in statistics and afforded opportunities to discuss the biases they may occur during the planning and data phases of the PPDAC cycle (i.e., before the analysis). Twelve students that took the pre- and post-assessment.

Total Pre-Post Gains

I began by looking at overall gains from the pre- to post-assessment. Figure 5.1 shows boxplots of the pre- and post-assessments for the study design unit. Individual student scores are shown by the points, the mean is shown by the dashed line, and the median is shown by the solid line.

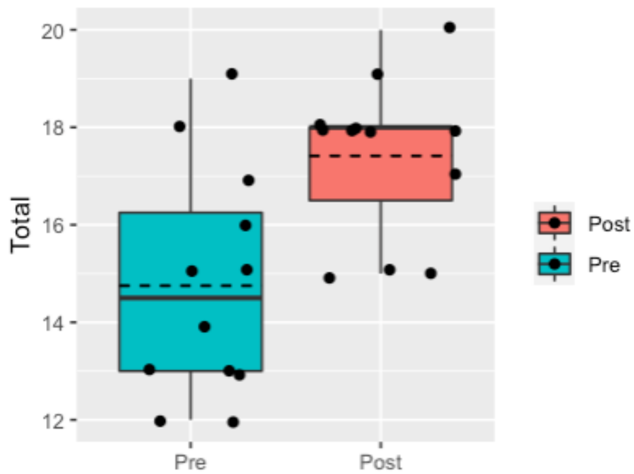


Figure 5.1: Boxplot of pre- and post-assessment scores in the study design unit

Overall, there was an increase in the mean and median from the pre- to the post-assessment. Particularly, Table 5.1 shows the pre and post means, standard deviations, paired t-test, Cohen's effect size, and normalized gains. The mean increased from about 14.75 (about 67%) to about 17.42 (about 79%), a 2.67-point difference (about 12%). In terms of statistical tests, the average gains score was small. However, the paired t-test, effect size, and gain of average suggest that there was a significant increase. Specifically, the paired t-test p-value may be interpreted as a statistically significant p-value, suggesting that the difference in means from the pre- to the post-assessment did not occur by chance. Furthermore, the effect size was large and the normalized gain was medium, suggesting that there is also a practical significance in the pre-post gains. In terms of the median, the paired Wilcoxon sign rank test p-value may also be interpreted as a statistically significant increase from 14.5 in the pre-assessment (about 66%) to

Table 5.1: Paired t-test, Cohen's effect size, and normalized gains in the study design unit for the total and by question type

Note:
[1] $p<0.5$*, $p<0.01$**, $p<0.001$***, $p<0.0001$****
[2] $|d|<0.5$ small, $|d|<0.8$ medium, otherwise large
[3] $g<0.3$ low, $g <=0.7$ medium, otherwise large (Hake, 1999)

		Pre		Post		Diff in Means	Paired T-Test			Cohen's Effect Size		Normalized Gains	
	Points	Mean	SD	Mean	SD		t statistic	p-value[1]	95% CI	d[2]	95% CI	Gain of Avg[3]	Avg Gains[3]
Total	22	0.67	(0.11)	0.79	(0.07)	0.13	0.12	0.012*	(0.73, 4.61)	1.33 (L)	(0.73, 2.58)	0.58 (M)	0.27 (S)
Multiple Choice	14	0.73	(0.11)	0.83	(0.13)	0.10	2.33	0.040*	(0.08, 2.76)	0.86 (L)	(0.04, 1.77)	0.61 (M)	0.32 (S)
Free Response	8	0.56	(0.20)	0.72	(0.03)	0.16	3.56	0.005**	(0.48, 2.02)	0.70 (M)	(0.04, 1.16)	0.56 (M)	0.22 (S)

18 in the post-assessment (about 82%), shown in Table 5.2. That is, about half of the students had a 66% or higher in the pre-assessment, but about half of the students had an 82% or higher in the post-assessment.

Table 5.2: Paired Wilcoxon the total score in the study design unit for the total and by question type

					Paired Wilcoxon Sign Rank		
	Points	Pre Median	Post Median	Diff in Medians	V statistic	p-value	95% CI
Total	22	14.5	18	3.5	60	0.018*	(1, 5)
Multiple Choice	14	10.5	12	1.5	46.5	0.057	(0, 3.5)
Free Response	8	4	6	2	43.5	0.012*	(0.50, 2.50)

Note:
For the paired Wilcoxon: $p<0.5^{*}$, $p<0.01^{**}$, $p<0.001^{***}$, $p<0.0001^{****}$

Furthermore, almost all students had an increase from the pre- to the post-assessment. There were three students who had a decrease, each of one point and each of which had relatively high scores. Each of these students also had an 18 or higher on the pre-assessment. One student also had the same score on the pre- and post-assessment (also a score of 18).

Gains by Question Type (Free Response or Multiple Choice)

I conducted a similar analysis to the total pre-post gains with the question type (free response or multiple choice). The goal of this analysis was to see if the gains varied by question type. Figure 5.2 shows a boxplot for the scores on the free response (left, eight points total) and multiple choice (right, 14 points total) questions in the pre- and post-assessments. Individual student scores are shown by the points, the mean is shown by the dashed line, and the median is shown by the solid line. Table 5.1 also shows the pre and post means, standard deviations, paired t-test, Cohen's effect size, and normalized gains. Table 5.2 shows the pre-post medians and paired Wilcoxon Sign Rank test. The results were similar to the total pre-post gains, with the paired t-test, paired Wilcoxon sign rank test, Cohen's D, and gain of average suggesting that there was some significant and practical growth from the pre- to the post-assessments.

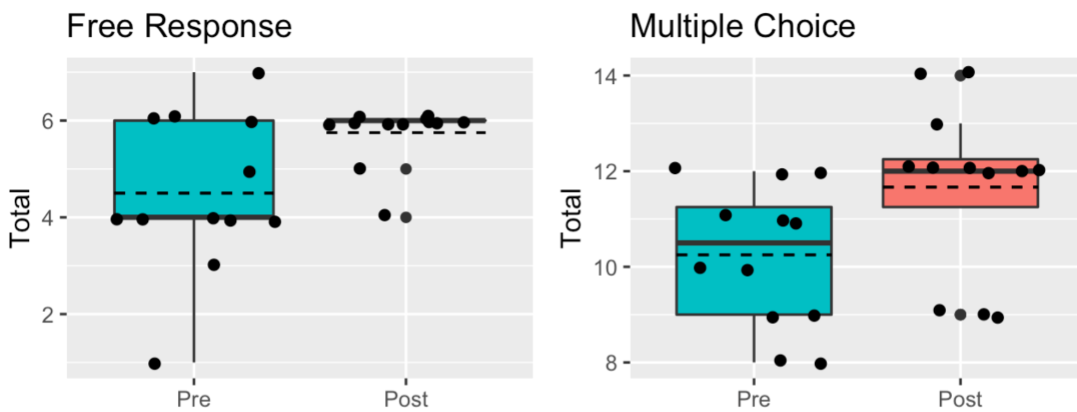


Figure 5.2: Boxplot for the free response (left) and multiple choice (right) questions of the study design unit

The free response questions were out of eight total points. Students had about 56% free response questions correct in the pre-assessment but 72% in the post-assessment, about a 16% increase. Notably, the standard deviation also decreased from about 20% in the pre-assessment to about 3% in the post-assessment, suggesting that the data was more concentrated around the mean in the post-assessment. The median also increased from about four (about 50%) to six points (about 75%), a two-point increase (about a 25% increase). The average gains were small, but the paired t-test and paired Wilcoxon sign rank test had p-values that may be interpreted as statistically significant (i.e., below 0.05) and a medium Cohen's effect size and gains of average. Thus, the free response questions had a statistically significant and practical increase. As a reminder, the sample size in this dissertation is relatively small. Future studies may aim to replicate this study with a larger sample. Nonetheless, the statistical tests may help provide an insight into the types of learning that occurred in the TE.

The multiple-choice questions had a similar increase than the free response questions. Particularly, students had an average of about 73% of the multiple-choice questions correct in the pre-assessment and an average of about 83% correct in the post-assessment, about a 10% increase. The median also increased, from about 10.5 (about 75%) to 12 (about 86%), a 1.5

(about 11%) increase. The paired t-test p-value (0.040) and paired Wilcoxon Sign Rank test p-value (0.057) were relatively large and the average of gains was small (implying a slightly significant increase), but there was a large effect size and medium gain of average (implying a practical increase). The mixed results may be indicative of the small sample size or because the students started off with a relatively high score in the pre-assessment. Nonetheless, all the results suggest that there was a meaningful increase from the pre- to the post-assessment.

Gains by Question

Finally, I looked at the overall gains by question. The goal of this analysis was to identify any specific topics that may have had varied growth (e.g., more growth on questions about sampling than about study design). All multiple-choice questions were one point. All free response questions had two parts, each of which were two points. Table 5.3 also shows the pre and post means, standard deviations, paired t-test, and Cohen's effect size by question and Table 5.4 shows the pre-post medians and paired Wilcoxon Sign Rank test by question.

There were four prompts that had no growth: Q1, Q5, Q8, and Q11. The first three were about sampling and all students answered them correctly in the pre- and post-assessment. The fourth question was about study design, and 11 (about 92%) of the students answered the question correctly in the pre-assessment and the same 11 answered the question correctly in the post-assessment. There were two questions that had a decrease from the pre- to post-assessment. Neither of the questions had a statistically or practically significant decrease. Both of the questions were related to sampling. In terms of questions that had significant growth, Q4, Q10, Q15a, Q15b, and Q16a all had a medium or large effect size, implying that there was practical growth. All of these questions were also related to study design. That is, students showed more

Table 5.3: Paired t-test, Cohen's effect size, and normalized gains in the study design unit by question

				Pre		Post		Means Diff	Paired T-Test			Cohen's Eff Size	
Q	Q Type	Topic	Pts	Mean	SD	Mean	SD		t-stat	p-value[1]	95% CI	d[2]	95% CI
1	Multiple Choice	Sampling	1	1	(0.00)	1	(0.00)	0	-	-	-	-	
2	Multiple Choice	Sampling	1	0.92	(0.29)	1	(0.00)	0.08	1	0.3388	(0.10, 0.27)	0.41 (S)	(0.47, 1.29)
3	Multiple Choice	Study Design	1	0.67	(0.49)	0.75	(0.45)	0.08	0.56	0.5863	(0.24, 0.41)	0.18 (S)	(0.48, 0.83)
4	Multiple Choice	Study Design Sampling	1	0.25	(0.45)	0.75	(0.45)	0.5	3.32	0.0069**	(0.17, 0.83)	1.11 (L)	(0.23, 1.98)
5	Multiple Choice	Sampling	1	1	(0.00)	1	(0.00)	0	-	-	-	-	-
6	Multiple Choice	Variation	1	0.33	(0.49)	0.58	(0.51)	0.25	1.39	0.1911	(0.14, 0.64)	0.50 (S)	(0.29, 1.28)
7	Multiple Choice	Random Assignment	1	0.58	(0.51)	0.67	(0.49)	0.08	0.56	0.5863	(0.24, 0.41)	0.17 (S)	(0.45, 0.78)
8	Multiple Choice	Sampling	1	1	(0.00)	1	(0.00)	0	-	-	-	-	-
9	Multiple Choice	Sampling	1	0.92	(0.29)	1	(0.00)	0.08	1	0.3388	(0.10, 0.27)	0.41 (S)	(0.47, 1.29)
10	Multiple Choice	Study Design Sampling	1	0.33	(0.49)	0.67	(0.49)	0.33	2.35	0.0388*	(0.02, 0.65)	0.68 (M)	(0.01, 1.34)
11	Multiple Choice	Study Design	1	0.92	(0.29)	0.92	(0.29)	0	-	-	-	-	-
12	Multiple Choice	Study Design Sampling Random Assignment	1	0.42	(0.51)	0.67	(0.49)	0.25	1.39	0.1911	(0.14, 0.64)	0.50 (S)	(0.29, 1.28)
13	Multiple Choice	Sampling	1	1	(0.00)	0.92	(0.29)	-0.08	-1	0.3388	(-0.27, -0.10)	-0.41 (S)	(-1.29, -0.47)
14	Multiple Choice	Study Design Sampling	1	0.92	(0.29)	0.75	(0.45)	-0.17	-1	0.3388	(-0.53, -0.20)	-0.44 (S)	(-1.40, -0.52)

Table 5.3: Paired t-test, Cohen's effect size, and normalized gains in the study design unit by question, Continued

				Pre		Post		Means Diff	Paired T-Test			Cohen's Eff Size	
Q	Q Type	Topic	Pts	Mean	SD	Mean	SD		t-stat	p-value[1]	95% CI	d[2]	95% CI
15a	Free Response	Study Design Random Assignment	2	0.63	(0.31)	0.88	(0.23)	0.25	2.57	0.0261*	(0.07, 0.93)	0.92 (L)	(0.04, 1.79)
15b	Free Response	Study Design Sampling	2	0.50	(0.00)	0.59	(0.19)	0.09	1.48	0.1661	(0.08, 0.41)	0.61 (M)	(0.32, 1.53)
16a	Free Response	Study Design Random Assignment	2	0.59	(0.42)	0.79	(0.33)	0.21	2.8	0.0172*	(0.09, 0.74)	0.53 (M)	(0.11, 0.94)
16b	Free Response	Study Design Sampling Variation	2	0.54	(0.26)	0.63	(0.23)	0.09	1	0.3388	(0.20, 0.53)	0.34 (S)	(0.39, 1.08)

Table 5.4: Median and paired Wilcoxon Sign Rank test gains in the study design unit by question

							Paired Wilcoxon Sign Rank	
	Question Type	Topic	Points	Pre-Median	Post Median	Diff in Medians	V statistic	p-value[1]
Q1	Multiple Choice	Sampling	1	1	1	0	-	-
Q2	Multiple Choice	Sampling	1	1	1	0	0	1.000
Q3	Multiple Choice	Study Design	1	1	1	0	2	0.773
Q4	Multiple Choice	Study Design, Sampling	1	0	1	1	0	0.020*
Q5	Multiple Choice	Sampling	1	1	1	0	-	-
Q6	Multiple Choice	Sampling Variation	1	0	1	1	3	0.233
Q7	Multiple Choice	Random Assignment	1	1	1	0	2	0.773
Q8	Multiple Choice	Sampling	1	1	1	-	-	-
Q9	Multiple Choice	Sampling	1	1	1	0	0	1.000
Q10	Multiple Choice	Study Design, Sampling	1	0	1	1	0	0.072
Q11	Multiple Choice	Study Design	1	1	1	-	-	-
Q12	Multiple Choice	Study Design, Sampling, Random Assignment	1	0	1	1	3	0.233
Q13	Multiple Choice	Sampling	1	1	1	0	1	1.000
Q14	Multiple Choice	Study Design, Sampling	1	1	1	0	7.5	0.424
Q15a	Free Response	Study Design, Random Assignment	2	1	2	1	4.5	0.041*
Q15b	Free Response	Study Design, Sampling	2	1	1	0	0	0.346
Q16a	Free Response	Study Design, Random Assignment	2	1	2	1	0	0.037*
Q16b	Free Response	Study Design, Sampling Variation	2	1	1	0	2.5	0.423

Note:
[1] p<0.5*, p<0.01**, p<0.001***, p<0.0001***

growth in the study design questions than the questions about variation, sampling, or random assignment.

Study Design Free Response Samples

There were two free response questions as part of this pre- and post-assessment. Both questions were from the LOCUS assessment (Jacobbe et al., 2014). As mentioned in the Methods chapter, each part of the question was two points (four points total per question) and the pre- and post-questions were very similar (minor changes in the quantities or context).

Question 15 (Study Design and Random Assignment)

The first free response question is shown in Table 5.5.

Table 5.5: Question 15 in pre- and post-assessments for the study design unit

Pre-Assessment	Post-Assessment
A department store manager wants to know which of two advertisements is more effective in increasing sales among people who have a credit card with the store. A sample of 100 people will be selected from the 5,300 people who have a credit card with the store. Each person in the sample will be called and read one of the two advertisements. It will then be determined if the credit card holder makes a purchase at the department store within two weeks of receiving the call. (a) Describe the method you would use to determine which credit card holders should be included in the sample. Provide enough detail so that someone else would be able to carry out your method. (b) For each person in the sample, the department store manager will flip a coin. If it lands heads up, advertisement A will be read. If it lands tails up, advertisement B will be read. Why would the manager use this method to decide which advertisement is read to each person?	A department store manager wants to know which of two advertisements is more effective in increasing sales among people who have a credit card with the store. A sample of 50 people will be selected from the 1,300 people who have a credit card with the store. Each person in the sample will be called and read one of the two advertisements. It will then be determined if the credit card holder makes a purchase at the department store within two weeks of receiving the call. (a) Describe the method you would use to determine which credit card holders should be included in the sample. Provide enough detail so that someone else would be able to carry out your method. (b) For each person in the sample, the department store manager will flip a coin. If it lands heads up, advertisement A will be read. If it lands tails up, advertisement B will be read. Why would the manager use this method to decide which advertisement is read to each person?

According to the LOCUS description of the question (Jacobbe et al., 2014), the first question was about recognizing a need for random selection and describing an appropriate method to select a random sample (part a) and explaining why random treatments are important (part b). In terms of grading part a, I used the LOCUS rubric from Jacobbe et al. (2014). Particularly, they stated that (comments added in double parentheses and bold is added for emphasis):

> **Part a:** An ideal response to part (a) **uses random selection** in a planned way to determine which of the store's 5300 credit card holders will be included in the sample. To be considered essentially correct for part (a) the response **must indicate how the random selection will be implemented** ((2 points)).
> A response that indicates the need for random selection, **but which does not provide an adequate description of how the random selection will be accomplished** (for example, a response that just says pick 100 people at random from the list) is considered partially correct ((1 point)).

This part of the question had an increase from the pre-assessment (average of 63%, median of 1) to the post-assessment (average of 88%, median of 2).

Table 5.6 shows sample student responses to part a and the points they were assigned. I chose the pre-assessment samples presented here because students did not always consider randomization (e.g., randomly selecting from the entire population), how randomization will be implemented (e.g., creating two groups), or following up after two weeks to compare across groups (as stated in the prompt). In particular, Robert, Jaime, and Kimberly all stated that there needed to be some randomization component to the sampling, but did not specify a type of randomization or how it would be accomplished. This was common for most of the students in the pre-assessment.

In the post-assessment responses presented here, students mentioned randomization as well as how the randomization process that they would use (e.g., flipping a coin). For example, Robert explicitly stated the he would select people using a "systematically random" sample where every "nth person would be chosen." Notably, he also states that he would limit the sample to people who have a credit card. Jaime and Kimberly also considered randomization, but suggested that she would use a coin to assign people to one of two groups. Although not explicitly stated, flipping a coin to assign groups could result in groups with different sizes (e.g., 20 in the treatment and 30 in the control group). Nonetheless, both of their methods are still appropriate because she randomly sampled from the entire population and randomly assigned groups and the prompt did not ask for equal-sized samples.

Table 5.6: Sample student responses to part a of Q15 (study design and random assignment) in the study design pre- and post-assessments

Student	Pre-Assessment	Points	Post-Assessment	Points
Robert	It should be chosen at random with no previous knowledge about when their last purchase at the store was to create a sense of randomization not a common patron	1	It should be randomly chosen throughout the day. Maybe every nth person should be chosen to make sure that its systematically random, but only count people who have a credit card. Then the first person with a CC goes in one group, then the second in the second group, then the third in the first group and so on	2
Jaime	The sample of 100 credit card holders should be randomly selected. They could choose 100 people on one of the busiest days of the week.	1	Assign a number to every customer with a store credit card and then use a computer to randomly select 50 numbers. Then flip a coin for every person in the group of 50 and assign them to either group A or group B. Call the customers and read them the advertisement. If they don't answer, follow up.	2
Kimberly	To create two groups and figure out which advertisement is more appealing to people	1	The best way to determine who should be in the sample is to have the 50 cardholders be randomly selected from the list of 1,300 people who have a credit card with the store, then make two groups by like flipping a coin. This way would reflect if the sales tactic actually works and bring people in to spend money.	2

In terms of grading part b (why would a manager choose flipping a coin to randomly assign people to decide which advertisement people would receive), I also used the LOCUS rubric from Jacobbe et al. (2014). Particularly, they stated that (comments added in double parentheses):

> **Part b:** An ideal response to part (b) provides an explanation that indicates that the random assignment to treatments in the experiment described **allows the store manager to conclude that which advertisement is read is the cause of any observed difference** in the proportion of people who make a purchase after hearing an advertisement (a cause-and-effect conclusion). The ideal response also gives the explanation in the **context** ((2 points)).
> If a correct explanation is given, **but the explanation is not in context**, the response is considered to be partially correct ((1 point)).
> Responses that indicate that random assignment tends to produce comparable groups or that good experiments include random assignment to treatments **but do not specifically tie this to the type of conclusion that can be drawn** are also considered to be only partially correct ((1 point)).
> Responses that argue only that this is a **"fair way" to assign customers** to the two experimental groups or that only talk about lack of bias are considered to be incorrect ((0 points)).

Part b did not have as large of an increase as part a. Particularly, as mentioned above, the average for the pre-assessment was 50% and the average for the post-assessment was 59%. The median was one for both the pre- and post-assessment (out of two points).

Table 5.7 shows sample student responses to part b and the points they were assigned. I chose these examples because they illustrate the different types of responses: (a) flipping a coin helps randomly assign people, (b) comparing across groups, and (c) creating generalizations from experimental designs. For example, in the pre-assessment, Elenai stated that flipping a coin will "give you better chance of splitting the sample group up evenly without controlling who gets read which Advertisement" (flipping a coin helps randomly assign people), but does not state how she will be able to compare across the different groups or make inferences about the effect of the advertisement. As a result, this response was assigned zero points. Elenai extends her thinking in the post-assessment, where she states that the two groups will help "compare the groups" (control or treatment). However, using the LOCUS rubric, she does not reference how the advertisements may have caused observed differences and, consequently, the response was assigned one point.

Table 5.7: Sample student responses to part b of Q15 (study design and random assignment) in the study design pre- and post-assessments

Student	Pre-Assessment	Points	Post-Assessment	Points
Elenai	Flipping a coin you have a 50/50 chance of getting heads or tails. Using this method will give you better chance of splitting the sample group up evenly without controlling who gets read which Advertisement.	0	By flipping a coin you have a 50/50 chance which randomizes the who will hear A or B. It gives an equal chance to hearing either of the advertisements to compare the groups.	1
Jaime	To create two random groups and figure out which advertisement is more appealing to people	2	He would use this method to make the sample be random and to see if there's any effects after two weeks but there is that possibility that the sample may not be evenly split. 50 people should do one and 50 should do the second advertisement.	2

In terms of Jaime's responses, he appeared to acknowledge how the randomization process may help understand a cause-and-effect relationship. Particularly, in the pre-assessment, he stated that the "two random groups" may help understand "which advertisement is more appealing to people." Admittedly, this response was challenging to grade because Jaime did not present a randomization process. Nonetheless, I followed the LOCUS rubric and Jaime referenced a cause-and-effect conclusion and referred to the context. Jaime had a similar response in the post-assessment, but specifically referenced seeing "any effects after two weeks." Although having groups of the same size is not necessary in experimental studies, Jaime also noted that flipping a coin may result in uneven group sizes and instead recommends having groups of the same size (50 in the control and 50 in the treatment). It is also worth noting that these decisions (what is necessary in the response, how many points each part of the response is, how partial points are assigned, if I should deduct points) is itself a social process, providing further illustrations of how data is a result of a social process and therefore susceptible to our own beliefs and perspectives.

Question 16 (Study Design and Sampling Variation)

The second free response question is shown in Table 5.8. According to the LOCUS description of the question (Jacobbe et al., 2014), this question was about recognizing a need for random selection and describing an appropriate method to select a random sample (part a) and explaining why sample variation should be taken into account (part b).

Table 5.8: Question 16 in the study design pre- and post-assessments

Pre-Assessment	Post-Assessment
A farmer conducted an experiment to find out whether a new type of fertilizer would increase the size of tomatoes grown on his farm. The farmer randomly assigned 10 tomato plants to receive the new fertilizer and 10 tomato plants to receive the old fertilizer. All other growing conditions were the same for the 20 plants. At the end of the experiment, the mean weight of tomatoes grown with the new fertilizer was 0.4 ounce heavier than the mean weight of the tomatoes grown with the old fertilizer. (a) Describe one method that the farmer could have used to randomly assign the 20 plants into groups of 10 each. (b) Based on the results, the farmer is convinced that the new fertilizer produces heavier tomatoes on average. Briefly explain to the farmer why simply comparing the two means is not enough to provide convincing evidence that the new fertilizer produces heavier tomatoes.	A farmer conducted an experiment to find out whether a new type of fertilizer would increase the size of pumpkins grown on his farm. The farmer randomly assigned 10 pumpkin plants to receive the new fertilizer and 10 pumpkin plants to receive the old fertilizer. All other growing conditions were the same for the 20 plants. At the end of the experiment, the mean weight of pumpkins grown with the new fertilizer was 0.4 ounce heavier than the mean weight of the pumpkin grown with the old fertilizer. (a) Describe one method that the farmer could have used to randomly assign the 20 plants into groups of 10 each. (b) Based on the results, the farmer is convinced that the new fertilizer produces heavier pumpkins on average. Briefly explain to the farmer why simply comparing the two means is not enough to provide convincing evidence that the new fertilizer produces heavier pumpkins.

In terms of grading part a, I also used the LOCUS rubric from Jacobbe et al. (2014). Particularly, they stated that (comments added in double parentheses and bold is added for emphasis):

> **Part a:** Describe a process for randomly assigning experimental units to treatments in an experiment.
> An ideal response to part (a) describes a way of assigning the 20 plants to the two fertilizers using **some form of random assignment**. To be considered essentially correct, the response needs to **identify how the random assignment would be carried out** and the method described would need to result in **two groups with 10 plants** in each group. Responses that are equivalent to pulling numbers from a box or hat need to specifically mention mixing in order to be considered essentially correct. ((2 points))
> Because the **question specified groups of equal size,** responses that describe methods that use random assignment **but that might result in groups of different sizes** (for example, flipping a coin for each plant to determine which fertilizer the plant would receive) are considered to be partially correct for part (a). ((1 point))
> Responses that **do not indicate a method of random assignment** (for example just saying "randomly pick 10 plants for the first fertilizer") but do describe a method that ensures that there are **10 plants in each fertilizer group** are also considered partially correct for part (a). ((1 point))

This part of the question had an increase from the pre-assessment (average of 59%, median of 1) to the post-assessment (average of 79%, median of 2).

All students provided responses that included two groups of the same size to part a. Table 5.9 shows two sample student responses to part a and the points they were assigned. These examples were selected to illustrate the diversity in the responses and some challenges with grading these responses. Elenai's responses were chosen because she did not consider an appropriate randomization method in pre-assessment but did in the post-assessment. Particularly, in the pre-assessment, she stated that she would create two groups depending on the size of the plant, where the "10 smallest plants" would be assigned to one group and 10 would be assigned to another group. In addition to this not being a random process, there may be some confounding effects if one group is mainly small plants and the other is mainly large plants or plants of all sizes. As a result, this response was assigned zero points. In the post-assessment, Elenai included a form of random assignment ("every other plant"). Assuming that the plants were randomly ordered (e.g., not from smallest to largest), this random assignment may be methodologically appropriate. As a result, this response was assigned two points.

Table 5.9: Sample student responses to part a of Q16 (study design and random assignment) in the study design pre- and post-assessments

Student	Pre-Assessment	Points	Post-Assessment	Points
Elenai	The farmer could have picked the 10 smallest plants along with assigning 10 more to see if the fertilizer could increase their size.	0	In order to keep growing conditions the same and having 10 plants receive new fertilizer and 10 receive old fertilizer he could do every other plant as long as there was enough space in between in order to not contaminate.	2
Jacky	Could have drawn from a jar with the numbers 1-20 in which corresponds with the plants, returning the number withdrawn back into the jar ensuring a true 1/20 probability every time we drew from the jar.	2	Using a simple sampling method would be the best way to assign the 20 plants into their groups. This could be done by labeling the plants from 1-20, and pulling 10 numbers from a jar with the number 1-20 written on them to determine which would receive the new fertilizer	2

Both of Jacky's responses included appropriate random sampling methods (randomly selecting ten numbers from a jar) that would result in two groups of the same size (as stated in the goal of the task). However, in the pre-assessment, Jacky also mentioned that she would return the numbers back into the jar to ensure "a true 1/20 probability every time we drew from the jar." Although adding the numbers back in the jar would ensure a 1/20 probability each time, it is not necessary for sampling. That is, she could pull 10 numbers from a jar, assign those to one group, and assign the remaining ten to another group (as suggested in her response to the post-assessment). In fact, adding the numbers back to the jar might take longer (e.g., she could pull the number 12 five times in a row and would have to continue until she gets 10 unique numbers).

Similar to Jaime's response, I did not penalize Jacky for providing additional information that was not necessarily needed because she still addressed the goals listed by the LOCUS rubric.

In terms of grading part b (Briefly explain to the farmer why simply comparing the two means is not enough to provide convincing evidence that the new fertilizer produces heavier vegetables), I also used the LOCUS rubric from Jacobbe et al. (2014). Particularly, they stated that (comments added in double parentheses and bold added for emphasis):

> **Part (b)**: Explain the need to take sampling variability into account when drawing conclusions based on data
> Part (b) asks students to explain why it is not appropriate to reach a decision based solely on the fact that one fertilizer group mean is greater than the other fertilizer group mean. An ideal response to part (b) recognizes that even if all plants received the same fertilizer, there would still be variability in tomato weights from one plant to another, and there is a need to determine if a difference of 0.4 ounce might be something that could be observed just by chance when there is no difference in the effect of the two fertilizers. To be considered essentially correct for part (b), the response must: (1) **refer to sampling variability or the variability introduced by random assignment of plants to fertilizers**, and (2) indicate that the **observed difference in averages might be due to chance alone** (the random assignment of plants to fertilizer groups). ((2 points))
> Responses that **only include one of these two required elements are considered to be partially correct for part** (b). ((1 point))
> **A response that does not include either of these two required elements** (for example, one that just says "you need to do a test"), is considered to be incorrect for part (b). ((0 points))

As mentioned above, the average for the pre-assessment was 54% and the average for the post-assessment was 63%. The median was one for both the pre- and post-assessment (out of two points).

Table 5.10 shows sample student responses to part 1 and the points they were assigned. These examples were chosen to highlight the diversity in the student responses: (a) referencing a small sample size, (b) confounding variables, and (c) repeated experiments to confirm that any differences in vegetable size are not due to chance. First, in the pre-assessment, Ellie recognized that the sample size was too small to create any meaningful conclusions. From my experiences, a small sample size is often an issue of concern. However, using the LOCUS rubric, Ellie did not refer to sampling variation, randomization, or how the observed differences might not be due to chance. As a result, her response was assigned zero points. Ellie also referenced having a "greater amount of plants" in the post-assessment but added that there may be "more variables than just the fertilizer" that may affect the growth of a vegetable. I interpreted this as Ellie acknowledging confounding variables as a source of variation. It is possible that Ellie was also considering how the difference was not due to chance. However, Ellie did not explicitly mention that the variation might be due to chance, so Ellie's response was assigned one point.

Table 5.10: Sample student responses to part b of Q16 (study design and random assignment) in the study design pre- and post-assessments

Student	Pre-Assessment	Points	Post-Assessment	Points
Ellie	The sample size is very small and it may not produce the same results. Increase the sample size to get better results.	0	there are more variables than just the fertilizer that counts in this experiment. It should have been taken on a greater amount of plants to really see if it as the fertilizer that made that difference	1
Robert	There could be other factors involved that cause the potatoes to be larger.	1	There is always going to be small differences in the data. He should repeat this experiment several times and if he discovers that the new fertilizer is consistently growing heavier proportions, then he could conclude that the new fertilizer produces heavier pumpkins.	2

Robert's response in the pre-assessment also mentioned that there "could be other factors involved," suggesting that there may be some confounding variables that created variation. This response was assigned one point. Robert extended his thinking in the post-assessment, where he said that "there is always going to be small differences in the data" (sampling variation). To confirm the differences, he suggested repeating the study "several times" to provide evidence for or against any cause-and-effect relationship between the new fertilizer and vegetables. Although he did not explicitly reference randomization in the repeated studies, he did reference repeating “this experiment” which included a random assignment process. Thus, Robert’s response was assigned two points.

Unit on Regression

The regression assessment was mainly about correlation and simple linear regression. However, we also discussed multiple linear regression, logistic regression, classification and regression trees, and random forests as well as various measures for model evaluation and comparisons throughout the course. Regression was chosen as a unit of focus for this dissertation because it provides the foundations of predictive modeling that are common in other machine learning algorithms. We also made explicit connections to the different phases of the PPDAC cycle throughout the unit. For example, we discussed how sampling (planning and data phase) may influence model prediction accuracy in the facial recognition algorithm (analysis). In total, 14 students took the pre- and post-assessment.

Total Pre-Post Gains

Similar to the analysis of the study design unit, I begin by presenting the gains from the pre- to post-assessment. Figure 4.4 shows boxplots of the pre- and post-assessments for the regression unit. Individual student scores are shown by the points, the mean is shown by the dashed line, and the median is shown by the solid line.

All students had an increase from the pre- to the post-assessment. Furthermore, only two students (about 14% of the students that took the pre- and post-assessment) had a 12 (about 67%) or higher in the pre-assessment, but 12 students (about 86% of the students that took the pre- and post-assessment) had a 12 or higher on the post-assessment.

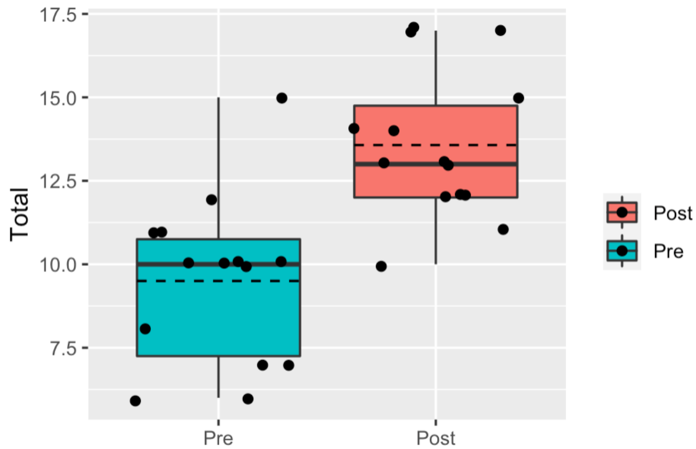


Figure 5.3: Boxplot of pre- and post-assessment scores in the regression unit

Overall, there was an increase in the mean and median from the pre- to the post-assessment. Table 5.12 shows the pre and post means, standard deviations, paired t-test, Cohen's effect size, and normalized gains the mean increased from about 9.54 (about 53%) to about 13.50 (about 75%), a 3.96-point difference (about 22%). In terms of test statistics, all the test statistics suggest that there was a significant or practical increase from the pre- to the post-assessment. Specifically, the paired t-test p-value may be interpreted as a statistically significant p-value, suggesting that the difference in means from the pre- to the post-assessment did not occur by chance. Furthermore, the effect size was large and, in fact, all the values in the 95% confidence interview were large. Further, the gain of average was large and the average gain was medium, suggesting that there is also a practical significance in the pre-post gains. In terms of the median, the paired Wilcoxon sign rank test p-value may also be interpreted as statistically significant increase from 10.5 in the pre-assessment (about 55%) to 13 in the post-assessment (about 72%), shown in Table 5.12. That is, about half of the students had a 55% or higher in the pre-assessment, but about half of the students had a 72% or higher in the post-assessment.

Table 5.11: Paired t-test, Cohen's effect size, and normalized gains in the study design unit for the total and by question type

Note:
[1] $p<0.5$*, $p<0.01$**, $p<0.001$***, $p<0.0001$****
[2] $|d|<0.5$ small, $|d|<0.8$ medium, otherwise large
[3] $g<0.3$ low, $g<=0.7$ medium, otherwise large (Hake, 1999)

		Pre		Post		Diff in Means	Paired T-Test			Cohen's Effect Size		Normalized Gains	
	Points	Mean	SD	Mean	SD		t statistic	p-value[1]	95% CI	d[2]	95% CI	Gain of Avg[3]	Avg Gains[3]
Total	18	0.53	(0.14)	0.75	(0.12)	0.23	6.62	0.001**	(2.74, 5.4)	1.71 (L)	(0.88, 2.54)	0.92 (L)	0.48 (M)
Multiple Choice	6	0.68	(0.27)	0.73	(0.23)	0.05	0.84	0.414	(-0.08, 0.17)	0.19 (S)	(-0.28, 0.66)	0.17 (S)	0.08 (S)
Free Response	12	0.45	(0.15)	0.77	(0.11)	0.32	8.71	<0.001***	(0.24, 0.39)	2.30 (L)	(1.26, 3.33)	1.40 (L)	0.57 (M)

Table 5.12: Paired Wilcoxon the total score in the study design unit for the total and by question type

					Paired Wilcoxon Sign Rank		
	Points	Pre-Median	Post-Median	Diff in Medians	V statistic	p-value	95% CI
Total	18	10	13	3	105	0.001**	(2.5, 5)
Multiple Choice	6	4	4	0	24	0.429	(-1, 2)
Free Response	12	5	9	4	105	0.001**	(3, 5)

Note:
For the paired Wilcoxon: p<0.5*, p<0.01**, p<0.001***, p<0.0001****

Gains by Question Type (Free Response or Multiple Choice)

I conducted a similar analysis to the total pre-post gains with the question type (free response or multiple choice). The goal of this analysis was to see if the gains varied by question type. Figure 5.4 shows a boxplot for the scores on the free response (left) and multiple choice (right) questions in the pre- and post-assessments. Individual student scores are shown by the points, the mean is shown by the dashed line, and the median is shown by the solid line.

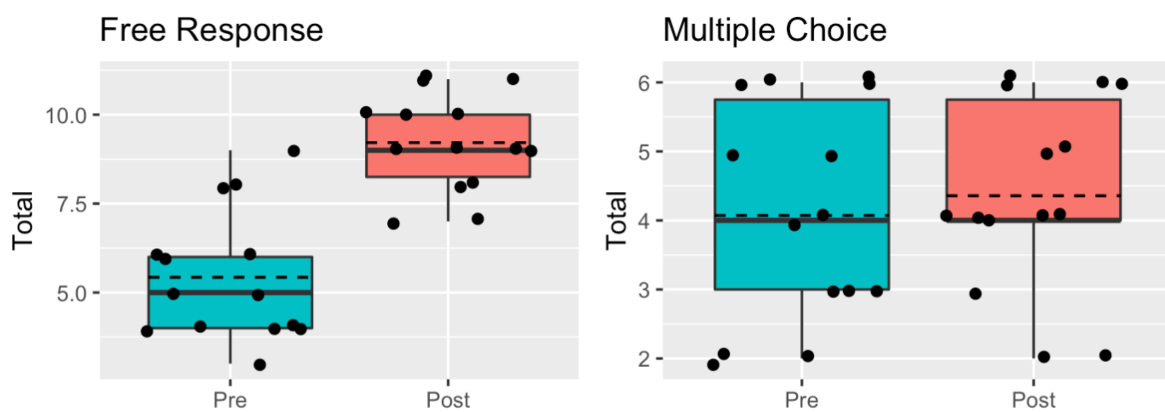


Figure 5.4: Boxplot for the free response (left) and multiple choice (right) questions of the regression unit

Table 5.11 also shows the pre and post means, standard deviations, paired t-test, Cohen's effect size, and normalized gains. Table 5.12 shows the pre-post medians and paired Wilcoxon Sign Rank test. The results were similar to the total pre-post gains, with the paired t-test, paired

Wilcoxon sign rank test, Cohen's D, and gain of average suggesting that there was some significant and practical growth from the pre- to the post-assessments.

The free response questions were out of 12 points. Students had about 45% free response questions correct in the pre-assessment but 77% in the post-assessment, about a 32% increase. The median also increased from about five (about 42%) to nine points (about 75%), a four-point increase (about a 33% increase). The paired t-test and paired Wilcoxon sign rank test had p-values that may be interpreted as statistically significant (i.e., below 0.05), there was a large effect size and large gain of averages, and a medium average gain. Combined, the statistical measures suggest that there was a statistically significant and practical increase from the pre- to the post-assessment.

The multiple-choice questions about regression did not have as large of an increase as the free response questions. Particularly, students had an average of about 68% of the multiple-choice questions correct in the pre-assessment and an average of about 73% correct in the post-assessment, about a 5% increase. The median was four (about 67%) for both the pre- and post-assessments. The paired t-test p-value and paired Wilcoxon Sign Rank test p-value were larger than 0.05, implying that there was no statistically significant change in the mean or median, respectively. Furthermore, the effect size, gain of average, and average of gains were small, implying that there was no practical change in the means. The statistical measures suggest that there was little to no increase in the multiple-choice questions of the regression unit.

Gains by Question

Finally, I looked at the overall gains by question. The goal of this analysis was to identify any specific topics that may have had varied growth (e.g., more growth on questions about sampling than about study design). All multiple-choice questions were one point. All free

Table 5.13: Paired t-test, Cohen's effect size, and normalized gains in the study design unit by question

				Pre		Post		Means Diff	Paired T-Test			Cohen's Eff Size	
Q	Q Type	Topic	Pts	Mean	SD	Mean	SD		t-stat	p-value[1]	95% CI	d[2]	95% CI
1	Multiple Choice	Nonlinear relationship	1	0.57	(0.51)	0.50	(0.52)	-0.07	-0.56	0.583	(-0.35, 0.2)	-0.14 (S)	(-0.65, 0.37)
2	Multiple Choice	Residuals	1	0.57	(0.51)	0.64	(0.50)	0.07	1.00	0.336	(-0.08, 0.23)	0.14 (S)	(-0.15, 0.43)
3	Multiple Choice	Nonlinear relationship	1	0.71	(0.47)	0.71	(0.47)	0.00	0.00	1.000	(-0.32, 0.32)	0.00 (S)	(-0.65, 0.65)
4	Multiple Choice	Correlation	1	0.86	(0.36)	0.93	(0.27)	0.07	1.00	0.336	(-0.08, 0.23)	0.21 (S)	(-0.23, 0.66)
5	Multiple Choice	Correlation	1	0.64	(0.50)	0.71	(0.47)	0.07	0.43	0.671	(-0.28, 0.43)	0.15 (S)	(-0.56, 0.85)
6	Multiple Choice	Compare relationships	1	0.71	(0.47)	0.86	(0.36)	0.14	0.81	0.435	(-0.24, 0.53)	0.34 (S)	(-0.55, 1.24)
7a	Free Response	Describe relationship	2	0.25	(0.43)	0.71	(0.26)	0.46	4.76	<0.001***	(0.26, 0.68)	1.24 (L)	(0.53, 1.95)
7b	Free Response	Prediction, Residuals	2	0.43	(0.43)	0.93	(0.18)	0.50	4.27	<0.001***	(0.25, 0.76)	1.46 (L)	(0.45, 2.48)
7c	Free Response	Predictions - extrapolation	2	0.21	(0.43)	0.46	(0.37)	0.25	1.71	0.110	(-0.07, 0.57)	0.63 (M)	(-0.2, 1.46)
8a	Free Response	Describe relationship	2	0.39	(0.21)	0.75	(0.26)	0.36	4.37	<0.001***	(0.18, 0.54)	1.50 (L)	(0.47, 2.53)
8b	Free Response	Compare relationships	2	0.61	(0.29)	0.82	(0.25)	0.21	2.48	0.028*	(0.03, 0.4)	0.79 (M)	(0.04, 1.54)
8c	Free Response	Compare relationships	2	0.82	(0.32)	0.93	(0.18)	0.11	1.15	0.272	(-0.20, 0.31)	0.41 (S)	(-0.36, 1.18)

Table 5.14: Paired t-test, Cohen's effect size, and normalized gains in the study design unit by question

							Paired Wilcoxon Sign Rank	
	Question Type	Topic	Points	Pre-Median	Post-Median	Diff in Medians	V statistic	p-value[1]
1	Multiple Choice	Nonlinear relationship	1	1	0.5	-0.5	4	0.7728
2	Multiple Choice	Residuals	1	1	1		0	1.0000
3	Multiple Choice	Nonlinear relationship	1	1	1	0	5	1.0000
4	Multiple Choice	Correlation	1	1	1	0	0	1.0000
5	Multiple Choice	Correlation	1	1	1	0	6	0.7656
6	Multiple Choice	Compare relationships	1	1	1	0	7	0.4840
7a	Multiple Choice	Describe relationship	2	0	1	1	0	0.0042**
7b	Multiple Choice	Prediction Residuals	2	1	2	1	0	0.0074**
7c	Multiple Choice	Predictions (extrapolation)	2	0	1	1	12	0.1113
8a	Multiple Choice	Describe relationship	2	1	1.5	0.5	0	0.0048**
8b	Multiple Choice	Compare relationships	2	1	2	1	4.5	0.0411*
8c	Multiple Choice	Compare relationships	2	2	2	0	2	0.3447

response questions had three parts, each of which were two points. Table 5.13 also shows the pre and post means, standard deviations, paired t-test, and Cohen's effect size by question and Table 5.14 shows the pre-post medians and paired Wilcoxon Sign Rank test by question.

There was one question that had no growth: Q3. This question was about determining whether we could use a regression line to predict data that did not show a linear relationship. Although there was no change from the pre- to post-assessment, the same students that had the correct answer in the pre-assessment did not all get the correct answer in the post-assessment. Furthermore, there was one question that had a slight decrease: Q1. This question was also about nonlinear relationships. It is possible that students believed that we could use a regression to predict nonlinear relationships because we discussed nonlinear regression and regression trees in the course. However, further research (i.e., follow-up interviews) would be required to confirm this hypothesis.

Regression Free Response Samples

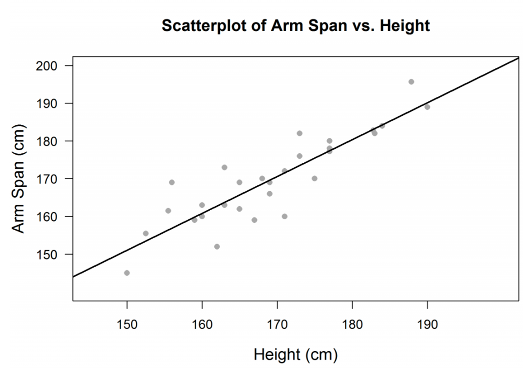
Similar to the study design unit, were two free response questions as part of this pre- and post-assessment. Both questions were from the LOCUS assessment (Jacobbe et al., 2014). As mentioned in the Methods chapter, each part of the question was two points (six points total per question). Beyond these items, there were no changes to the free-response questions.

Question 7 (Describe Relationship, Prediction, and Residuals)

The first free response question in the regression unit is shown in Figure 5.5. According to the LOCUS description of the question (Jacobbe et al., 2014), this question was about recognizing that the slope of a least-squares regression line represents an average change in the y variable associated with an average one unit increase in the x variable (part a), finding a predicted value and its residual given the equation of a line (part b), interpreting the residual (part b), and recognizing the limitations of extrapolation (predicting outside of the range of data).

The heights (in centimeters) and arm spans (in centimeters) of 31 students were measured. The association between x (height) and y (arm span) is shown in the scatterplot below. The equation of the least-squares regression line for this association is also given.

estimated armspan = 4.5 + 0.977height



(a) If Mike is 5 cm taller than George, what is the expected difference in their arm spans? Show your work.
(b) Jane is 158 cm tall and has an arm span of 154 cm. Rhonda is 163 cm tall and has an arm span of 165 cm. Does the least-squares regression line give a more accurate predicted value for Jane or Rhonda? Explain.
(c) Doug is 210 cm tall. Would you use this least-squares regression line to predict his arm span? Explain.

Figure 5.5: Question 7 in the pre- and post-assessments for the regression unit

In terms of grading part a, I used the LOCUS rubric from Jacobbe et al. (2014). Particularly, they stated that (comments added in double parentheses and bold is added for emphasis):

> **Part a:** An ideal response to part (a) recognizes that the slope of the least-squares regression line can be interpreted as the **expected change in arm span associated with a 1 cm increase in height**. Students who understand this interpretation could then just multiply the given slope by 5 to obtain the expected difference in arm span for two people whose height differed by 5 cm. ((4.885 cm, or rounded)) Responses that take this approach and that **provide an explanation or include supporting work** are scored as essentially correct for part (a). ((2 points))
> While it was anticipated that students would take the approach described above in answering this question, the majority of students chose instead to assume heights for Mike and George with the height from Mike being 5 cm greater than the height for George. These heights were then used in the equation of the least-squares regression line to obtain predicted arm spans, and the difference in predicted arm spans was then calculated. While this approach is a lot more work, it leads to a correct answer, and responses using this method to obtain correct predicted values and the correct difference in predicted values are also scored as essentially correct for part (a).

> Responses that used one of the two methods described but **which included errors in the calculations** needed to determine the expected difference were scored as partially correct for part (a). ((1 point))

Part a had an increase from the pre-assessment (average of 25%, median of 0) to the post-assessment (average of 71%, median of 1).

There were four types of responses across the pre- and post-assessments for part a (If Mike is 5 cm taller than George, what is the expected difference in their arm spans? Show your work): the difference between Mike and George's arm span is (a) 5cm, (b) about 9.385 cm, (c) about 4.885 cm (the correct answer), and (d) another quantity. Most of the students inputted their work (e.g., showed their steps) as the response. Since this question was more about calculating a specific value, I present and explain the different methods that students used and reference student responses to illustrate some of their reasoning. Furthermore, it is worth noting that this question was mathematical in the sense that it entailed understanding and calculating rates of change, highlighting the intersectionality between mathematics and statistics.

First, the most common incorrect answer was that the difference between Mike and George's arm span was 5cm (about half of the students gave this response). For example, Ellie stated that "The expected difference of their arm spans are 5 cm. I got this answer because there is a 5 cm height difference between the two people." Similarly, Regina stated that "5 cm difference because it's the relationship between arm length and height." One possible reason for why students stated that the difference between Mike and George's arm span was 5cm is because they believed that the difference in their height was the same as the difference in their arm spans. For example, if the difference between Mike and George's height is 6cm, then, under this method, the difference in their arm span is also 6cm. Similarly, if the difference in their height was 22 cm, then the difference in their arm span is also 22cm. However, the height and arm span are not directly related like this. Using any of the two approaches mentioned by LOCUS, students should have found the difference in their arm span is 4.885cm.

The second most common incorrect answer was that the difference between Mike and George's arm span was about 9.385 cm. It appeared that the students who provided this answer because "if you plug in 5 for height, it would show the difference between Mike and George" (Sam). This would result in:

$$\begin{aligned} \text{estimated armspan} &= 4.5 + 0.977*\text{height} \\ &= 4.5 + 0.977*(5)\text{ cm} \\ &= 4.5 + 4.885\text{ cm} \\ &= 9.385\text{ cm.} \end{aligned}$$

However, students who used this approach may have interpreted the slope as the difference in heights between two people rather than the average increase in the arm span given an increase in the height. Specifically, this approach is finding the estimated arms pan of someone who is 5cm tall. That is, using the provided equation, someone who is 5cm tall has an estimated arm span of 9.38cm. Thinking of the practical implications and the context, this may have raised red flags because (a) a person is likely not 5cm tall and (b) the arm span of a person is likely not almost twice their height.

A third common answer (and the most common in the post-assessment) was correctly identifying the difference in the arms spans as 4.885cm. Students used three methods to find 4.885cm. Samples are shown in Table 5.15. First and similar to the description from LOCUS, most of the students appeared to have chosen two heights that are within the range of the data and 5cm apart, then found the height for Mike and George, then found the difference. For example, Eric chose to "input George as 150cm and Mike as 155cm," got 151.05 and 155.93 for their heights, and found that the difference was about 4.8cm. Although not as common, the second method that students used to find the difference between the arm spans of George and Mike was a generalized version of the previous method, where one height was x and the other height was $x+5$. For example, Elena estimated the height of George as 4.5 + .977(x) and the height of Mike as 4.5 + .977(x+5). Subtraction both equations result in a height of .977(5) or 4.885cm. The third method appeared by less than five students across the pre- and post-assessment where students found the difference by multiplying the slope by five cm. It is unclear why students multiplied the slope by 5cm, but one possible explanation is that they understood the regression slope as the average increase in the arm span given a one unit increase in the height. Thus, if there is a five unit increase in height, then the arm span increases by the product of five and the slope, or 5 cm * .977 = 4.885cm.

Table 5.15: Sample student responses to part a of Q7a in the regression pre- and post-assessment

Student	Pre or Post	Response	Description of the method
Eric	Pre	There is a difference of 4.8 cm in their arm length I input George as 150cm and Mike as 155cm and got 151.05 and 155.93 and after I subtracted them and that's how I got 4.8 difference in arm length	Select to heights that are within the range of the data and 5cm apart
Elenai	Post	If you make x their height (being equal) but add the 5 cm to Mike you end up with two equations that are almost identical. George = 4.5 + .977(x) while Mike = 4.5 + .977(x+5) the expected difference comes with the .977(5) = 4.885	Generalized version of the previous method, where one height is x and the other height is $x+5$.
Josue	Post	5 times the slope= 4.885	Definition of regression slope as the average increase in the y variable given a one unit increase in the x variable
Madelyn	Pre	4.885 cm (5 cm*.977 =the difference in armspan)	
Madelyn	Post	4.985 - take the difference in their highs and multiply by .977	

Finally, there were responses that were not 5cm, 9.385cm, or 4.885cm. In particular, 3.6 cm, 3.85 cm, and 3.908cm all appeared once. The responses did not show their full work and were therefore assigned zero points. However, it is possible that these responses resulted from miscalculations or typos and should have been assigned one point. Hence, the importance of showing work.

In terms of grading part b (Does the least-squares regression line give a more accurate predicted value for Jane or Rhonda? Explain), I also used the LOCUS rubric from Jacobbe et al. (2014). Particularly, they stated that (comments added in double parentheses and bold added for emphasis):

> **Part (b):** asks students to determine which two predictions based on the least-squares regression line is more accurate. An ideal response to part (b) **indicates that the prediction of Rhonda's arm span is more accurate that the prediction of Jane's arm span** and **provides justification for this choice.** There are two ways that a student could provide a correct justification. One possible justification is based on **calculating predicted values and residuals and then noting that the absolute value of the residual for Rhonda is less than the residual for Jane**, indicating that the predicted arm span is closer to the actual arm span for Rhonda. Responses that provide a justification based on this method are considered to be essentially correct for part (b). ((2 points))
> Responses based on this method that include **errors in calculating** the predicted values or the residuals are considered to be partially correct for part (b). ((1 point))
> A second approach that could be used to support the choice of Rhonda in part (b) uses the given scatterplot and least squares line. Students using this method use the information on height and arm span for Rhonda and Jane to **plot points on the scatterplot**. They then note that the point that corresponded to Rhonda's height and arm span is **closer to the least-squares line** than the point that corresponded to Jane's height and arm span. Because predicted arm spans are points on the least-squares line, this means that the predicted arm span would be closer to the actual arm span for Rhonda. Responses based on this method are scored as essentially correct for part (b) provided that they **include an explanation and show the two relevant points drawn on the scatterplot**. ((2 points))
> Responses based on this method that include **errors in plotting the points** on the scatterplot are considered to be partially correct for part (b). ((1 point))

As mentioned above, the average for the pre-assessment was 43% and the average for the post-assessment was 93%. The median was one for the pre-assessment and two for the post-assessment (out of two points).

Table 5.16 provides two samples that were representative of almost all of the student responses (over 90% of the responses in the pre- and post-assessment).

Table 5.16: Sample student responses to part a of Q7b regression pre- and post-assessments

Student	Pre-Assessment	Points	Post-Assessment	Points
Jacky	The regression line provides a value that is quite close to her height so Rhonda in this situation would have the more accurate predicted value.	1	The least-squares regression line gives a more accurate predicted value for Jane than Rhonda. In Jane's case, her residual is .87 while Rhonda's residual is 1.25.	1
Derrick	I would say Rhonda since she has closer measurements	1	Rhonda is more of an accurate prediction than Jane. When you run the numbers through the equation you get the jane's estimated arm span is 158 which is a about a 4 cm difference. However when you put Rhonda's numbers you get that the predicted arm span is 163 which is about a2 cm difference.	2

In particular, Jacky stated how the "regression line provides a value that is quite close to her ((Rhonda's)) height" and Derrick stated that "I would say Rhonda since she has closer measurements." It is possible that both Jacky and Derrick were referencing the residuals (with statements like "quite close" and "closer measurements") that describe the distance between an observed and predicted value. As a result, both Jacky's and Derrick's responses were assigned one point. It is worth noting that it is unclear if students had learned about residuals in previous statistics classes and, in fact, no students mentioned "residuals" in the pre-assessment. However, it was more common for students to mention "residuals" and show their calculations in the post-assessment. For example, Jacky explicitly referred to calculating a "residual" and Derrick described the approximate difference between the observed and predicted value. Notably, it appears that Jacky may have had some calculation errors which, using the LOCUS rubric, resulted in her response being assigned one point instead of two. Nonetheless, it was clear that most of the students were using or referencing residuals in the post-assessment, especially given that 93% of the responses to this question in the post-assessment were correct.

In terms of part c (Would you use this least-squares regression line to predict his arm span?), the LOCUS rubric was:

> **Part (c):** Part (c) asks students if it is reasonable to use the least-squares regression line to predict the arm span for Doug, an individual with a height of 210 cm. An ideal response to part (c) recognizes that 210 cm is quite a bit greater than the height of the tallest person in the group of 31 students that were used to develop the equation of the least-squares line. Because this represents an **extrapolation** beyond the range of the data, an essentially correct response to part (c) includes a statement that **the least-squares regression line should not be used to predict Doug's arm span**. ((2 points))
> Responses that indicate that it is not reasonable to use the equation of the least-squares regression line to predict Doug's arm span but **which do not specifically link this decision to Doug's height being outside the range of the data** used to develop the equation are considered to be only partially correct for part (c). ((1 point))
> Responses that do not include an explanation or that indicate that it is OK to use the least-squares regression line to predict Doug's arm span are considered incorrect for part (c). ((0 points))

As mentioned above, the average for the pre-assessment was 21% and the average for the post-assessment was 46%. The median for the pre-assessment was 0 and the median for the post-assessment was 1 (out of two). Notably, extrapolation was not greatly discussed in the course, partially because I ran out of time the day I was going to talk about extrapolation. However, I did briefly discuss it when I introduced the homework because there were some homework questions about extrapolation.

Table 5.17 provides two samples that were common across the pre- and post-assessments. First, many students in the pre- and post-assessment stated that it was possible to predict Doug's arm span even though it was outside the range of the heights that were given. For example, Josue stated that it was possible to predict Doug's arm span because "it is the line that best predicts the arm span" and "because the points are almost linear." Jaime also said that it was possible to predict Doug's arm span and provided an estimate that was likely calculated using the provided equation. However, Josue and Jaime both recognized that this question was about extrapolation and, as a result, it was not appropriate to predict Doug's arm span. In particular, Josue stated that Doug's "height is not in range" and Josue stated that Doug's height is "higher than every person in the graph."

Table 5.17: Sample student responses to part a of Q7c regression pre- and post-assessments

Student	Pre-Assessment	Points	Post-Assessment	Points
Josue	yes, because it is the line that best predicts the arm span. It works because the points are almost linear	0	No, because the height is not in range.	2
Jaime	Yes, 210cm	0	No because he is higher than every person in the graph. He wouldn't be on the least-squares regression line.	2

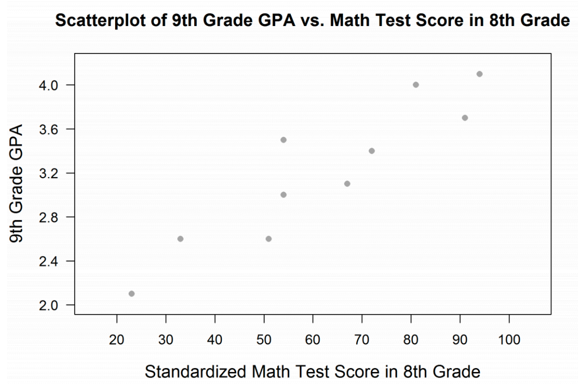
Question 8 (Describe and Compare Relationships)

The second free response question in the regression unit is shown in Figure 5.6. According to the LOCUS description of the question (Jacobbe et al., 2014), this question was about interpreting bivariate numerical data (part a), comparing scatterplots to determine which has a stronger relationship (part b), describing the strength of the relationship (part b), and choosing between two variables to predict a value of interest (part c).

In terms of grading part a, I also used the LOCUS rubric from Jacobbe et al. (2014). they stated that (comments added in double parentheses and bold is added for emphasis):

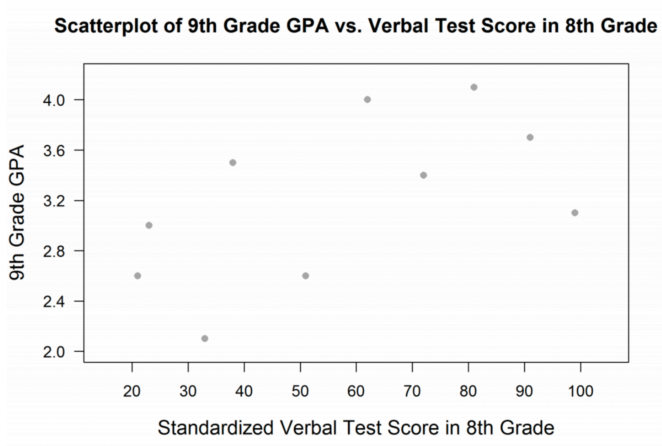
> **Part (a):** An ideal response to part (a) notes that the relationship between standardized math test score in 8th grade and 9th grade GPA is **approximately linear and that the relationship is strong and positive** (higher values of 9th grade GPA tend to be paired with higher 8th grade math test scores). **The ideal response also includes context—it is not enough to just say "strong, positive, and linear."** This means that the ideal response indicates that the relationship is strong, linear, and positive and is in context. ((2 points))
> If a response is missing one or two of these four elements, it is considered partially correct for part (a). ((1 point))
> If it is missing more than two, it is considered incorrect for part (a). ((0 points))

A random sample of 10 high school students was selected to investigate the relationship between standardized test scores in 8th grade and GPA (grade point average) in 9th grade. The scatterplot below shows the relationship between standardized math test scores in 8th grade and GPA (grade point average) in 9th grade.



(a) Based on scatterplot, describe the relationship between standardized math test scores in 8th grade and GPA (grade point average) in 9th grade.

For the data on standardized math test score in 8th grade and GPA in 9th grade, the value of the correlation coefficient is r = 0.92. The scatterplot below shows the relationship between standardized verbal test scores in 8th grade and GPA (grade point average) in 9th grade.



(b) For the data on standardized verbal test scores in 8th grade and GPA in 9th grade, will the value of the correlation coefficient be greater than, less than, or about the same as r = 0.92? Explain.

(c) If you want to predict 9th grade GPA, which variable would you use as a predictor— 8th grade standardized math test score or 8th grade standardized verbal test score? Explain.

Figure 5.6: Question 8 in the pre- and post-assessments for the regression unit

This part had an increase from the pre-assessment (average of 39%, median of one) to the post-assessment (average of 75%, median of 1.5), where more students referenced at least three of the four parts of the ideal response (strong, positive, linear, and interpreted in the context).

Table 5.18 provides two sample student responses to part a. It was common for students to refer to the context in the pre-assessment, but not all three of the other features (strong, positive, and linear) that would make their responses an ideal response according to the LOCUS rubric. For example, in the pre-assessment, Madelyn only interpreted the relationship in the context, relating a students' score on the standardized verbal test to their 9th grade GPA. Since the response only included one of the four features, the response was assigned zero points. Ellie referenced two of the features: the context and the linear relationship. Thus, Ellie's response was assigned one point (partial credit).

Table 5.18: Sample student responses to part a of Q8a regression pre- and post-assessments

Student	Pre-Assessment	Points	Post-Assessment	Points
Madelyn	the higher people scored the higher their gpa.	0	positive, fairly strong. As standardized math scores go up GPA goes up too.	1
Ellie	There is a linear correlation between the two scores. The higher the score in 8th grade the higher the gpa in 9th grade.	1	the relationship between 8th grade test scores and 9th grade gpa is a positive, strong, linear relationship. Higher test scores are paired with higher gpas therefore giving a strong, positive, linear relationship.	2

Students added more details about the relationship in the post-assessment. For example, Madelyn noted that the relationship was "positive, fairly strong," and interpreted the relationship in the context. However, following the LOCUS rubric, she did not explicitly mention that there was a linear relationship, so her response was assigned one point. Ellie noted all four features (strong, positive, linear, and interpreted in the context). As a result, her response was assigned two points.

In terms of grading part b (will the value of the correlation coefficient be greater than, less than, or about the same as r = 0.92?), I also used the LOCUS rubric from Jacobbe et al. (2014). they stated:

> **Part (b)**: Part (b) asks students to indicate whether the correlation coefficient for the data displayed in the scatterplot of 9th grade GPA versus 8th grade verbal test score would be greater than 0.92, the given value of the correlation coefficient for the data displayed in the scatterplot of 9th grade GPA versus 8th grade math test score. An ideal response to part (b) correctly indicates that the value of the **correlation coefficient would be less than 0.92** and **provides an explanation** based on the fact that the approximate linear relationship in the second scatterplot (GPA versus verbal test score) is **weaker than the approximate linear relationship in the first scatterplot** (GPA versus math test score). ((2 points)) Responses that correctly indicate that the correlation coefficient for GPA versus verbal test score will be less than 0.92 but that provide an **explanation that is considered weak or incomplete** are considered to be partially correct for part (b). ((1 point))

As mentioned above, the average for the pre-assessment was 61% and the average for the post-assessment was 82%. The median for the pre-assessment was one and the median for the post-assessment was two (out of two).

Table 5.19 provides two sample responses that were common across the pre- and post-assessment. First, in the pre-assessments, almost all the students correctly identified that the correlation will be less than 0.92. However, students provided general statements about the correlation, like the points "don't have a strong correlation" (Jaime) or "these points do not seem to be correlated" (Robert). However, students added more details to their response to this question in the post-assessment, typically in terms of the (a) linearity or (b) spread of the points. For example, Jaime specified that the relationship "is not as linear as the 9th grade graph ((with a correlation of 0.92))" and therefore would have a correlation less than 0.92. Robert focused on the spread of the points as a measure of the strength of the correlation, where the correlation would be "less than (0.92)) because dots in graph are more spread out." That is, Robert correctly identified that a larger spread (assuming that the axes units are the same) implies larger correlation.

Table 5.19: Sample student responses to part a of Q8b regression pre- and post-assessments

Student	Pre-Assessment	Points	Post-Assessment	Points
Jaime	less than 0.92, since the x and y values in the second scatter plot don't have a strong correlation like the first one	1	The relationship of the data in the graph is not as linear as the 9th grade graph, so the correlation coefficient for this graph (above) would be less than r=0.92.	2
Robert	Less than, these points do not seem to be correlated.	1	less than because dots in graph are more spread out, less accurate	2

Finally, in terms of grading part c (If you want to predict 9th grade GPA, which variable would you use as a predictor— 8th grade standardized math test score or 8th grade standardized verbal test score? Explain.), the LOCUS rubric from Jacobbe et al. (2014) states:

> **Part (c):** Part (c) asks students to choose between 8th grade math test score and 8th grade verbal test score as a predictor of 9th grade GPA.
> An ideal response to part (c) chooses math test score as the predictor and justifies this choice based on a comparison of the **strength of relationship** between 9th grade GPA and each of the **two potential predictors**. ((2 points))
> Responses that choose math test score as the predictor but provide an **explanation that is weak or incomplete** are considered to be only partially correct for part (c). ((1 point))
> Also considered partially correct for part (c) are responses that give a good explanation of the role that strength of the relationship plays in making a choice between predictors but that **do not actually make a choice**. That is, they fail to actually state that math test score is the chosen predictor. ((1 point))
> Responses that do not include an explanation for the stated choice or that provide an explanation that is not based on the data displayed in the given scatterplots are scored as incorrect for part (c). ((0 points))

The average for the pre-assessment was 82% and the average for the post-assessment was 93%. The median for the pre- and post-assessment was two (out of two). This question had relatively high accuracy when compared to the other free response questions.

Table 5.20 provides three sample responses that are representative of the other student responses. In the pre-assessment, there were students like Caden that listed which relationship they would use but did not provide an explanation. These responses were assigned zero points. There were also cases where students explained a correlation of a relationship, but did not compare the strengths of the correlation as a justification for choosing one over the other. For example, Jaime stated that he would use 8th grade standardized math tests because the "relationship is quite strong ((with 9th grade scores))" but did not relate to correlation between verbal test scores and 9th grade scores. Responses like this were assigned one point. I also included Madelyn's statement because she did not explicitly reference the strength (e.g., saying "it shows a stronger correlation" instead of "it shows more correlation"), but "more" was interpreted as a description of the strength. Responses like this were assigned two points.

Table 5.20: Sample student responses to part a of Q8c regression pre- and post-assessments

Student	Pre-Assessment	Points	Post-Assessment	Points
Caden	math test score	0	I would use the standardized math test score because the data is more linear when compared to the 9th grade gpa	2
Jaime	8th grade standardized math test, given the correlation coefficient of 0.92 the relationship is quite strong between the two variables	1	8th grade standardized math test scores, since the correlation coefficient is closer to 1 then the other, which means that the predictions are close to the actual points/scores.	2
Madelyn	math score because it shows more correlation.	2	the math scores will be more accurate. because the r or correlation between the math scores and gpa is stronger than the verbal test score and gpa.	2

Similar to the other free response questions, students provided more details to their responses to this question in the post-assessment. For example, Caden's response referenced the linearity of relationships, where the relationship that is more linear with 9th grade GPA (math test scores) was preferred over the other (verbal test scores). Jaime and Madelyn's responses also suggested that they were comparing relationships in the post-assessment, noting that math test scores appeared to have a stronger correlation with 9th grade GPA than the verbal test scores.

Chapter 6: RQ3 (Statistical and Data Scientific Practices)

Here, I present the findings for RQ3a: How do participants' statistical practices evolve through the course of a teaching experiment that uses a social justice-oriented approach to teaching statistics? The goal of this question is to document how participants engaged with the statistical investigation cycle, focusing on the intersection between statistical and data scientific practices with critical practices. As discussed in Chapter 3 (Methods), I conducted pre-teaching experiment task-based interviews with five students, but only four students (Monse, Elenai, Jaime, and Robert) completed the post-teaching experiment task-based interviews. Only the four students that completed the pre and post interviews are included in this analysis. During the task-based interview, students were asked to use the Common Online Data Analysis Platform (CODAP; The Concord Consortium) to explore data about local schools. The Task is shown in Appendix 3. The task is similar to Gould et al. (2017) and has three main parts: (a) reading an article introducing the problem context, (b) answering follow-up questions about the article that includes a thought experiment about designing a statistical study, and (c) responding to a statistical open-ended task using the CODAP software. The open-ended task in part (c) was:

> Task: DHUSD would like you to make one or two recommendations that would address the following questions. Which three schools should we visit and why?

Table 6.1 shows the number of statistics courses that students took before the summer TE, gender and racial or ethnic identity, if the students are a pre-service mathematics teacher (PSMT), and the students' response to how comfortable they are with talking about race and racism in mathematics classrooms. The gender and racial or ethnic identity questions were asked as free-response questions. Furthermore, Monse, Elenai, and Jaime are all pre-service mathematics teachers. The last question was asked during the beginning activity of the second class.

Table 6.1: Demographics of students who participated in the task-based interviews

Student Name	# statistics courses taken before this course	Gender Identity	Racial/ Ethnic Identity	PSMT	I am comfortable talking about race and racism in mathematics classrooms
Monse	1	Female	Mexican-American	Yes	Strongly Agree
Elenai	1	Female	Bi-racial	Yes	Strongly Agree
Jaime	2	Male	Latino	Yes	Slightly Agree
Robert	2	Male	White	No	Strongly Agree

In the rest of this chapter, I begin by presenting an overview of how the students engaged with phases of the PPDAC Statistical Investigation Cycle (Problem, Plan, Data, Analysis, Conclusion) in the pre-interview and compare that to the post-interview. The goal of this part of the analysis is to describe the different pathways and trajectories that students took in the pre- and post-interview, focusing on how the goals that students had when engaging with the task. I then present the different practices that students exhibited during the interviews, highlighting the practices that emerged in the post-interview.

PPDAC Statistical Investigation Phases

The phases of the PPDAC cycle were each coded using elements of Miles et al.'s (2020) tactics for generating meaning in qualitative data and their implications for coding (discussed in Chapter 3). An overview of the qualitative coding process is shown in Figure 6.1. For the PPDAC cycle, I began using the PPDAC phases as the cluster codes. I also included an Other code for instances that were not related to the task (e.g., when I would give students permission to share their screen, when dogs made guest appearances on our video calls, when students asked questions about how to use CODAP). Since the data was collected for the students (e.g., they did not need to go and find a data source), I grouped the Plan and Data phases together. I often relied on to what students were saying (e.g., "I'm just looking through the data to see what's here" was coded as the Plan / Data phase because students were gaining familiarity with the data) and what students were doing on CODAP (e.g., making a visual related to the statistical question that students provided was coded as part of the Analysis phase). There were also moments when students were working with CODAP without saying what they were doing. In those cases, I usually asked students to clarify what they were doing in the last few seconds or minutes, which was then used to code the previous segment.

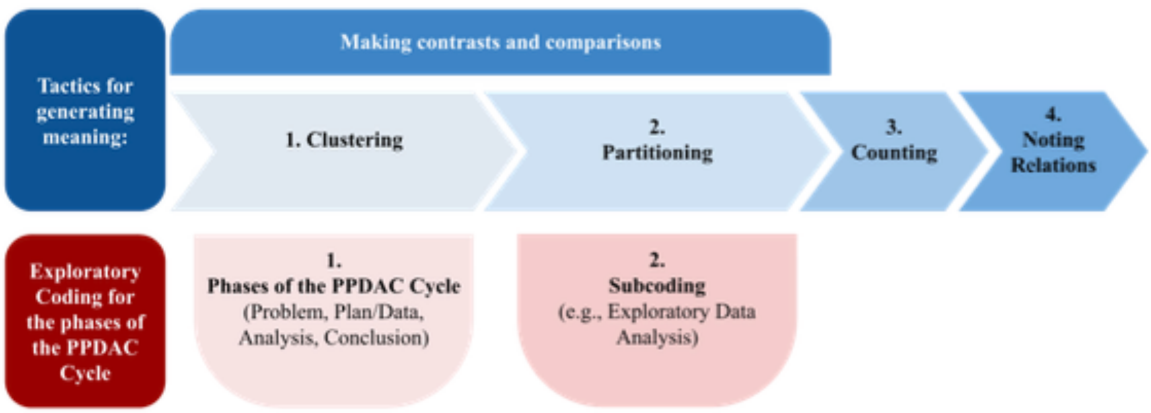


Figure 6.1: Overview of the qualitative coding process for the PPDAC phases in the pre- and post-task-interviews

The next phase of the analysis included partitioning any possible cluster codes. This happened after all of the pre- and post-interviews were analyzed. Here, I inferred that students were engaging in what may be considered an Exploratory Data Analysis (EDA). The EDA is a specific form of analysis where statisticians and data scientists may perform an initial investigation of the data to discover anomalies, check statistical assumptions, generate descriptive statistics, make any preliminary graphical representations, or explore patterns. This typically occurs before formal analyses related to a specific research question or goal. Thus, I added a separate code for analyses that were more exploratory than formal. In this dissertation, the main difference between the EDA and Analysis phases was that students explicitly stated that they were analyzing data to be familiar with the data and not necessarily to answer a particular research question.

The cluster and partitioning phases of the analysis resulted in the code definitions and examples shown in Table 6.2. The colors in Table 6.2 correspond with the colors that are used for the respective codes in the remainder of the analysis.

Table 6.2: Code definitions and examples for the phases of the PPDAC cycle

Code	Definition	Examples
Problem	Identifying statistical questions and the context	Identifying the statistical question (i.e. question about variation) to be addressed, identifying the systems or structures at hand
Plan / Data	Considering how data is collected, defined, stored, and cleaned	Identifying what data will be needed to address the question of interest, what tool (e.g., survey) or procedures (e.g., sampling, randomization) to be used, addressing missing data, data formatting and storage, planning the analysis
EDA	Initial analysis about the underlying structure of data, not necessarily related to a research question	Discover anomalies, checking statistical assumptions, generating descriptive statistics, making any preliminary graphical representations, exploring any other patterns (usually within a group)
Analysis	Identifying patterns related to the question, related to the research question at hand	Identify patterns in the data that are directly tied to the question of interest, enacting the analysis, generating hypotheses. May include interpreting findings that are not directly related back to the problem context (e.g., interpreting the correlation but not identifying three schools)
Conclusion	Summarizing, communicating, and relating findings back to the problem context	Relating the findings back to the original question (identifying three schools to visit). Preparing conclusions and presenting information to others. May include follow-up questions

Overall Time Spent Across the Statistical Investigation Phases

Figure 6.2 shows the time that each student spent on each phase that was coded and the combined total. All timestamps that were coded as Other were removed from this analysis. The total time (in seconds) for each student is shown in the black text above the bar. In the pre-interview, time spent in the Problem phase ranged from 5% (Robert) to 25% (Monse) of the total coded time and the Plan/Data phase ranged from 6% (Robert) to 30% (Jaime). Monse and Robert were the only students that engaged in any type of Exploratory Data Analysis (EDA; 12% and 10%, respectively). Between 40% to 79% of all the students' coded time was spent in the Analysis phase. Notably, Robert did not spend any time in the Conclusion phase in the post-interview.

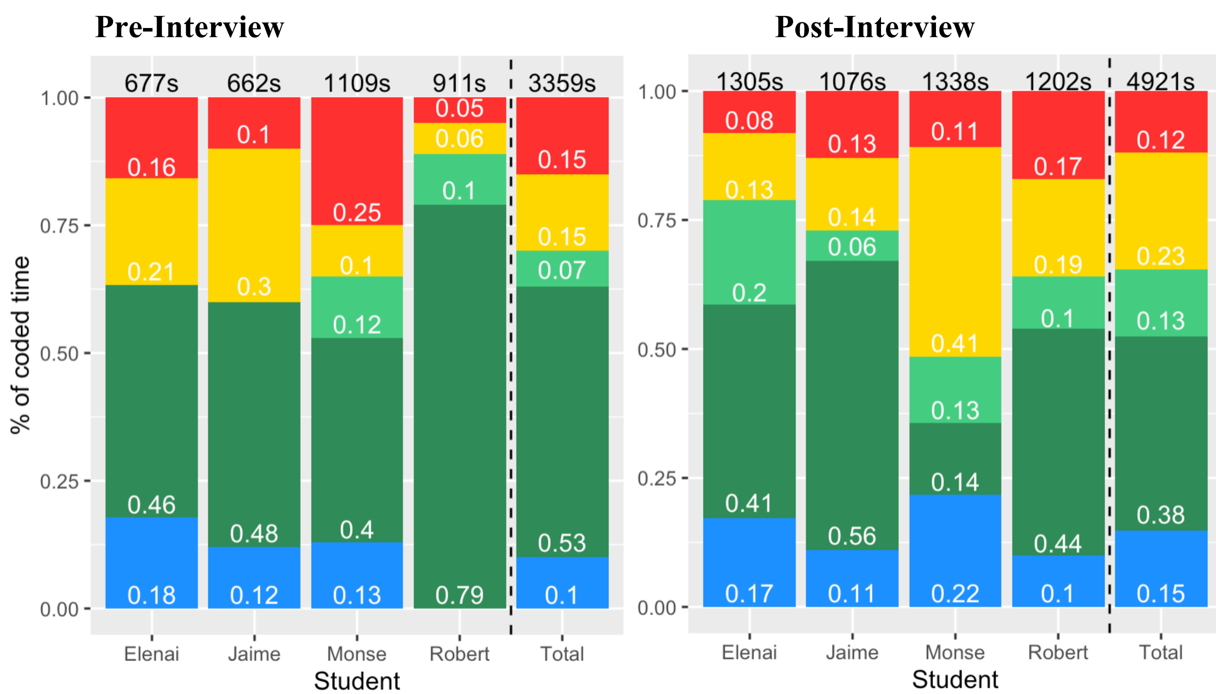


Figure 6.2: Time (in seconds) spent on each of the coded phases of the statistical investigation phases in the pre- and post-task-based interviews. The total time in seconds is shown in black text above each bar

In the post-interview, all students spent between 8% (Elenai) and 17% (Robert) of their total coded time in the Problem phase and all students spent time in the EDA phase. Combined, all students spent between 67% (Monse) and 76% (Jaime) of their total time in the Plan/Data, EDA, and Analysis phases. However, compared to the pre-interview, there was more variation in the time spent in each phase. Particularly, the Plan/Data phase ranged from less than 3% (Elenai) to 41% (Monse), the EDA phase ranged from 6% (Jaime) to 20% (Elenai), and the Analysis phase ranged from 14% (Monse) to 56% (Jaime).

Comparing across the pre- and post-interviews, Elenai and Jaime spent less time in the Plan/Data phase in the post-interview than the pre-interview. One possible interpretation for why the Plan/Data phase decreased is because the data used in the post-interview was similar to the data used in the pre-interview and the data was a subset of the data used in class, which may be why they asked fewer questions about data definitions in the post-interview than the pre-interview. Monse spent the most time in the Plan/Data phase during the post-interview (about 41% of the total code time). This may be because she took time to get familiar with the data in the beginning of the post-interview whereas she did not take the time in the pre-interview.

In terms of the EDA phase, all students spent time in the EDA phase in the post-interview whereas only two students (Monse and Robert) spent time in the EDA phase in the pre-interview. Monse's and Robert's time in the EDA phase stayed approximately the same.

Finally, in terms of the Analysis phase, Monse spent about 40% of the total coded time during pre-interview in the Analysis phase but only 14% of the total coded time during the post-interview in the Analysis phase. One interpretation of why Monse spent less time in the Analysis phase is that she allocated most of her time to getting familiar with the data (either by looking through the spreadsheet in the Plan/Data phase or making graphs in the EDA phase). In fact, the Plan/Data and EDA phases may have served as a preliminary analysis that helped her efficiently analyze data. Similarly, Robert's time in the Analysis phase decreased from 79% to 44%. Similar to Monse, this may be because Robert spent more time getting familiar with the data and engaging in an exploratory data analysis that helped facilitate a formal analysis.

Transitions and Pathways in the Statistical Investigation Phases for Each Student

To further explore how students engaged with the different phases that were coded, I created a timeline showing the time spent on each code, shown in Figure 6.3. This is similar to how Gould et al. (2017) tracked pathways and transitions within their data cycle template. The goal of presenting Figure 6.3 is not to identify a "correct" or "successful" approach to engaging with data or to compare students but rather to identify new transitions and pathways that emerged in the post-interview. As mentioned above, all timestamps that were coded as Other were removed from this analysis.

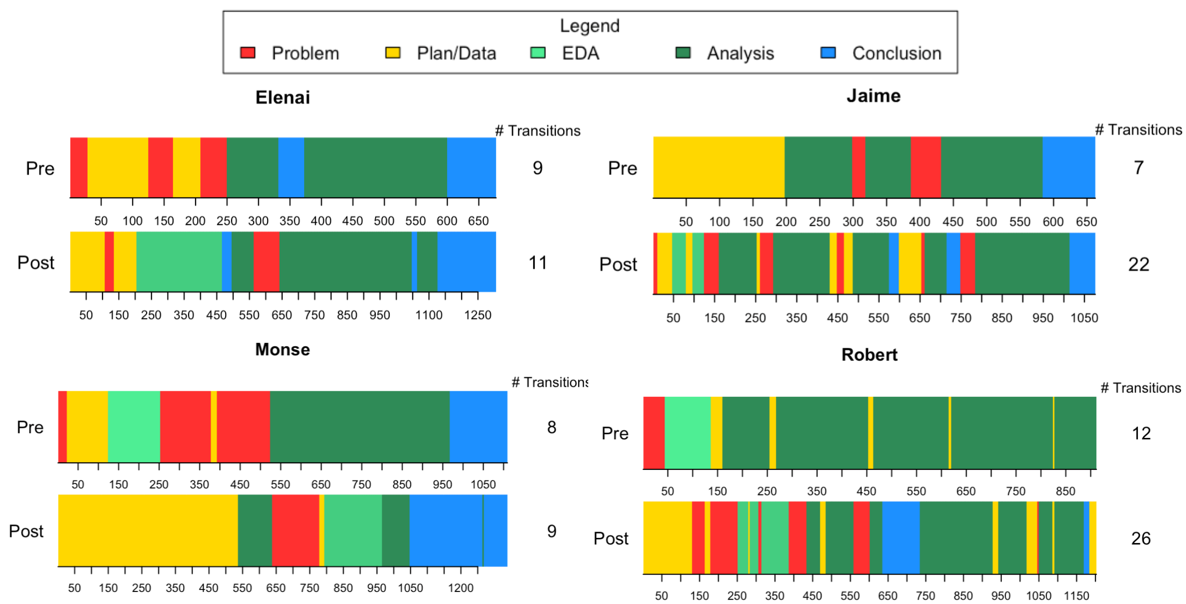


Figure 6.3: Timeline of the PPDAC phases in the pre- and post-interviews for all students

In terms of the transitions between for each student, all students increased in the number of phases that they transitioned across from the pre- to the post-interview to varied extents. In particular, both Elenai and Monse transitioned across eight phases in the pre-interview and nine in the post-interview. Robert had the largest number of transitions in both the pre-and post-interviews, increasing from 12 transitions in the pre-interview to 26 in the post-interviews. This may be because Robert had multiple research questions about which school to visit in each interview that built on each other (e.g., was looking at different variables), whereas Elenai and Monse were focused on one research question (e.g., one set of data) in both the pre- and post-interview. Finally, Jaime had the largest increase in the number of transitions, from seven in the pre-interview to 22 in the post-interview. Similar to Robert, Jaime's post-interview included multiple research questions, each appearing to have their own PPDAC cycle. As one may expect, the number of transitions appears to be associated with the number of research questions that the students had throughout the interview, where fewer research questions are associated with fewer transitions and more research questions are associated with more transitions.

In terms of the different pathways for each student, it is important to note that the PPDAC phases describe different ways of interacting with data, but the phases are not necessarily sequential. This is reflected in both the pre- and post-interviews, where students often alternated through multiple phases of the PPDAC cycle before reaching a conclusion. To further explore each students' statistical investigation pathways in the interview, I provide a summary of each of the students' pre- and post-interviews and compare and contrast both of their interviews.

Elenai Pre and Post PPDAC Phases

Elenai began by cycling between the Plan/Data and Problem phases in the first third of the interview. In fact, a theme across all interviews was that there was a lot of interaction between the Plan/Data and Problem phases (e.g., students would look through the data then formulate research questions depending on what was given). A transcript of this interaction is shown below in Table 6.3 with the coded PPDAC phases on the right. Prior to the interaction in Table 6.3, Elenai spent about 34 seconds "just looking through the data" (Elenai). Then, she began by talking about the racial context of the data, noting that there are some "bad" (in air quotes, Line 1) neighborhoods that teachers may not want to work in. This statement may be

Table 6.3: Elenai cycling through the Problem and Plan/Data phases in the pre-interview

Line	Speaker	Utterance	PPDAC Phase
1	Elenai	So part of what I've learned is that they use race to look at, kind of, like how students are doing in their school. Because way back when, I know that if you were a minority, then you were basically in the school of your neighborhood, **and the majority of those neighborhoods were bad ((makes air quotes gesture)) neighborhoods, and most good teachers did not want to work there. So you were not given equal opportunities as others (that) didn't live in your neighborhood.**	**Problem**
2	Elenai	Now, what I want to ask and, what is percent FRPM?	**Plan / Data**
3	Kevin	Yeah, so FRPM is free and reduced priced meals	
4	Elenai	What’s the BIPOC?	
5	Kevin	Black, Indigenous, People of Color. In this data, everyone who is not identified as White.	
6	Elenai	Alright. Cool. Well, so one question that could be asked, or that I would ask to find out more on would be: for schools that have higher percentages of Black, Indigenous, and People of Color, are they doing as well as other schools that the (BIPOC) percentages are lower?	**Problem**
7	Kevin	I'm going to add that into the chat just so we have it for reference. But we're gonna go ahead and start off with that question. Wait, actually, how is that going to help you find which schools to visit?	
8	Comment	28s pause	
9	Elenai	So, this is percent that met or above eighth grade level English?	**Plan / Data**
10	Kevin	Yea, it’s the percent of students that met or exceeded the standards for English or Math, for sixth grade, seventh grade, and eighth grade.	
11	Elenai	Okay. (8s pause) So then, I could compare just those two? That's English. I would want to know English or Math, though. Okay yea. That’s my question.	**Problem**
12	Comment	Elenai begins analysis	

Note: ((Parenthesis)) are used to show comments. Bold is added to emphasize statements. [Brackets] are used to show text that may be missing from the text interpreted as a deficit statement because it may suggest an assumption about certain neighborhoods being less than others. However, it is unclear if Elenai is suggesting that there are "bad" neighborhoods or that her use of air quotes is indicative of the acknowledgement these are deficit narratives that surround educational contexts. Then, Elenai asked some clarifying questions about acronyms (FRPM in Line 2, BIPOC in Line 4) before beginning to frame a question for the task (Line 6). She then went back to clarify some data definitions (Line 9) before confirming that the question from Line 6 was her final question (Line 12).

Elenai spent the remainder of the interview in the Analysis and Conclusion phase, focusing on the same question from Line 6. Her analysis mainly included creating scatter plots and looking for any patterns in those plots (mainly, linear relationships). Elenai ended the interview without providing three schools to visit (the initial question of the task). However, once the task ended, I asked Elenai to expand on some of her decisions, shown in Table 6.4.

Table 6.4: Elenai reflecting on the analysis in conclusion during the pre-interview

Line	Speaker	Utterance
1	Kevin	Okay. We're done with [the problem-solving part of] the interview. Thank you so, so much. I was curious, if I was a teacher and I came to you and I said. Let me try to phrase this. I said, I think that the more Students of Color we have, the lower test scores are going to be. How would you react to that or what are your thoughts on that?
2	Elenai	**According to the numbers, I think you're right**. Sadly. But I would say, according to that, maybe we need to reach those students by doing different things in order to reach those students to increase our scores. Yeah, it's a great, a great problem that we need to work on as teachers…**like it's not bad to have more Students of Color, so it's just weird to say that**

One direct interpretation of this statement is that Elenai may agree that increasing the proportion of BIPOC students decreases the students' score. If so, then this statement may be considered deficit because it places the responsibility on groups of people, makes an attribution to race, and does not account for the larger historical structures of social injustice. However, Elenai qualified her response with “According to the numbers,” which may be interpreted as her disassociating from the claim. That is, she is noting a difference from what she believes and what the numbers are implying. Similarly, she explicitly states how “weird” it is to associate bad test scores with Students of Color. Combined with other statements during the interview (e.g., saying that “I know that. there's so much talk about,” “these kids are more than just numbers,” or “I don’t think that these numbers have, like, the full picture” when talking about grade disparities), it is possible that Elenai was navigating the tensions in the problem-solving part of the interview between what the “numbers” are implying and her own personal beliefs (e.g., staying away from deficit interpretations) and identity (e.g., as a bi-racial, Black and White, woman). It is important to note that I did not explicitly ask participants how their personal beliefs or identity shaped how they engaged with the data. It would be worth including a question about this in future iterations of this interview.

In the post-interview, Elenai spent about 198 seconds (or about the first 16% of the coded time) cycling between the Problem and Plan/Data phases. The majority of this time focused on Elenai drafting questions for the task, shown in Table 6.5. She began by asking clarifying questions about free and reduced priced meals (FRPM, Line 1) and the percent of students that scored a 150 or higher on a standardized assessment (PctCE150, Line 3). She identified a question in Line 5 about comparing teachers’ teaching experience to different data. She spent the rest of this part clarifying data definitions and identifying potential data that may be interesting to relate to the teachers’ teaching experience.

Elenai then moved on to an *EDA* that consisted mainly of looking at the spread of the data. She then revisited her question and continued on a formal analysis comparing the teachers’ experience to other data. Turning to the goal of the task (identifying three schools to visit),

Table 6.5: Elenai cycling through the Problem and Plan/Data phases in the post-interview

Line	Speaker	Utterance	PPDAC Phase
1	Elenai	What is percent FRPM?	**Plan / Data**
2	Kevin	Percent of students that qualified for free or reduced priced lunch	
3	Elenai	What is P-c-t-CE 150?	
4	Kevin	The percent of students that got 150 or more on some standardized test, which I think was like a cut off, like an important cut off	
5	Elenai	Okay, so I found it interesting that first your teachers, I would want to see where first year teachers are compared to the percentage of People of Color, test scores. (17s pause)	**Problem**
6	Elenai	Yeah. And then I'd want to do the same thing with third year teachers and students.	
7	Elenai	What's teacher absent and teacher this year?	**Plan / Data**
8		Teacher absent is a percent of teachers that were absent, the average. The average percent per day. And, what was the other one?	
9	Elenai	Teachers year.	
	Kevin	The percent of teachers that returned for another year. From year one to year two, whatever year one was.	
11	Elenai	The num teachers. What does that do?	
12	Kevin	The number of teachers at the school	
	Comment	Elenai continues to look through the spreadsheet	

Elenai ended this part of the analysis by suggesting that they should visit schools with high test scores. However, she also added that (bold added for emphasis, double parentheses added for comments):

> I would say that it ((the top three schools)) **don't represent our population**. So, yeah, we can look at them, but it doesn't-, you won't find what we're looking for in order to improve for every school. What we do need to look at is where some of the, where some of the schools are. So, what I would look at let's see, say we're looking at ((school district)), **we don't have to leave the district to go to a White school or a better school.** We just have to go to a school that's doing better than us, but similar to us, and see what, what's different with them, why is it working for them, but it's not working with us. Yeah, and that's the whole thing if we're, if we're, if our discussion is about making things more equitable because of the numbers, then you want to look at, where are they showing that it is equitable?

> And what schools are doing right. **Not just go to where it's more White or they have higher test scores because whatever**. But you want to keep, I would want to **keep it in the same ballpark**. And just going to White schools or rural or whatever, it's not the same. Yeah because ((school name)) has the lowest percentage of Students of Color and they have the higher test scores, but there's other things like money too.

Here, Elenai ultimately decided that the schools that she would identify to visit depend on what schools we are trying to compare and should go beyond comparing test scores. For example, she states that the three schools might not "represent our population" and provides an example of one school that has high test scores but is majority White. Similarly, she states that going to schools that are "more White or they have higher test scores" may not be in the same "ballpark" as other students. Statistically, Elenai may be referencing the tensions in the problem-solving part of the interview with the generalizability of which three schools she decides to visit and how representative the schools are, where visiting schools with the three highest test scores may lead to some racial biases that set schools that are majority White as a standard to be compared to.

She expands on these tensions after the problem-solving part of the interview. I asked her the same question that I asked at the end of the pre-interview, shown below in Table 6.6. Unlike the pre-interview, Elenai explicitly says that saying that increasing Students of Color decreases test scores "is not right" because there's "so much more than the student." She states that other factors include but are not limited to the school, curriculum, access, and tests. Turning to the first design feature of the course (DF1: Reflect on structures of social injustices), one interpretation of Elenai's statement is that she is considering the social, political, cultural, and historical factors of social and racial injustices and, in doing so, shifting the responsibility away from individual students towards the larger structures of social and racial injustices.

Table 6.6: Elenai reflecting on the analysis in conclusion during the post-interview

Line	Speaker	Utterance
1	Kevin	So, then, second question, how would you react to someone that says that increasing Students of Color decreases your test scores?
2	Elenai	I would say that we need to figure out what the school's aren't doing or what what's, or how are we losing that? You could say that, sure. **But you need to look at what's going on**. **Saying, oh it's because, it's because of Students of Color is not right because there's so much more**. There's the **curriculum**, and access, and the tests, and everything. **It's so much more than the student.**

In terms of similarities between the pre- and post-interviews, Elenai spent about the first third of each interview drafting questions and getting familiar with the data, either by asking questions about data definitions (pre- and post-interview) or through an EDA process (post-interview). The Analysis and Conclusion phases also remained relatively the same, consisting of creating visuals and interpreting those visuals. Furthermore, both the pre- and post-interview questions that Elenai asked were statistical questions and Elenai ended each interview by focusing on some of the tensions in the racialization of data (e.g., associations between the proportion of students identified as Hispanic or Latino and test scores).

In terms of differences between the pre- and post-interview, Elenai spent about 141 seconds (about 12% of the total coded time) in the EDA phase in the post interviews whereas she did not spend any time in the EDA phase in the pre-interview. Furthermore, Elenai did not provide three schools to visit in the pre-interview, but did provide an approach for how she would identify three schools during the pre-interview that focused on representation and generalizability. Her approach to identifying three schools may be interpreted as one that was accounting for generalizability and cautious of the racial biases that may be implicit in the three schools that she selects. Finally, Elenai appeared to note some of the tensions and have a more anti-deficit approach in the post-interview. For example, in the pre-interview, Elenai noted that there were "bad" schools that teachers may not want to work at, but as noted above it was unclear if she had a critique of that deficit statement. However, in the post-interview, Elenai had a more clearly anti-deficit (and possibly anti-racist) approach where she explicitly critiqued interpretations that may perpetuate negative stories about BIPOC students. Rather, she notes that there is “so much more than the student.”

Jaime Pre and Post PPDAC Phases

Jaime spent the first 434 seconds (almost two thirds of the total coded time) in the Plan/Data and Problem phases during the pre-interview. Notably, the majority of this time (about 369 seconds) was Jaime asking questions about the definitions of data (e.g., what acronyms stood for) and getting familiar with the data (looking through the spreadsheet to see what data was available and the format of the data) before going straight into an analysis. In this first part, Jaime did not show evidence of any type of analysis besides identifying the data that was provided. If there was evidence of analysis, this would have been coded as EDA. Further, in this part of the interview, Jaime also drew on his experiential knowledge to make sense of the data and where the schools were located, shown in Table 6.7.

Table 6.7: Jaime drawing on his experiential knowledge about the local context to make sense of the data

Line	Speaker	Utterance	PPDAC Phase
1	Jaime	I'm trying to see where these schools are located. So a couple of them are inner city. But that's not really reflected in the data.	**Problem**
2	Kevin	Wait, how do you know?	
3	Jaime	Yeah well at least from what the school that I've heard of	

This moment shows an interaction between the *Problem* and *Plan/Data* phase where he was relating the collected data to what he knows about the local context. For example, he knows that some schools may be identified as inner city, but did not see that classification in the data.

After continuing to get familiar with the data, Jaime stated his question for the task:

> Okay um. Well, I guess, based on like the first observation that we were making the-, focusing on just race in general, with the Hispanic and Latino students, the Black and Indigenous, People of Color and the correlation with like the free and reduced lunch Program, as- I wonder if those numbers that, if they run pretty high, like that correlation. And with test scores.

Jaime was interested in seeing if there was a correlation between BIPOC with the percent of students that received Free and Reduced Priced Meals (FRPM, a proxy for socioeconomic status) and test scores. Jaime interpreted each graph in terms of their being a “strong relationship” or not, but did not make explicit connections to the goal of the task (identifying three schools) or name any tensions between some of the relationships. For example, he said that schools with “more Hispanic (students), (have) lower scores” but did not name any tensions or note possible deficit interpretations.

For the remainder of the interview, Jaime was in the Analysis and Conclusion phases. Notably, Jaime’s analysis did not include any visuals and was primarily ordering the columns and “going down the table to see how the numbers are similar” (Jaime). He ended the interview by summarizing and interpreting his analysis. In terms of the goal of the task (identifying three schools to visit), Jaime did not identify three schools.

Jaime cycled through many more phases in the post-interview (seven in the pre-interview and 22 in the post-interview). One possible reason for this is that Jaime cycled through more than one question, each which built on the previous question(s), shown in Figure 6.4. In the first 123 seconds and similar to the pre-interview, Jaime started off by referring to his knowledge about the local context. In the EDA phase, Jaime also graphed test scores in relation to race and ethnicity data and free or reduced priced meals (FRPM). This was followed by Jaime cycling through the Plan/Data phase (mainly asking about data definitions) and an exploratory data analysis (mainly looking at the distribution of data).

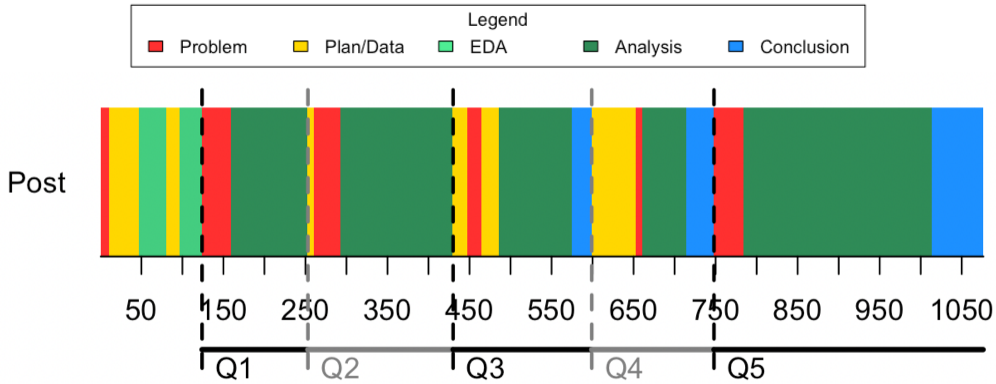


Figure 6.4: Timeline of the PPDAC phases in Jaime's post-interview by question that was asked

The remainder of the interview included exploring four different questions. The four prompts are shown in Table 6.8. The first question was generally about how racially and ethnically diverse schools were. After creating some visuals, Jaime asked the second question about school funding, focusing on who gets the most funding. The second question was about comparing funding of institutional aids across magnet school classification.

Table 6.8: Five questions explored by Jaime in the post-interview

Question	Utterance
Q1	Okay, so I guess for this question, we could, one of the questions would be: How, or what's the diversity? How diverse are the school districts?
Q2	Okay, so looking at schools with, what about the amount of funding that they get? Who gets the most funding by magnet or not? ((institutional aids))
Q3	Which schools have higher person expenditures? ((institutional aids, teachers, and staff))
Q4	I guess, one of the questions would be what would cause this school to receive or to have more, to get more spending money? Or not cause, but like who gets the most money? ((institutional aids, teachers, and staff))
Q5	So yeah what would cause this school in particular that's within the same district as all these other schools to be above average, more than the other schools? ((institutional aids, teachers, and staff))

The analysis also included creating a visualization to compare distributions. This second question prompted the third question (generally about the spread of school expenditure, including expenses for institutional aids, teachers, and staff), which prompted the fourth question (distribution of school expenditure for teachers) and the fifth question (how school expenditure for teachers varied across different data like FRPM, BIPOC, number of counselors, and number of AP classes offered). The third, fourth, and fifth analysis were also followed by a short conclusion where Jaime summarized and interpreted the findings and briefly connected them to the goal of the task (identifying three schools).

After the interview, I asked Jaime to clarify why he selected race, ethnicity, and funding as his primary data to investigate. He said:

> I think it's more, **more on the political side**, like how certain districts, or have certain schools receive certain funding because they're in nicer neighborhoods and have higher scores as opposed to like inner city schools that have lower funding for their schools. Or, I guess, property taxes and that. **So I wanted to see if funding was the issue with tests, not like just being Hispanic**…I think that before I did Hispanic and test scores but **that doesn't explain it all, but like funding is a bigger issue that maybe explains it more**. Like schools with more resources might have better scores, so I wanted to see that.

This response indicates that Jaime recognizes that the educational data is political. For example, from the EDA phase in the beginning of the interview, he noticed that test scores varied by demographic markers (race and ethnicity, FRPM). However, he stayed away from making conclusions that relate race and ethnicity data to test scores because it "doesn't explain it all." This may partially be because he wanted to stay away from creating any possible deficit interpretations, or perhaps may have been cautious about how others would interpret his findings. Instead, he related funding to test scores because it "maybe explains more" than looking at data on race or ethnicity. This may be motivated by the first design feature of the TE (DF1: Structures of Injustice) where students were encouraged to look at larger systemic or structural causes of injustices instead of placing responsibility on individual people or groups of race or ethnicity.

It is worth noting that it is unclear if Jaime initially thought that race and ethnicity was related to funding. Statistically, that may have implied that race or ethnicity and funding were confounding variables (e.g., proportion of Hispanic or Latino students may affect the funding or vice versa). However, as part of the third question (Which schools have higher expenditures?), he did note that there was no relationship between race and ethnicity with the school's funding and proceeded to use funding in the remainder of the analysis. Thus, it is possible that he was generally interested in data about systems or structures (funding) over demographic or individual data (race or ethnicity). Furthermore, other students often had questions with multiple parts (e.g., Elenai compared the school's average teacher experience with the percentage of BIPOC students, test scores, and other data of interest), but Jaime's questions were new questions that built on each other. That is, new questions emerged as Jaime analyzed data and they were more specific as he progressed (from general expenditure to how teacher expenditure varied across different data) and included more data (from institutional aids to all types of expenditure). Personally, this is reflective of my own experiences engaging with data, where I typically start with a goal that snowballs into different analyses.

In terms of similarities between the pre- and post-interviews, Jaime began both interviews by drawing on his knowledge about the local context (e.g., schools that may be identified as "inner city") and looking through the data to see if the data included those classifications. In terms of differences between the pre- and post-interviews, Jaime had more transitions between phases in the post-interview (22) than the pre-interview (7). However, the post-interview also included five separate questions that built on each other and got more complex as he progressed. For each question, Jaime generally followed a pattern where he cycled between the Problem and Plan/Data phase before Analyzing data, then often ending with a short

Conclusion. A second important difference was that Jaime used more visuals. In particular, he did not use any visuals in the pre-interview, but made at least two visuals for each analysis in the post-interview. Finally, Jaime noted some tensions with how race and ethnicity was used in this data. Particularly, it appears that Jaime was cautious about noting any correlations or implying causation between race or ethnicity and test scores. Thus, he focused on using funding instead of race or ethnicity.

Monse Pre and Post PPDAC Phases

Monse spent almost the first half of the interview (523 seconds of the total coded time, or about 47%) getting familiar with the data and problem context. Particularly, in the first 252 seconds of the total coded time, she began by clarifying the goal of the task, and noting that "I'm not 100% sure what it means when it's talking about, like, highlight awesome work" (reading the prompt). She then added, "okay, umm, let me see what data we have" and engaged in an EDA phase where she mainly looked at the range (mainly, the minimum and maximum values of the data) of the given data in the spreadsheet. Notably, Monse did not create any visuals. Statistically, Monse may have been thinking about what data is needed to respond to the task ("highlighting awesome work") as well as if the data was robust enough to respond to the task and data measurement (e.g., how is "awesome work" quantified).

In the following 271 seconds, Monse described the problem context, asked some clarifying questions about the data definitions, and stated her overall goal. Specifically, she said:

> Okay, I'm looking to see because, if like, we're looking at students' academics, right? We'll look at good scores. But then I'm also **looking at, like, the composition** of the school right. So like here, this one is like, okay, like most of them don't have free or reduced lunch, so there's other **opportunities that students have, right, when they have a higher income,** which could correlate to the scores, right? So, like this one would be a school if you're like, okay let's look at a middle-class school, **I'm assuming, just because there's not a lot of free reduced lunch students.** And then yeah, you would look at it and be like, okay is

> it ethnically diverse? Which like half of them are Black, Indigenous, or People of Color. Which is like you know, okay. Well, I don’t know actually. I don’t know where this area is … ((lists racial and ethnic breakdown)) Because I mean I guess you could always say, if you really wanted to just be like these are the high achieving schools and you just pick the ones with the highest scores, right? There wouldn't be any need to add ethnicity or like income status, right? **But then that wouldn’t represent schools.** That would just represent, based on your resources, what you can get. Of course, like a school, who has more resources and their parents have more money, you're going to be paying for those extra classes, right? You're going to go take those tutoring classes. **You're going to do all of this extra stuff that students in low-income areas don't always have the opportunity to do.**

Monse’s main goal was to look at the schools “composition” and identify three schools that were representative of other schools. Conversely, she did not want to pick three schools that were outliers because that “wouldn't represent ((other)) schools.” For example, she states that schools in higher income areas may also have higher resources, like extra classes and tutoring, which may lead to higher test scores. Statistically, Monse may have been thinking about confounding variables (discussed in the class) and the generalizability or representation of the schools that she picks. In terms of confounding variables, Monse appeared to be implying a relationship between test scores and income, where higher income is associated with additional resources, which may increase test scores. Thus, by selecting the schools with the three highest test scores, Monse may have been suggesting that we may also be selecting schools with higher incomes and schools that are not necessarily representative of other schools.

Monse then continued to the Analysis phase. Similar to Jaime, Monse did not create any visuals. Rather, she ordered the columns and was trying to find schools that were in the middle 50% of different school demographics (e.g., race and ethnicity, FRPM, expenditure). For each variable that Monse looked it, she removed the lower and upper 25% of the data to ultimately have a data set that she said was “the middle 50%.” Her overall approach was similar to Elenai’s and Jaime’s post-interviews in that she wanted to find three schools that were representative of the target population. However, Elenai and Jaime were interested in finding schools that were similar to a specific school whereas Monse was interested in identifying schools that were broadly representative of the entire dataset. Monse ended in the Conclusion phase by providing three schools that were generally in the middle 50% of the data that she looked at, but had some variation within those three schools. For example, she identified one school that was a charter school (with a relatively low number of students enrolled), one school that had a relatively high number of BIPOC students and students that qualified for FRPM, and one school that had relatively high enrollment. It is not clear why she chose these criteria. Statistically, this may be interpreted as a non-probabilistic stratified sampling (a sampling method discussed in class), where she first stratified to include the middle 50% of the data, then stratified again based on her criteria of interest.

In the post-interview, Monse spent the majority of her time (793 seconds, about the first 60% of the interview) exploring the data before finalizing her question to task and goal. Specifically, she started off by spending about 535 seconds in the Plan/Data phase looking through the spreadsheet, writing down data that she thought she would be interested in, and asking clarifying questions about data definitions. After removing columns of data that she was not interested in, she started to look at the overall spread of the data. This was similar to the pre-interview where she ordered the column and removed what appeared to be the lower and upper 25% of each data. Also similar to the pre-interview, she did not create any visuals.

After subsetting the data, Monse began to describe her overall approach to responding to the task (identifying three schools to visit). She stated that “like the first time, I want the middle of the group again to make sure it’s like similar to other schools.” However, this time she specifically wanted to focus on the proportion of students that enrolled in AP classes. As with the pre-interview, her focusing on the middle 50% of the data may have been guided by her interest in identifying schools that were representative of the entire dataset. Furthermore, later in the interview, she clarified that she wanted to see schools that have high proportions of students who are identified as BIPOC or qualify for FRPM because "AP classes are important for college….especially for Students of Color or poorer students." Since the data was about the number of students who were enrolled in an AP course (not the proportion), she also created a new variable for the proportion of students who enrolled in AP courses out of the total enrollment at the school. She then went into the Analysis phase, where she found schools that had a high proportion of students that enrolled in AP classes, students identified as BIPOC, and students who qualified for FRPM.

Throughout the interview, Monse also asked follow-up research questions that she thought would help her identify other schools. These questions were inspired by the data, but would not be answered with the data that was collected. She revisited these questions in the Conclusion phase and raised new questions for future analysis. For example, she stated that they could "observe, like, how their classes, how the math classes are and why students are confident." Monse also had questions about how data was collected because she noticed that some schools reported that there were more students that enrolled in AP courses than the total number of students enrolled in the school. She decided that it may have been a typo or that maybe the number of students enrolled in the AP course was not a unique count (e.g., students may be counted twice if they were enrolled in two or more AP classes).

In terms of similarities between the pre- and post-interview, Monse had approximately the same approach to identifying three schools: subset the data to include the middle 50% of data of interest, use selecting criteria to identify three schools within the subsetted data. Also, Monse was the only student that did not create any visuals in the pre- or post-interview, relying mostly on ordering the data in the spreadsheet. She was also the only student that hid data from the spreadsheet, primarily to subset the data to the middle 50%. In terms of differences, Monse spent more time in the Plan/Data and Problem phases and less time in the Analysis phase post-interview. This may be attributed to at least three reasons. First, Monse used a similar approach to identifying three schools as her pre-interview which, consequently, may have sped up her analysis process. Second, Monse spent a relatively large amount of time getting familiar with and preparing data for the analysis (the first 965 seconds, or the first 72% of the interview) that helped her identify data of interest. Third, and related to the second reason, Monse knew which data she wanted to look at prior to formally engaging with the analysis in the post-interview, whereas she added more data and relationships through creating variables in the pre-interview as new interests emerged.

Robert Pre and Post PPDAC Phases

As mentioned before, Robert had the most transitions in the pre-interview (12 in total). After asking some questions about the acronyms and looking through the data, Robert spent the remainder of the interview (752 seconds, about 83% of the total interview) cycling between the Plan/Data and Analysis phase, which contributed to why he had so many transitions. Specifically, Robert began by first stating:

> I was kind of wondering if, like, the lower the salary, the more absent teachers would be, but it doesn't look like it. I'm trying to think of, like, what things would be good to look at, but I don't know. I'm not finding anything (7s). I want to find stronger correlations.

One interpretation of Robert's approach is that he was engaging in a form of *p*-value hacking where he was trying to find correlations that may be presented as statistically significant. However, about 743 seconds into the interview, Robert stated that "no correlation is a correlation." Although he was interested in finding "stronger correlations," this may suggest that he was still considering weak correlations as informative relationships. Furthermore, Robert appeared to create hypotheses about which variables would have a correlation by drawing on his experiential knowledge about what he knew about the context, then analyzed the data to confirm or reject his hypothesis. This experiential knowledge and hypothesis driven approach was similar to how other students decided to select and analyze data, but Robert was more explicit about whether or not his conjectures about the expected relationships were true. For example, while making scatterplots of different data, he would say things like "that's not what I expected," "I didn't think it would be this way", "I was kind of expecting this one to be a stronger correlation," and "this is kind of what I expected." Thus, rather than *p*-value hacking (or only be interested in *p*-value hacking), Robert's approach may have (also) been centered around referencing correlations (estimated by looking at the linearity between variables in scatterplots) to confirm or reject his hypothesis about which variables were associated.

It is also important to note that Robert ended the interview without providing a list or method to identify which schools to visit (the goal of the task). Furthermore, Robert identified the correlation for all the pairs of data he analyzed and determined the strength of the correlation. However, he did not interpret the correlation in the context (e.g., would only identify the strength as "weak," "medium" or "mild", or "strong). Since the Conclusion phase requires that students interpret the findings in the context or connect the findings back to the goal of the task, Robert did not spend any time in the Conclusion phase during the pre-interview.

In the post interview, Robert also began by looking at the summary of the data that was provided. He then asked clarifying questions about the data (mainly, data definitions) and stated that he wanted to "find correlations again (6s) I'm going to just look around for a little bit and see if there's any correlation before I start talking" (Robert, 247 seconds into the interview). This began an EDA phase where Robert was making scatterplots. This appeared to be a formal analysis, similar to his statistical investigation process in the pre-interview. However, about 433 seconds into the interview, he specified that "I want to pick correlations with grades ((eighth grade math test scores))." This then prompted a formal data analysis where he made scatterplots and identified the strength of each correlation. He specifically looked for "things that are already, I think, like would have a correlation, yeah, so now I'm just going to throw stuff in there." This was similar to the pre-interview in that he was drawing on his experiential knowledge about the topic to create and test hypotheses about the data. Table 6.9 shows examples of these hypotheses.

Table 6.9: Robert's hypothesis about data that would be correlated with grades

Question	Utterance	Comments
Q1	I'll probably find a correlation between this and, like a **charter school, grades**. Charter is like just a nicer school basically, right?...Like **they have more money, and resources**, and, yeah?...**so they probably have higher scores?**	Initially, Robert thought that charter schools were private schools, and therefore would have higher test scores because private schools may have more resources
Q2	Total enrollment is probably something good to look at. **Maybe total enrollment and like something to do with grades**. I wonder if there's a-, I'm not, I'm not sure what I would expect there. But maybe there's a correlation between, like, how many people go to school and like how good people are, or how good people are at that school…**like my high school was pretty big and pretty under-resourced**	Robert assumed that larger schools were under-resourced, which may influence the quality of education and the test scores
Q3	How about, let's do the amount who qualify **for free and reduced meals and grades and races**.	Robert wanted to see if there were any correlations between "data about equity" (FRPM and race or ethnicity)
Q4	I wonder what the correlation between like, I mean I'm assuming-. Actually, let's throw it in here. I'm going to do, so I'm going to do **math scores in sixth grade ((x-axis)) and math scores in eight grade ((y-axis))**. I'm curious if, but, I'm assuming it would be like a line, like **diagonal**. Or, I don't know, let's see	Robert later clarifies that he's looking to see how close the data is to a linear relationship and if the y-intercept is different from zero.

Robert chose the first question because he thought that charter schools were a specific type of private school. Later we clarified that charter schools are publicly funded but run independently of the local school district. Robert assumed that Charter schools were "nicer school(s)" that have "more money, and resources" that may be associated with higher scores on standardized assessments. Notably, Robert did not necessarily critique the use of standardized assessments, but he did imply that income and other resources may influence scores. The second question was similar in that it was also about resources. Particularly, Robert reflected on his high school experiences where he went to a school with high enrollment that was under-resourced. This is similar to the QuantCrit tenet of drawing from experiential knowledge to guide analyses.

The third question happened later in the interview after Robert had stated that "I wish there was, I would just be looking to see if there was something about diversity." He then said that FRPM and the race and ethnicity data was similar to what he wanted, but he wanted something that was "on a scale of zero to 10 of, like, how diverse the school was." Since that data was not available, he instead used the FRPM and each race and ethnicity data separately. He started with FRPM demographic data, then the proportion of students identified as Black, then the proportion of students identified as Hispanic or Latino students (Figure 6.5). For each graph, he noticed that there was a negative correlation between the demographic data and the test scores (e.g., low test scores associated with schools that have higher proportions of students identified as Hispanic or Latino).

After looking at the graph for a few seconds, Robert said:

> What would give a reason, why this is true? **I don't think it's because they're Hispanic and so they're bad at math**, right. Like you can't say that, **it sounds bad.** But it could have to do with, like, the school, so I'm assuming…((pauses to look at graph in Figure 6.5)) **So it probably just has to do with the quality of the school**, but **I don't know if I have enough information to really say that right now**. But like **definitely not the students, not because they're Hispanic.**

This analysis and interpretation was similar to how Elenai and Jaime were cautious about interpreting the negative relationships between BIPOC students and the scores on the standardized assessments. Particularly, Robert did not place the responsibility on the students or groups of students (e.g., saying that they have low test scores because they are Hispanic or Latino) and rather considered the larger structures at play (e.g., quality of the school). This is directly related to the first design feature (DF1: Reflect on structures of social injustices) that was designed for students to focus on the larger structures of social (in)justice rather than placing the responsibility or blame on an individual person or group of people. Furthermore, Robert also added that having a metric on the quality of the school would be helpful, but he did not have that data. Similar to how he wanted a measure for diversity, he ended this analysis by saying that "I also wish that we had something about, like, the quality of school, also like from zero to 10…but, like, I don't even know what that would be."

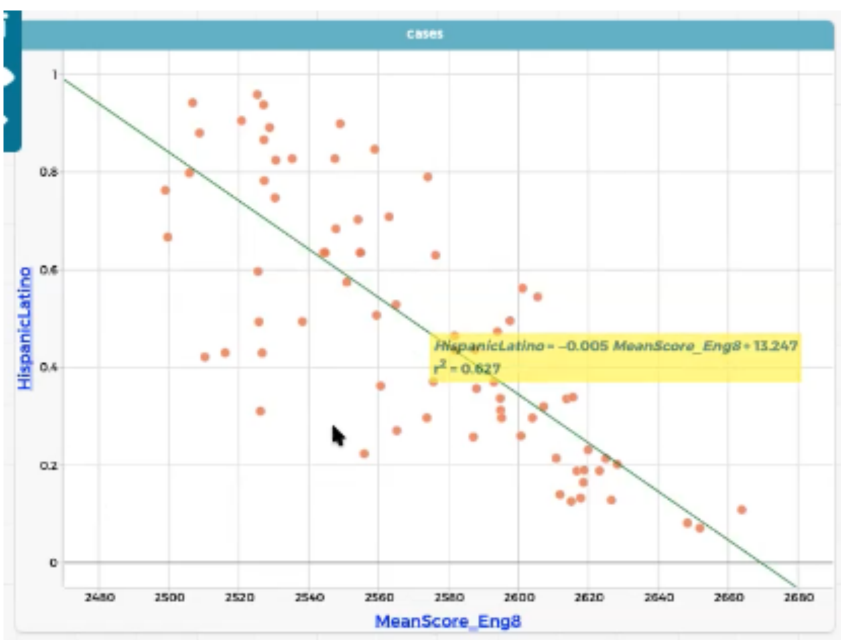


Figure 6.5: Scatterplot of the mean assessment score for eight grade and proportion of students identified as Hispanic or Latino in Robert's post-interview

The final analysis was relating the sixth-grade mathematics scores with the eighth-grade mathematics scores. The scatterplot plot showed a line that was similar to the line y = x. When interpreting the slope, Robert was considering how linear the relationship was and how close the y-intercept was to zero. Although CODAP does not provide a formal regression analysis (e.g., with test statistics, p-values, confidence intervals), Robert used the scatterplot and regression line provided as a form of a regression analysis. For example, for the slope, he clarified that "the slope doesn't matter, or, actually, I think it just has to be positive" because a positive slope shows an increase in the eighth-grade test scores given an increase in the sixth-grade test scores. Conversely, a negative slope would show a negative relationship that implies that an increase in the sixth-grade test score was associated with a decrease in the eighth-grade test score. That is, the higher the sixth-grade score, the lower the eighth-grade score.

For the y-intercept, Robert was curious to see if the y-intercept was at least zero, or if students were "learning consistently." This interpretation was possible because we assumed that the assessments used for this data were aligned to the appropriate grade level. Specifically, Robert said that a school that had an average of 160 in the sixth-grade test and the eighth-grade test may be interpreted as a school that "met standards before and after" whereas a school that had 160 in sixth grade and 180 in eighth grade "met the standards, then exceeded the standards." Here, Robert appears to be using similar language to the achievement level classifications of standardized assessments (e.g., standard not met, standard nearly met, standard met, standard exceeded). Thus, one interpretation of a y-intercept of zero with a line that has a slope of approximately one (as is with this case) is that the school achievement classification stayed the same. A y-intercept greater than zero might indicate that the school achievement classification increased and a y-intercept smaller than zero might indicate that the school achievement classification decreased. Ultimately, the y-intercept was approximately zero.

Robert ended the interview without providing three schools to visit or a method to identify three schools. It is unclear why he did not provide three schools, but one possible reason was because he was so invested in the analysis and interpretation of the scatterplots. This is similar to my experiences engaging with data, where I often get lost in the data and run across analyses that are really interesting and tell a great story, but may not necessarily address the initial goal. In this sense, it is possible that a data scientific and statistical practice is engaging with data in a way that is relevant with the main goal while also keeping track of new analyses that may be worth exploring in future work.

In terms of similarities between the pre- and post-interview, Robert was the only student who spent time looking at the summary of the data that was provided. All other students chose to look through the spreadsheet in CODAP in the pre- and post-interview. Additionally, Robert was the only student that used any statistics in the pre- and post-interviews, mainly correlations. A third similarity between the pre- and post-interview is that Robert did not provide a list of schools or a method to identify which schools to visit (the goal of the task). It is unclear why Robert did not provide a list or method to identify schools to visit. In terms of differences, Robert had a narrower focus in the post-interview (comparing demographics to standardized test scores) whereas he was looking for any correlations in the pre-interview. Furthermore, Robert also engaged in his own form of regression analysis in the post-interview that was similar to some of the analysis that we had discussed in class.

Summary of the PPDAC Phase

I end this part of the analysis by comparing and contrasting overall patterns in the pre- and post-interviews. A summary of selected patterns across the pre- and post-interviews is shown in Table 6.10.

Table 6.10: Summary of selected patterns across the pre- and post-interview

Observation	Pre-Interview	Post-Interview
Provided three schools or method to find three schools that considered generalizability and having a representative sample	Monse	Elenai Jaime Monse
Had multiple questions that build on each other		Jaime Robert
Engaged in an Exploratory Data Analysis (EDA)	Monse	Elenai Jaime Monse Robert
Used visuals in the analysis	Elenai Robert	Elenai Jaime Monse Robert
Used statistics measures (e.g., correlation)	Robert	Robert
Used regression lines or informal regression analysis		Robert
Noted tensions between race and racism with data analysis	Elenai Monse	Elenai Jaime Monse Robert

First, in terms of how students' approached the task, one student chose to identify schools that were generalizable or representative of a target population during the pre-interview, and three students took on the same approach during the post-interview. This often included selecting schools that were racially, ethnically, and economically representative of most schools or, conversely, avoiding selecting schools that were majority White or higher income because they may not be representative of other schools. This focus on generalizability and representation may have been guided by the facial recognition example from class where researchers talk about the algorithmic biases associated with poor sampling techniques. This is explained further in the next chapter.

Second, Jaime and Robert both had multiple questions that built on each other in their analysis. This is similar to my experiences analyzing data, where questions emerge and get more specific as I analyze data, similar to a snowball effect. Third, and as mentioned above, more students engaged in a form of exploratory data analysis in the post-interview than the pre-interview. The EDA phase often acted as a bridge between the Plan/Data and Analysis phases that allowed students to begin to get familiar with the data while also specifying their research question and planned analysis. This may have also led to a more efficient Analysis phase, which may help explain why students spent an average of 54% of the coded time in the Analysis phase during the pre-interview but 38% in the post interview (Figure 6.2).

In terms of the analysis itself, all students used a type of visual in the post-interview analysis whereas only two students used visuals in the pre-interview. Often, these visuals were first created in the EDA phase to help students look at the spread of data or find any preliminary relations (e.g., between test scores and race or ethnicity) that led to research questions. Three students also completed their analysis without any formal statistical measures (e.g., correlation). Additionally, Robert engaged in a type of regression analysis in the post-interview that was similar to what we did in class, but did not include any test statistics or p-values, suggesting that a form of data analysis is possible without p-values, confidence intervals, or other test statistics, all of which were not available on CODAP.

Finally, there were some differences with how students considered race, racism, and social justice in the pre- and post-interviews. Specifically, in the pre-interview, all four students used data that showed that there was a negative correlation between a school's demographics (e.g., race or ethnicity, proportion of students that qualify for FRPM) and educational equity or a measure of “quality” (as measured by the standardized assessment data that was given). That is, their analysis may be used to identify a social injustice (or engage in “gap-gazing” or even make deficit claims about students). However, Elenai was the only student that explicitly noted tensions with implying a causal relationship between race or ethnicity and the school’s average tests scores. Specifically, she stated that “saying, oh it's because, it's because of Students of Color is not right because there’s so much more than the student.” One interpretation of Elenai’s statement is that she identified a social injustice and clarified Students of Color are not less capable than their White counterparts but, instead, educational inequities should be situated within the systemic and structural injustices. This is related to the first design feature (DF1: Reflect on structures of social injustices) that was intended to encourage students to focus on structural or systemic factors of social injustices rather than placing the blame or responsibility on individuals or groups of people.

In the post-interview, Elenai, Jaime, and Robert stated that they wanted to avoid implying a correlation or causation between demographic data and test scores that may lead to a deficit rhetoric. As a result, Elenai noted that it again that it was “not right” to blame students, Jaime used data about the structure or systems that may be related to school test scores (funding), and Robert suggested a follow-up study that looks at factors of the schools’ quality of education (and recognizes the difficulties with quantifying that). Monse took a different approach, where she was interested in identifying schools where students were taking and passing AP classes and also had high proportions of students that were identified as BIPOC or qualified for FRPM. In other words, she followed an asset-based approach where she was interested in amplifying some of the possible outcomes in communities (especially Communities of Color) rather than only identifying educational gaps. This may have been influenced by the course activity where we learned about anti-deficit and deficit-based research questions.

In other words, in the pre-interviews, most students focused on identifying social injustices. In the post-interviews, all four of the students also wanted to avoid painting a static picture of educational (in)equities, suggesting that educational equity cannot only be captured by standardized assessments and considering ways to account for the larger social, cultural, political, and historical contexts of educational equity in this statistical investigation. Thus, it is possible that students showed a shift in their critical statistical and data scientific consciousness, where most of the students focused on identifying social injustices in the pre-interview but engaged with more dimensions of praxis in the post-interview that included reflection on the social injustices and provided avenues to address those social injustices in the statistical investigation.

Statistical and Data Scientific Practices

Here, I extend the PPDAC analysis to focus on the different practices that emerged during the pre- and post-interview. I used a similar coding process to that of the PPDAC cycle. A summary of the coding process is shown in Figure 6.6. I began by using a priori codes from the literature review and pilot study, but allowed for other practices to emerge using the constant comparison process. All new codes were initially coded under an *Other* cluster. After the cluster codes were coded for all pre- and post-interviews, I partitioned the cluster codes into smaller codes. These cluster codes were also guided by prior research, but I created new codes and categorized them as needed. I then counted and noted relations between the codes, focusing on identifying codes that emerged in the post-interview but not the pre-interview and identifying any shifts in the distributions of codes.

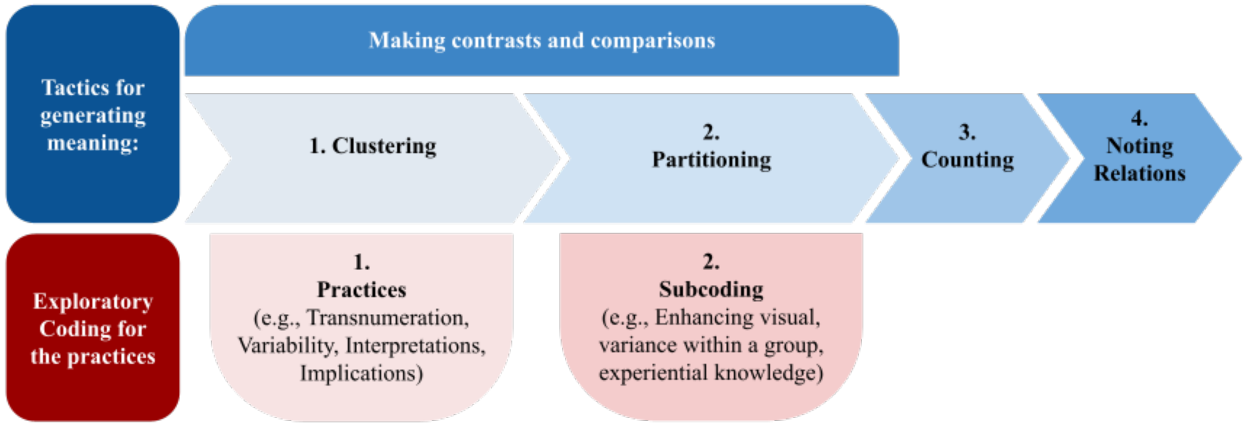


Figure 6.6: Overview of the qualitative coding process for practices in the pre- and post-task-based interviews

Overall Time Spent on Practices

Code definitions and examples for the practices that emerged in the pre- and post-interviews are shown in Table 6.11. The practices are categorized under six clusters: context, transnumeration, variability, data, questions, and conclusions. In this analysis, I focus on the practices that were evident in the post-interview but not the pre-interview. Figure 6.7 shows the percent of coded time for each practice in the pre- and post-interview for all the students combined.

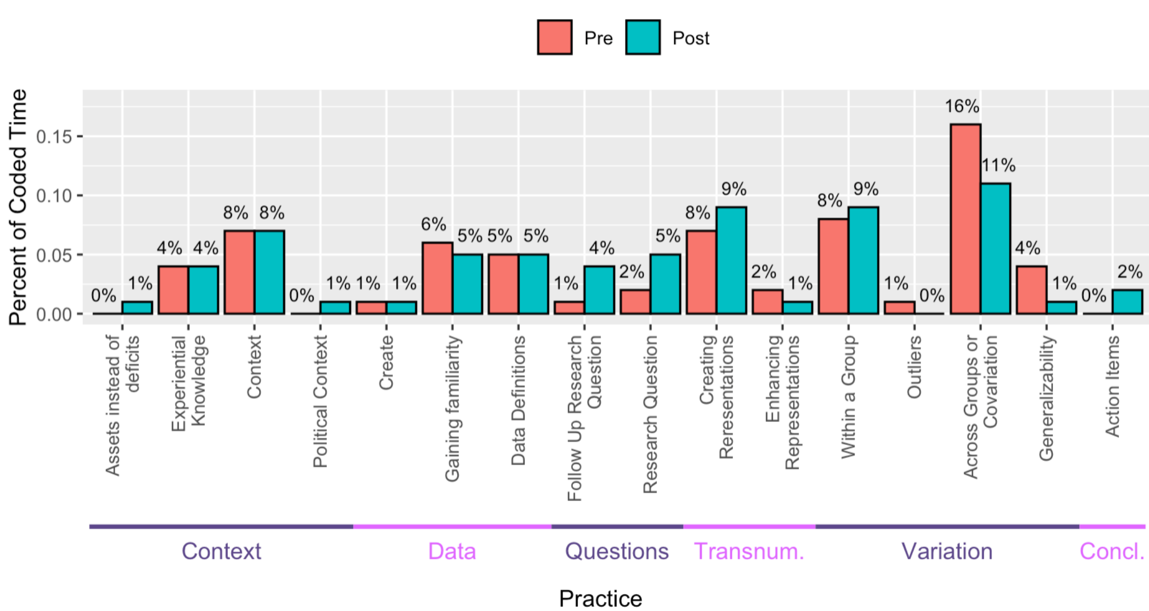


Figure 6.7: Percent of coded time for each practice in the pre- and post-interviews

Table 6.11: Statistical practice that emerged in the pre- and post-interview

Cluster Code	Subcode	Definition	Example	Explanation
Context				
	Assets instead of deficits	Focusing on highlighting the strengths of communities or schools	we don't have to leave the district to go to a White school or a better school. We just have to go to a school that's doing better than us, but similar to us, and see what, what's different with them, why is it working for them" (Elenai)	Student focused on highlighting school assets within a community instead of searching for schools outside of the community
	Experiential knowledge	Drawing on their personal experiences to make sense of the problem context	I'm trying to see where these schools are located. So a couple of them are inner city. But that's not really reflected in the data" (Jaime)	Student drew on their experience about the local context to characterize schools, and noted that what they know based on their experiences is not coded in the given data
	Context	Making connections between the data or analysis to the problem context	that's why I always wanted to look at that that are free and reduced lunch, because I didn't see it, but it's like 11%, so it would now, just be interesting to see, right, because it's like what kind of resources or teachers are they hiring?" (Monse)	Student made a connection between the percent of students that qualify for free or reduced priced meals and its implications on the socioeconomic context (e.g., access to resources, types of teachers)
	Political context	Considering the sociopolitical and/or racialized nature of data (Gillborn et al., 2018; Weiland et al., 2017), focusing on systemic	saying that these schools are bad because they have higher Students of Color is not right because it's about, they don't have adequate access for everyone. It's racist if you blame the students" (Elenai)	Student considered how gap-gazing my reinforce system narratives and states that it is racist to blame students, shifting the blame away from the students

Table 6.11: Statistical practice that emerged in the pre- and post-interview, Continued

Cluster Code	Subcode	Definition	Example	Explanation
Variation				
	Generalizability	Considering the generalizability of data (Franklin et al., 2007; Lee & Tran, 2015), including how representative a sample is of a target population	" And what schools are doing right. Not just go to where it's more White or they have higher test scores because whatever. But you want to keep, I would want to keep it in the same ballpark" (Elenai)	Student wanted to select a school that was representative of a target population, possibly considering some of the implications of generalizing from a school that's majority White to a school that's majority Students of Color
	Outliers	Considering or noting possible outliers in the data and how they might influence the	" There's like a few outliers, but for the most part, it is lower" (Elenai)	Student noted that there were some points on a scatterplot that did not follow the overall trend, and therefore considered them outliers
Data				
	Gaining Familiarity	Students spend time looking at the given data	Spending independent time to explore the given data	Students would explain that they were spending time to look through the data before beginning any type of analysis
	Creating Data	Creating new data, typically using existing data	Using the number of students that enrolled in AP courses and the schools total enrollment to create a new variable for the proportion of students that enrolled in AP courses at the school (Monse)	Used given data to create a data that was easier to compare. In particular, it may be more appropriate to compare proportions instead of counts
	Definitions of Data	Asking for definitions of data	Students asked questions about what the data measured or the acronym definition	Students clarified quantities or symbols

Table 6.11: Statistical practice that emerged in the pre- and post-interview, Continued

Cluster Code	Subcode	Definition	Example	Explanation
Questions				
	Research Questions	Asking statistical and/or research questions that are analyzed in the interview	Explicitly stating a research question, which was often also a statistical question	Students explicitly stated the research question they were going to address in their analysis
	Follow-Up Research Questions	Asking new research questions that emerged from the analysis phase	" I guess to me would be interested to see like how their math classes look. Because obviously students are confident enough to take those ap math classes, so I would look at the one that has….I would be curious to observe, like, how their classes, how the math classes are and why students are confident" (Monse)	Student presented avenues for future research. In this case, the student wanted to observe a classroom, which may entail a qualitative study
	Action Items	" the curriculum can help with equity too" (Elenai)	Students provide avenues to advance social and/or racial justice	Student suggested that one way to achieve educational equity may be by reconsidering which curriculum is used

It is worth noting that there were moments in the video that were not coded with a practice code and there were moments that were coded with multiple practice codes, so the sum of the codes across the pre- or post-interviews will not add up to 100%. However, Figure 6.7 is helpful for identifying which practices emerged in the post-interview and if there were any significant shifts in the types of practices across the pre- and post-interview. For example, about 1% of the coded time in the pre-interview went towards students asking follow-up research questions, whereas 4% of the coded time in the post-interview went towards asking follow-up research questions. Similarly, 0% of the coded time in the pre-interview was coded as Action Items whereas 2% of the post-interview was coded as Action Items.

There were three practices that were evident in the post-interview but not the pre-interview: *Political Context*, *Assets Instead of Deficits*, and *Action Items*. All three practices were related to different elements of praxis. The Political Context and Assets Instead of Deficits codes were related to the reflection part of praxis. Particularly, the Political Context code was assigned to moments where students were shifting the conversation away from placing responsibility or blame on individuals or groups of people towards situating the social injustice within the larger social, political, cultural, and historical contexts (similar to DF1: Reflect on structures of social injustices). The Assets Instead of Deficits was assigned to moments where students showed evidence of an asset-based approach to analyzing data, such as aiming to highlight some of the strengths of communities instead of gap-gazing. The Action Items code is related to the action part of praxis. This code was applied to moments where students provided follow-up avenues to achieve social justice. Below, I provide examples of what these practices were like in the post-interviews.

Political Context

The political context code was assigned to instances where students considered the social, political, cultural, and historical contexts in the data (Gillborn et al., 2018; Weiland et al., 2017), focusing on systemic structures of social inequities. All students showed evidence with this practice during the post-interview, especially when interpreting negative correlations between school demographics (e.g., BIPOC, FRPM) with test scores. Particularly, students often noted possible deficit interpretations that may imply a causal relationship between the proportion of students that were identified as BIPOC or qualified for FRPM with the test scores, where a deficit interpretation may imply that higher proportions of students identified as BIPOC or qualify for FRPM may lead to lower test scores. However, students also noted that there are other social, political, cultural, or historical contexts that may be associated with the discrepancies (e.g., quality of education, finances, tutoring).

For example, at the end of the post-interview, Elenai stated that saying "oh, it's ((lower test scores)) because of Students of Color is not right because there's so much more…it's so much more than the student." Then, she added that

> saying that these schools are bad because they have higher Students of Color **is not right because it's about, they don't have adequate access for everyone**. **It's racist if you blame the students.** So, to me, I could, I could say that that's the conclusion because you're making recommendations based off, off what you think you're saying, but it's not really that. As a mother, I know that **it's also about the quality of education, redlining, … and all the other messed up stuff**…

Here, Elenai recognized that analysis may lead to racist implications (e.g., putting the "blame" on students for not performing well on standardized assessments). Instead, she noted that there are other factors (e.g., quality of education, redlining) that may lead to differentiated learning experiences.

Robert took a similar approach that aimed to shift the conversation away from a possible deficit interpretation to analyze the larger structural factors behind social injustices. For example, when analyzing data about the proportion of students that were classified as Hispanic or Latinx and test scores, Robert stated that

> What would give a reason, why this is true? **I don't think it's because they're Hispanic and so they're bad at math**, right. Like you can't say that, **it sounds bad.** But it could have to do with, like, the school, so I'm assuming…((pauses to look at)) **So it probably just has to do with the quality of the school**, but **I don't know if I have enough information to really say that right now**. But like **definitely not the students, not because they're Hispanic.**

Here, Robert noted that explicitly shifting the interpretation away from blaming Hispanic or Latino students for low test scores to talking about other factors that may influence test scores.

Monse and Jaime noted similar tensions. They named possible deficit interpretations, but then shifted the conversations towards analyzing the larger structures of social injustices. In this sense, students were engaging in *praxis*, where the statistical analysis served as a way for students to begin to identify social injustices using data but then also as help start conversations about the larger social structures at play and, eventually, some possible ways to achieve social justice.

Assets Instead of Deficits

The *Assets Instead of Deficits* practice code was used to describe instances where students focus on highlighting strengths of a community or school. Monse, Robert, and Elenai all engaged with this practice during the post-interview. For example, at the end of the post-interview, Elenai was summarizing her overall process:

> say we're looking at ((school district)), **we don't have to leave the district to go to a White school or a better school. We just have to go to a school that's doing better than us, but similar to us, and see what, what's different with them, why is it working for them**, but it's not working with us. Yeah, and that's the whole thing if we're, if we're, if our discussion is about making things more

> equitable because of the numbers, then you want to look at, where are they showing that it is equitable? **And what schools are doing right**. **Not just go to where it's more White or they have higher test scores because whatever**. But you want to keep, I would want to **keep it in the same ballpark**. And just going to White schools or rural or whatever, it's not the same. Yeah because ((school name)) has the lowest percentage of Students of Color and they have the higher test scores, but there's other things like money too.

This comment was assigned the *Generalizability*, *Political Context*, *Assets Instead of Deficits* code. Specifically, Elenai was interested in identifying schools that were similar to a school that she was hypothetically at to allow for a fair comparison (*Generalizability*). Elenai was also considering the political context. For example, one interpretation of Elenai's comment was that she is being intentional about which schools to visit and the messages that she may be sending by positioning as some "better" schools. Further, she noted previously in the interview that the schools that had higher test scores had a lower proportion of students identified as BIPOC, but that those discrepancies may be attributed to other factors like income or additional resources (*Political Context*). Thus, she suggests that it may be beneficial to not "just go where it's more White or they have higher test scores" and instead recommended identifying schools that are "doing better than us" (e.g., as measured by standardized assessments since that's the data that was provided) but also a school that is "similar to us" (the hypothetical school that she is working at) to show what "schools are doing right" (*Assets Instead of Deficits*). That is, it is possible that Elenai wanted to highlight some of the strengths of schools that are majority BIPOC rather than engaging in a color-blind analysis that may inherently (whether implicit or explicit) position schools that are majority White as "better schools" than schools that are not majority White.

Monse was also interested in identifying schools that were majority BIPOC and had a relatively high proportion of students that enrolled in AP courses. Table 6.12 shows the interaction when Monse was describing her overall process to identify three schools.

Table 6.12: Monse describing why she wanted to identify three schools that were majority BIPOC

Line	Speaker	Utterance
1	Monse	Let's look at why your kids are like this, right? So I will actually want to see one of those ((schools or classrooms)), but it would be more of a compare and contrast, right? Like what are you doing that's different from, like let's say this one ((school A)), right? **They both have, have high PoC, but why does this one have more AP students. What can we learn from them?**
2	Kevin	Why does it matter that it ((school A)) has high PoC?
	Monse	Because, like, **we can learn from *la gente* [the people]**. Like, tenemos cosas [we have things] and people that know their ((stuff)). **Everything doesn't have to be about White people.**

Similar to Elenai, Monse had previously mentioned that she did not want to select schools that were majority White because they may not be generalizable to all schools and because the analysis may send implicit deficit messages that position schools that are majority White as better than schools that are majority BIPOC. In fact, she may be trying to decenter Whiteness ("Everything doesn't have to be about White people") by learning from and amplifying the strengths of *la gente* instead. It is worth noting that this quote may also be telling of the shared ethnic and linguistic identity that I had with Monse, especially as people who identify with the larger Latinx or Chicanx community. This may be evident by the use of Spanish and English as well as the term "la gente" which may refer to Communities of Colors.

As a third example, Robert included similar analyses in his post-interview about demographic data (e.g., BIPOC, FRPM) and student test scores. After expressing some frustration about how the analysis "sounds bad" and that some of the discrepancies may be attributed to larger structural issues with the quality of education instead of racial or ethnic groups, Robert was interested in seeing if there was “something good” growth in the standardized assessments. He specifically related the sixth-grade mathematics scores to the eighth-grade mathematics scores. When interpreting the graph, he noted:

> They're still **learning consistently**. Which is what you would want to see in this, right? **So the scores might be lower, but they’re still learning**. They might even be learning more…**it’s not fair to say that they’re not, they’re not learning if they have lower scores, they’re still learning consistently, or maybe even more**. I don’t know if it’s more, but they’re definitely learning at least as much.

One interpretation of this analysis is that Robert was interested in findings that showed some type of asset rather than only focusing on analysis that may perpetuate negative assumptions, as noted by his frustration with the prior analysis that “sounds bad.” Robert may be specifically challenging an assumption that students in schools with lower test scores are not learning. In fact, he states that “it’s not fair to say that they’re not, they’re not learning if they have lower scores.” Rather, he provided evidence that suggests that the students are “still learning consistently, or maybe even more.” In fact, if Robert was ultimately looking for an analysis that would not perpetuate negative assumptions, it is possible that Robert’s entire approach was anti-deficit. However, a follow-up interview would have been needed to confirm this conjecture.

Action Items

Finally, the *Action Items* code was assigned to moments where students provided avenues for future work that would achieve social justice. It is worth noting that the goal of the task (identifying three schools) was not necessarily about social justice. However, since all students considered aspects of educational equity during their post-interview (e.g., relating race or ethnicity to test scores), it was common for students to also mention ways to achieve educational justice. Specifically, Monse, Elenai, and Jaime all brought ways to address the educational injustice that they noticed.

For example, after the interaction in Elenai talked about how it is "so much more than the student," Elenai continued to talk about her experiences observing classrooms and with her children is that "the curriculum can help with equity too." Specifically, she noted that one possible way to help achieve equity is to look at schools that have high test scores and possibly adopt those across the district. She did note that she does not know if "that will solve all the problems, it probably won't, but it's a start." Nonetheless, she was looking for avenues to help achieve educational equity across the school district by possibly identifying existing resources.

Since Monse was interested in visiting schools that high a relatively high proportion of students that were identified as BIPOC and were enrolled in AP classes, she stated that she was interested in learning more about the different "professional development support, or like what supports teachers need to help all students take AP classes, well, like, if they want. I feel like that would be one way to achieve equity here." Later, Monse also noted that schools can have "AVID tutors to help tutor students in math or science." That is, she was considering how schools could support teachers to help all students enroll in AP classes and provide tutoring as a pathway to achieving social justice.

Finally, at the end of the interview, Jaime stated that taking a deeper look at how funding and resources are allocated may help improve educational equity across the school district. Specifically, he said

> Money talks. Maybe we can give schools, or like I don't know how the taxes work, but like is there a way to **give more money to schools that need more help**? Like what if everyone at this school had a TA? Or like a translator?

It is possible that Jaime was reflecting on an equity versus equality idea, where equality implies that everyone gets the same funding and resources per student but equity is about sending funding and resources where they are most needed. Jaime ended this by stating that he is not familiar with how schools are funded, but suggesting that there may be some legislative opportunities to help all students.

Summary of the Statistical and Data Scientific Practices

I end by summarizing the practices across four themes: (a) goals of the task, (b), the role of experiential knowledge, (c) tensions between race, racism, and the statistical investigation cycle, and (d) action. First, in terms of the goal of the task, most of the students focused on identifying social injustices during the pre-interview (e.g., noting that schools with high proportions of Students of Color had lower average scores on standardized assessments). In the post-interviews, students appeared to focus on or note the structures that shaped or lead to social injustices (DF1: Reflect on structures of social injustices). In doing so, they may have avoided painting a static picture of educational equity (e.g., suggesting that educational equity cannot only be measured by standardized assessments) and sometimes named the larger social, cultural, political, and historical contents of educational equity in the task.

There were also some shifts in the role of experiential knowledge. In particular, during the pre-interview, experiential knowledge mostly came up when students were talking about which data they would select for the analysis, often selecting data that is traditionally associated with equity (e.g., race or ethnicity, free or reduced priced meals). For example, Jaime stated that "[w]e care about diversity, so I want to use Students of Color." Students also drew on their experiential knowledge to select data in the post-interview, but also drew on their experiential knowledge about the context to situate the data within a larger sociopolitical context (e.g., drawing on their own experiences with education and what entails educational equity). This may be interpreted as students drawing on their funds of knowledge (González et al. 2005; Moll et al., 2005), either from their own experiences with the context or what they learned about in other classes (e.g., students who referenced learning about Critical Race Theory in another class). For example, after the task portion of the interview, Monse stated that "[f]rom observing classrooms and as a student, I know that there's so much more to a student than a test. What about how empowered they are? What about how confident they are?"

I also noticed that there were some differences in how students addressed some of the tensions in the data, especially in regards to data about education achievement and equity and how that may lead to gap-gazing. Admittedly, the interview was designed to bring up those tensions. In the pre-interview, all students looked at data on race or ethnicity and scores on standardized assessments. When noting that there was a negative relationship between both data, Elenai noted that it felt "weird" to imply that Students of Color have low test scores and seemed to differentiate between what the data implied and what she believed. Similarly, Robert noted at the end of the interview that "I guess it's right, it's what the data says." However, in the post-interview, students named those tensions, often stating that an argument may be "deficit" or "not asset-based" and explaining why. In fact, it is possible that students were looking to engage with data an anti-racist way (action taken specifically to combat racism instead of generally combat deficit perspectives), but more evidence would be needed to support that claim.

Finally, there were some differences in the role of action while engaging with the task. In particular, students did not provide avenues for social change in the pre-interview. In addition to students focusing on identifying social justices and not accounting or naming the structures of injustice at play, this may be interpreted as students having limited engagement with the reflection and action components of praxis. In fact, after the task ended, Elenai noted that she ended the task by stating "not knowing what to do next. Like I care ((about educational equity)), but like now what?" One interpretation is that there was not a sense of agency. However, in the post-interview, students used the data as a launchpad for future analysis, some of which would require qualitative work. Thus, it is possible that students showed a shift in their critical statistical and data scientific consciousness, where most of the students focused on identifying social injustices in the pre-interview but engaged with more dimensions of praxis in the post-interview that included reflecting on the social injustices, challenging dominant ideologies, and providing avenues to address those social injustices in the statistical investigation

Chapter 7: RQ4 (Focusing Phenomenon)

Research Question 4 adds evidence of how students developed understanding of race and racism in the context of data science. In particular, the research question was:

Research Question 4: Focusing Phenomenon
a. How do elements of the TE contribute to the students' understanding of race and racism in the context of statistics and data science?

This section used elements of the focusing phenomena framework (Lobato et al., 2003, 2013) to coordinate how aspects of the classroom environment (e.g., design features, tasks, tools, and the teacher) directed students' attention towards understandings of race and racism in the context of data science.

Overview of Analysis

There were four phases to the analysis: (a) identifying Centers of Focus in student responses to four prompts about data neutrality, (b) identifying focusing interactions that may have directed students' attention to statistical, data scientific, or pedagogical features of the course, (c) describing the features of the task, including affordances and constraints of the task, and (d) describing the ways in which classroom participation is organized and regulated by classroom norms. I also add a fifth analytical pass that makes connections with the design features presented in Chapter 4.

Analytical Pass 1: Identifying Centers of Focus

First, I began by looking at student responses to prompts about race neutrality in statistics and data science (shown in Table 7.1), noting properties of features of race neutrality that students mentioned in their responses (Centers of Focus, CoFs). As mentioned in the methods section, CoFs were "the properties, features, regularities, or conceptual objects that students notice" (Lobato et al., 2013, p. 814). I considered CoFs that at least one student included in their responses to the four prompts in Table 7.1. However, it is worth noting that it is possible that there were CoFs that were not included in the responses and that one students' CoF may have not been a CoF for their group, breakout room, or class. Rather, the CoFs identified in the four prompts were used to guide the analysis of the following analysis of classroom data.

Table 7.1: Questions about data neutrality used to identify the centers of focus in the TE data with short descriptions of the class when the question was assigned

Class	Class 1: Pre-Survey	Class 3: Homework	Class 6: Homework	Class 12: Post-Survey
Question	**1**. A student says, "numbers speak for themselves." Do you agree or disagree? Please explain.	**2**. A friend tells you that "data can't be racist, numbers don't lie." Do you agree or disagree with this statement? Explain	**3**. One of your colleagues states that "technology is politically neutral, therefore data is politically neutral." Do you agree or disagree? Explain	**4**. A student says ,"numbers speak for themselves." Do you agree or disagree? Please explain.

All four prompts were assigned at specific points in the course. Particularly, although race and racism were discussed throughout the entire course, the four prompts from Table 7.1 were asked in preparation for lessons where we were designed to summarize the role of race and racism in the PPDAC cycle. This was designed intentionally so that students would be prepared for the class discussion and so that I could analyze their individual responses to data neutrality as well as the class discussion that followed about race and racism in the PPDAC cycle. Furthermore, the first question was assigned after Class 1 when students were introduced to the class and the goal of the class, but before any class activities about race and racism in the context of statistics and data science. The second question was assigned after students had read about Critical Race Theory and talked about it as a class and was assigned alongside a reading about Quantitative Critical Race Theory (QuantCrit). At this point, we had also discussed the role of the problem context in statistics and data science (the Problem phase of the PPDAC). The third question was assigned for the Class 6 homework. By this point, students had learned about sampling, randomization, study designs (the Planning and Data phases of the PPDAC cycle), and also watched a video on algorithmic bias as part of their homework. The last question was assigned during Class 12 after students had learned about different regression models (the Analysis phase of the PPDAC cycle). This last question was not debriefed as a class. The Conclusion phase of the PPDAC cycle was discussed throughout the class (e.g., role of researcher positionality when writing reports, impact of sampling on generalizability, how to interpret statistical models).

Furthermore, I chose to look at the responses to the four prompts about data neutrality instead of the task-based interviews for two reasons. First, I only had access to interviews from four students whereas I had more data about the individual student responses to class assignments. This allowed me to identify potential understandings of the role of race and racism in the PPDAC cycle that were not evident in the interview, then coordinate those with Centers of Focus and classroom data. Second, the students' responses to the prompts may have been more revealing about student learning than the interviews and noted specific examples from the class lessons and activities. For example, there were three practices that emerged in the post-interview but not the pre-interview (political context, assets instead of deficits, and action items). However, there were over 15 different codes for students' views of data neutrality that arose from these prompts, and most students identified specific activities (e.g., videos, readings, discussions) related to those views.

Analytical Pass 2: Identifying Focusing Interactions

Second, I looked for instances in the class interactions where the CoFs were highlighted (e.g., labeling, marking, annotating, and displaying student work to the entire class; Lobato et al., 2013). I related each highlighting instance to Mathematical Language Routines (DF3:

Communicate). It is worth noting that each instance of highlighting was not part of an MLR but, rather, I looked for highlighting in the MLR episodes. The goal of this part of the analysis was to identify discourse practices that may have directed students' attention to a particular CoF (focusing interactions).

As mentioned in the methods section, I chose to focus on highlighting student contributions because there were few instances of quantitative dialogue or renaming in the field notes. Part of why there were few instances of quantitative dialogue may be because mathematics is different from statistics and data science. In future work, another pass of this analysis could define and look for statistical and data scientific dialogue (e.g., dialogue related to variation, context, data wrangling or other attributes that are statistical and not necessarily mathematical).

Analytical Pass 3 and 4: Describing the Features of the Task and Nature of the Activity

The third analytical pass entailed identifying the features of the mathematical task related to a particular CoF. This included providing a description of the task as well as noting any possible affordances, constraints, and modifications for future and in-person iterations of this study. Finally, the fourth analytical pass required describing the classroom norms that may have influenced participation dynamics and, consequently, what students notice in the activity. This included describing general classroom norms and how those norms may be related to the students' roles, the teacher's role, and the students' noticing in the activity.

Results of CoF Analysis

Analytical Pass 1: Identifying Centers of Focus

The goal of the first analytical pass was to identify the CoFs across the four homework questions about data neutrality. This helped identify CoFs that emerged across the course (i.e., CoFs that appeared in the Class 3 homework question but not the Class 1 homework question). Similar to the interviews from Research Question 3, I used elements of Miles et al.'s (2020) tactics for generating meaning in qualitative data and their implications for coding to identify the Centers of Focus. An overview is shown in Figure 7.1.

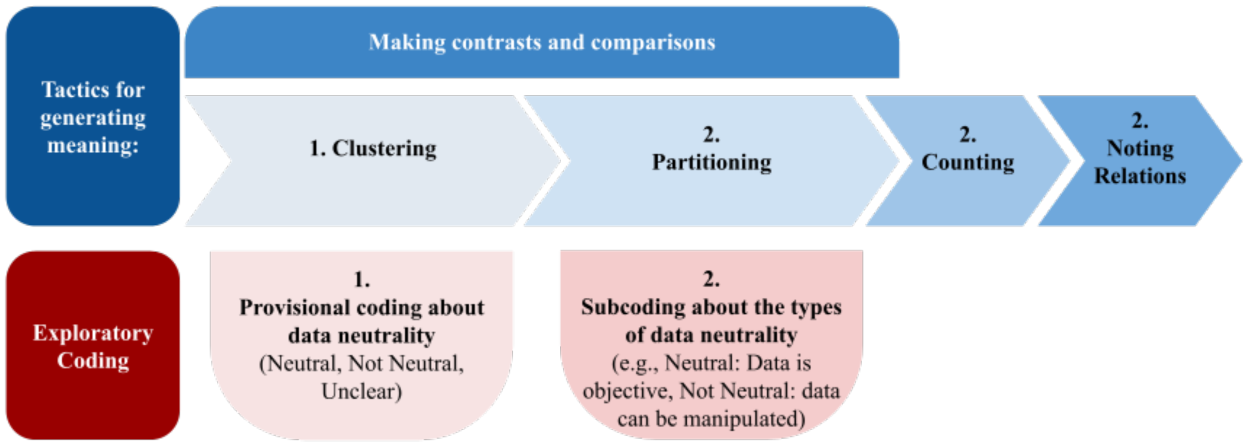


Figure 7.1: Tactics for generating meaning from qualitative data and exploratory coding analysis used for Research Question 4 (Miles et al., 2020)

Coding for Centers of Focus

The process began by clustering data for each student across each of the four prompts from Table 2.1. In this research question, this was implemented using *provisional coding* (Miles et al., 2020) to create cluster codes that described whether students' responses suggested that data was neutral, not neutral, or both. These three cluster codes were a priori codes that were informed by the pilot study responses to the pre- and post-survey questions about data neutrality. Although I was open to including new cluster codes, the a priori codes were enough to cluster all the student responses. Furthermore, The *Neutral* cluster code included responses that data was removed from any social, cultural, political, or historical influences. That is, "numbers speak for themselves" regardless of the surrounding context. Conversely, the *Not Neutral* cluster code included responses that referred to the social, cultural, political, or historical contexts involved in the PPDAC process. Finally, the *Both* cluster code included responses that suggest that data are numerical or technological objects that are inherently neutral, but there may be social, cultural, political, or historical contexts that influence how data is collected or used.

Next, I partitioned the clusters. Miles et al. (2020) described partitioning as the process of subdividing data to avoid data blurring. The goal is thus to identify smaller clusters of data that are all related under a broad theme but have differentiated features. This entailed *subcoding* the clusters to include details on students' view of data neutrality. Furthermore, student responses may include more than one characteristic of how data is neutral or not neutral. Thus, student responses may have more than one subcode attached. After all the subcodes were coded, I also noted which phase of the PPDAC cycle the subcode most addressed.

Central to the clustering and partitioning process is making contrasts and comparisons. To enact this process, I drew on the *constant comparison* process of grounded theory (Corbin & Strauss, 1990; Strauss & Corbin, 1994). The constant comparison is an iterative and reflective process where we create conjectures about the data, compare them to other parts of the data, and modify categories as needed (Corbin & Strauss, 1990; Strauss & Corbin, 1994). This helps further identify categories and codes, their characteristics, and relationships with each other. The constant comparison process may also help bring awareness to some of the emerging themes and guard against biases that may emerge from being restricted by fixed or given categories.

The final two phases include counting data and noting relations. For this research question, this included counting the clusters and subcodes within each cluster and noting any relations with the subcodes. Of particular interest was identifying when new subcodes appeared in the class and how often they appeared.

Example of Coding the Center of Focus. As an example, consider the response from Elenai shown below. The question was "A friend tells you that 'data can't be racist, numbers don't lie.' Do you agree or disagree with this statement? Explain." Bold emphasis is added to highlight different characteristics of data neutrality and the text inside the brackets show the codes assigned to the respective bolded text.

> **Data can be racist when it sets out to show disproportion through deficit framework [Researcher Bias]**. When studies look at failure rates instead of achievement rates it can be very racist. You can **use data and spin it in such a way that it becomes racist [Manipulate Conclusions]**. Depending on how you **collect your data, how many people are surveyed, types of people surveyed [Sampling]** can all make a difference in what backs or creates your "evidence". If the methods of collection are not biased the **results will not be reflective of the target population creating racist data [Sampling]**.

First, the cluster code assigned to this response was *Not Neutral* because the student states that "Data can be racist." This is not to say that the student did not think that data can be both neutral and not neutral, but rather they provide evidence for why data is not neutral and, in this case, can lead to racist outcomes. Then, there were three subcodes assigned to this student response: (a) *Researcher Bias*, (b) *Manipulate Conclusions* (Manp. Concl.), and (c) *Sampling*.

The *Researcher Bias* subcode was used to describe statements where students suggest that the statistician or data scientists may have expectations or preconceived beliefs that influence their engagement with the data at any point during the statistical investigation cycle and lead to potential biases. The students may also state whether the bias is implicit or explicit. In Elenai's case, they are specifically referring to statisticians or data scientists drawing from deficit frameworks that may support racist narratives (e.g., gap-gazing research that places responsibilities for differences in measures of achievement on Students of Color rather than accounting for the larger structural social injustices that cause differentiated learning experiences).

The *Manipulate Conclusions* subcode was assigned to statements where students suggest that data or numbers can be manipulated, usually to advance the researchers' personal agenda and typically during the conclusion phase of the statistical investigation cycle. This is different from the *Researcher Bias* code because the statistician or researcher may be specifically making an effort to manipulate data and is specific to the conclusion phase of the PPDAC cycle rather than an overall framing. In Elenai's case, they are referring to how statisticians and data scientists may "use data or spin it in such a way that it becomes racist."

Finally, the *Sampling* subcode was used to describe statements where students suggest that sampling may lead to potential biases (e.g., over- or underrepresentation), often related to the generalizability of the research. This is related to the planning and data phases of the PPDAC cycle. In Elenai's case, she refers to biases in data collecting and sampling that may influence the generalizability of the study.

Centers of Focus in the four prompts About Data Neutrality

The CoFs identified in the four prompts about data neutrality shown in Figure 7.2. The evidence for all codes were student responses to the four prompts about data neutrality. The counts for the clusters (neutral, both, not neutral) represent the total number of students that provided responses that were assigned to that particular cluster. These counts add up to the total number of students that submitted the assignment. The counts for the subcodes were the total number of instances that were assigned that description. Since student responses could be assigned multiple subcodes, the number of students in a cluster may not be the same as the number of instances of that subcode for each cluster. Furthermore, subcodes highlighted in yellow emerged first in the respective homework assignment. For example, the *Context - Historical Political* (Context - Hist. Polit.) subcode appeared first during the Class 3 homework under the *Both* and *Not Neutral* clusters. The subcodes highlighted in red had already appeared in the homework assignments before, but were under a different cluster. For example, the

Class 1: Pre-Survey		Class 3: Homework		Class 6: Homework		Class 12: Post-Survey		Phase in the PPDAC Cycle
Cluster (# students)	Subcode (# instances)	Cluster (# students)	Subcode (# instances)	Cluster (# students)	Subcode (# instances)	Cluster (# students)	Subcode (# instances)	
Neutral (6)	Objective (6)	Neutral (1)	Objective (1)	Neutral (2)		Neutral (2)	Objective (2)	Conclusion
					Well-intended (2)			PPDAC
Both (1)		Both (6)	Researcher Bias (1)	Both (2)	Researcher Bias (2)	Both (0)		PPDAC
			Context - Hist. Polit. (2)					PPDAC
			Sampling (1)					Plan-Data
	Outside Factors (1)							PPDAC
	Data are neutral (2)		Data are neutral (5)					PPDAC
			Interpreters (3)					Conclusion
			Manipulate Concl. (3)					Conclusion
			Objective (1)		Objective (2)			Conclusion
Not Neutral (4)	Researcher Bias (2)	Not Neutral (5)	Researcher Bias (1)	Not Neutral (8)	Researcher Bias (2)	Not Neutral (12)	Researcher Bias (5)	PPDAC
							Q Framing (2)	PPDAC
			Context (1)				Context (1)	PPDAC
			Context - Hist. Polit. (2)				Context - Hist. Polit. (3)	
								PPDAC
	Survey Qs (1)							Plan-Data
	Sampling (3)		Sampling (4)		Sampling (1)		Sampling (1)	Plan-Data
	Outside Factors (1)						Outside Factors (1)	PPDAC
	Manip. Concl. (3)		Manip. Concl.(3)		Manip. Concl. (2)		Manip. Concl. (1)	Conclusion
					Algorithmic Bias (6)		Algorithmic Bias (2)	PPDAC
					Agency (2)			Conclusion
	Interpreters (1)							
							General Bias (1)	PPDAC
Other (3)	Other (3)	Other (1)	Other (1)					

Figure 7.2: Centers of Focus for the four prompts about data neutrality

Key: Yellow highlights are subcodes that emerged first in the respective assignment. Red highlights are subcodes that may have appeared in other clusters (neutral, not neutral, both) in previous assignments, but appeared first in a cluster for the respective assignment.

Researcher Bias appeared under the *Not Neutral* cluster in the Class 1 homework, but then also appeared under the *Both* category in the Class 3 homework. The phase that each subcode most corresponds to is shown on the right.

There were eight codes that emerged after the Class 1 homework: (a) Context - Historical and Political, (b) Interpreters, (c) Well-Intended, (d) Encoded Bias, (e) Agency, (f) Question Framing, (g) Context, and (h) General Bias. Below, I describe the eight codes and provide examples of student responses from the students that gave consent to use their data for research. A full description of the Centers of Focus are shown in Table 7.2 to Table 7.5, including a description, the phase of the PPDAC cycle, and examples.

The *Context - Historical and Political* subcode emerged after the Class 3 homework. The Class 3 homework was “A friend tells you that ‘data can't be racist, numbers don't lie.’ Do you agree or disagree with this statement? Explain.” This subcode was used to describe statements where students suggest that some of the social, cultural, historical, or political contexts may help understand or create biases. This subcode may be applied to any phase of the PPDAC cycle. There was only one student who gave consent, Jacky, who was assigned this subcode. Particularly, Jacky stated:

> At first, I did agree because numbers are objective and are not sueded by opinions. But I failed to look at the bigger picture, **the historical and political context that may be surrounding the data**. So in a way, I would agree that **data cannot be racist but the data interpreters can be for neglecting to look at the bigger picture.**

This statement was coded under the *Both* cluster. One interpretation of Jacky’s statement is that “interpreters” (the audience, statistician, or data scientists) may implicitly or explicitly interpret

Table 7.2: Centers of Focus on the first question about data neutrality (Class 1): “A student says, ‘numbers speak for themselves.’ Do you agree or disagree? Please explain”

Cluster Code	Subcode	PPDAC Phase(s)	Description	Examples
Neutral	Objective	Conclusion	Students suggest that data are factual, definitive, or another form of objective truth or evidence (e.g., to support a claim). Related to the conclusion phase of the statistical investigation cycle.	Yes because numbers are factual and can be proven (Ellie) Numbers are objective, they are definitive. (Jacky) Yes, numbers such as dates and amounts are evidence used frequently in courts and daily life. (Margarita)
Both	Data are neutral	PPDAC	Students generally suggest that data are neutral and not a social construct. This is not specific to a particular phase of the statistical investigation cycle. This is similar to students suggesting that data are objective, but they do not explicitly make claims about how objective data are.	Numbers speak for themselves could pertain to data and evidence and I would agree… (Elenai)
	Outside Factors	PPDAC	Students suggest that there may be outside factors (e.g., latent variables, sampling, survey bias, outliers) that may cause biases, but do not state specific examples of outside factors. These outside factors may influence any phase of the statistical investigation cycle	…But what you do not see is what limiting factors may be hindering the data. (Elenai)
Not Neutral	Researcher Bias	PPDAC	Students suggest that the statistician or data scientists may have expectations or preconceived beliefs that influence their engagement with the data at any point during the statistical investigation cycle and lead to potential biases. May occur on purpose or not on purpose.	…Not only that, but when conducting statistical surveys, there are biases that researchers have that will reveal themselves in the numbers… (Robert)

Table 7.2: Centers of Focus on the first question about data neutrality (Class 1): "A student says, 'numbers speak for themselves.' Do you agree or disagree? Please explain", Continued

Cluster Code	Subcode	PPDAC Phase(s)	Description	Examples
Not Neutral	Survey Questions	Plan / Data	Students suggest that survey questions may lead to potential biases (e.g., leading questions, loaded questions, two-in-one questions). Related to the planning and data phases of the statistical investigation cycle.	and if they ask leading questions… (Robert)
	Sampling	Plan / Data Conclusion	Students suggest that sampling may lead to potential biases (e.g., over- or underrepresentation), often related to the generalizability of the research. Related to the planning and data phases of the statistical investigation cycle.	Numbers mean nothing without knowing the source [e.g., sample] which they came and who or what it was sampled from (Caden) …This happens when samples are taken from specific communities… (Robert)
	Outside Factors	PPDAC	Students suggest that there may be outside factors (e.g., latent variables, sampling, survey bias, outliers) that may cause biases, but do not state specific examples of outside factors. These outside factors may influence any phase of the statistical investigation cycle	…when researchers ignore other possible causes… (Robert)
	Manipulated Conclusion	Conclusion	Students suggest that data or numbers can be manipulated, usually to advance the researchers' personal agenda and typically during the conclusion phase of the statistical investigation cycle.	Numbers can be easily twisted for your own purpose... (Robert) I would have to disagree with this student as numbers can be manipulated for whatever purpose the person needs… (Jaime)

Note: information in brackets is added for interpretation

Table 7.3: Centers of Focus on the second question about data neutrality (Class 3): “A friend tells you that ‘data can't be racist, numbers don't lie.’ Do you agree or disagree with this statement? Explain”

Cluster Code	Subcode	PPDAC Phase(s)	Description	Examples
Neutral	Objective	Conclusion	Statements suggest that data are factual, definitive, or another form of objective truth or evidence (e.g., to support a claim). Related to the conclusion phase of the statistical investigation cycle.	I don't believe that data can be racist because data is just a gathering of straight facts about all kinds of information, so for it to be racist is kind of absurd. (Ellie)
Both	Researcher Bias	PPDAC	Statements suggest that the statistician or data scientists may have expectations or preconceived beliefs that influence their engagement with the data at any point during the statistical investigation cycle and lead to potential biases. May occur on purpose or not on purpose.	…However, researchers definitely can have bias when collecting/analyzing data. (Robert)
	Context - Historical Political	PPDAC	Statements suggests that some of the social, cultural, historical, or political contexts may help understand or create biases at any point of the statistical investigation cycle	At first, I did agree because numbers are objective and are not sueded by opinions. But I failed to look at the bigger picture, the historical and political context that may be surrounding the data… (Jacky)
	Sampling	Plan / Data Conclusion	Statements suggest that sampling may lead to potential biases (e.g., over- or underrepresentation), often related to the generalizability of the research. Related to the planning and data phases of the statistical investigation cycle.	…Data could also be limited to certain subjects which may not cover the whole topic and/por research topic. (Margarita)
	Data are neutral	PPDAC	Statements generally suggest that data are neutral and not a social construct. This is not specific to a particular phase of the statistical investigation cycle. This is similar to students suggesting that data are objective, but they do not explicitly make claims about how objective data are.	…So in a way, I would agree that data cannot be racist… (Jacky) I agree on this because data consists of quantitative information (Margarita) I agree that inherently, numbers do not lie… (Robert)

Table 7.3: Centers of Focus on the second question about data neutrality (Class 3): “A friend tells you that ‘data can't be racist, numbers don't lie.’ Do you agree or disagree with this statement? Explain”, Continued

Cluster Code	Subcode	PPDAC Phase(s)	Description	Examples
Both	Interpreters	Conclusion	Statements suggest that people may have expectations or preconceived beliefs that influence their interpretation of data conclusions. May occur on purpose or not on purpose.	…but the data interpreters can be for neglecting to look at the bigger picture. (Jacky)
	Manipulated Conclusion	Conclusion	Statements suggest that data or numbers can be manipulated, usually to advance the researchers' personal agenda and typically during the conclusion phase of the statistical investigation cycle.	However, this data could be used in several methods which may make it seem untrue… (Margarita)
Not Neutral	Researcher Bias	PPDAC	Statements suggest that the statistician or data scientists may have expectations or preconceived beliefs that influence their engagement with the data at any point during the statistical investigation cycle and lead to potential biases. May occur on purpose or not on purpose.	Data can be racist when it sets out to show disproportion through deficit framework. When studies look at failure rates instead of achievement rates it can be very racist… (Elenai)

Table 7.3: Centers of Focus on the second question about data neutrality (Class 3): “A friend tells you that ‘data can't be racist, numbers don't lie.’ Do you agree or disagree with this statement? Explain”, Continued

Cluster Code	Subcode	PPDAC Phase(s)	Description	Examples
Not Neutral	Sampling	Plan / Data Conclusion	Statements suggest that sampling may lead to potential biases (e.g., over- or underrepresentation), often related to the generalizability of the research. Related to the planning and data phases of the statistical investigation cycle.	Numbers are inherently racist because numbers can be given without any sort of context and can be collected out of bias. I mean that whole races of people can be excluded or the ones that are collected can be from POC people that are outliers. (Caden) …Depending on how you collect your data, how many people are surveyed, types of people surveyed can all make a difference in what backs or creates your “evidence”. If the methods of collection are not biased the results will not be reflective of the target population creating racist data. (Elenai) ..Along with that if you survey people about a topics they do not know about then that would might show how the population might not be interested in that topic yet the people that might know the topic would show it is important (Jaime)
	Manipulated Conclusion	Conclusion	Statements suggest that data or numbers can be manipulated, usually to advance the researchers' personal agenda and typically during the conclusion phase of the statistical investigation cycle.	…You can use data and spin it in such a way that it becomes racist… (Elenai) …numbers can be manipulated to say pretty much whatever they want… (Jaime)

Table 7.4: Centers of Focus on the third question about data neutrality (Class 6): “One of your colleagues states that ‘technology is politically neutral, therefore data is politically neutral.’ Do you agree or disagree? Explain”

Cluster Code	Subcode	PPDAC Phase(s)	Description	Examples
Neutral	Well-intended	PPDAC	Statements suggests that statisticians, data scientists, or researchers are well-intended and do not aim to use data in a racist way	…I don't believe that whoever creates some sort of technology, creates it to be racist. I believe that it was made to be neutral. (Ellie)
Both	Researcher Bias	PPDAC	Statements suggest that the statistician or data scientists may have expectations or preconceived beliefs that influence their engagement with the data at any point during the statistical investigation cycle and lead to potential biases. May occur on purpose or not on purpose.	I disagree since technology could be created by different types of people. There may be several types of people who create algorithms and/or programs which are used to support their political stance whether they are for or against a certain topic… (Margarita)
	Objective	Conclusion	Statements suggest that data are factual, definitive, or another form of objective truth or evidence (e.g., to support a claim). Related to the conclusion phase of the statistical investigation cycle.	However, data is neutral. Data are facts deprived from people, communities, and more. Data could be something simple such as measuring the circumference of your head. This example shows that data has no political stance. (Margarita)
Not Neutral	Researcher Bias	PPDAC	Statements suggest that the statistician or data scientists may have expectations or preconceived beliefs that influence their engagement with the data at any point during the statistical investigation cycle and lead to potential biases. May occur on purpose or not on purpose.	I will hardly ever agree that something is politically neutral in the data world. Because time and time again human errors have shown racism and bias in not only who they sample for data but what they sample them for…. (Caden) Data must be collected and therefore data can have bias based on those who collect it. (Robert)

Table 7.4: Centers of Focus on the third question about data neutrality (Class 6): “One of your colleagues states that ‘technology is politically neutral, therefore data is politically neutral.’ Do you agree or disagree? Explain”, Continued

Cluster Code	Subcode	PPDAC Phase(s)	Description	Examples
Not Neutral	Manipulated Conclusion	Conclusion	Statements suggest that data or numbers can be manipulated, usually to advance the researchers' personal agenda and typically during the conclusion phase of the statistical investigation cycle.	We as people have been programmed to believe that technology is politically neutral and therefore data is politically neutral but that is not true. We have found that technology is made up of algorithms and they can be flawed or easily manipulated and reflect inequalities not neutrality. (Elenai) Before I would have agreed but lately, with the exposure to the subjectivity of data interpretation, I would have to disagree. It’s true that 10 means 10 but then we forget how behind these sets of data is a human being interacting with it and coming to conclusions. (Jacky)
	Algorithmic Bias	PPDAC	Statements refer to algorithmic bias that encodes systemic biases (racism, sexism, other forms of discrimination) that create differentiated outcomes	…Humans are inherently bias and because they are the ones creating these machines and learning systems, their creations are also biased. (Caden) We as people have been programmed to believe that technology is politically neutral and therefore data is politically neutral but that is not true. We have found that technology is made up of algorithms and they can be flawed or easily manipulated and reflect inequalities not neutrality. (Elenai)

Table 7.4: Centers of Focus on the third question about data neutrality (Class 6): “One of your colleagues states that ‘technology is politically neutral, therefore data is politically neutral.’ Do you agree or disagree? Explain”, Continued

Cluster Code	Subcode	PPDAC Phase(s)	Description	Examples
	Algorithmic Bias (Continued)	PPDAC		I disagree with it fully… machine(s) learns from the humans' writing so if there is a bias that the humans have written it then it will leak into the code. Like the University that had the machine-learned what was, it had a bias to spot White faces better and that was due to the lack of other faces that the machine was shown as examples…. (Jaime)
	Agency	Conclusion	Statements show agency by showing possible ways to mitigate algorithmic bias. Note: this is not asked in the question, but emerged as code since a student mentioned this.	…so we should be able to teach the machine better if we give other examples. (Jaime)

Table 7.5: Centers of Focus on the third question about data neutrality (Class 12): “A student says, ‘numbers speak for themselves.’ Do you agree or disagree? Please explain.”

Cluster Code	Subcode	PPDAC Phase(s)	Description	Examples
Neutral	Objective	Conclusion	Statements suggest that data are factual, definitive, or another form of objective truth or evidence (e.g., to support a claim). Related to the conclusion phase of the statistical investigation cycle.	I agree. Numbers are facts, they have no opinions and state the obvious. (Margarita)
Not Neutral	Question Framing	Problem	Statements suggest that the statisticians’ or data scientists’ research questions (e.g., anti-deficit or deficit) may influence their engagement with the data and lead to potential biases. Usually occurs in the beginning of the statistical investigation process (the problem phase). This is similar to the subcode on researcher bias, but specifically considers research questions.	Numbers no longer speak for themselves. It is the formation of deficit and anti-deficit statements or questions that create the numbers. (Elenai)
	Context	PPDAC	Statements suggest that the context provides meaning to the data and may help identify any potential biases. This may occur at any point of the statistical investigation cycle	No. Numbers are just numbers but in order to interpret, we must look at the bigger picture. This might show if there is anything wrong with the data or how to interpret the data. (Jacky)
	Outside Factors	PPDAC	Statements suggest that there may be outside factors (e.g., latent variables, sampling, survey bias, outliers) that may cause biases, but do not state specific examples of outside factors. These outside factors may influence any phase of the statistical investigation cycle	No they don't. Numbers are inherently biased because they do not account for outliers or uncontrollable variable (Caden)
	General Bias	Conclusion	Statements suggest that the data or numbers may come from bias research, but do not specify the type of bias	No, the numbers can come from bias research (Robert)
	Manipulated Conclusion	Conclusion	Statements suggest that data or numbers can be manipulated, usually to advance the researchers' personal agenda and typically during the conclusion phase of the statistical investigation cycle.	I would disagree as numbers can be manipulated and do not always show the real data (Jaime)

data under racist perspectives if they are not considering the historical and political context from which the data comes from.

The *Interpreters* subcode was also applied to Jacky's statement from the Class 3 homework. Particularly, the interpreters subcode was used to describe statements where students suggest that people may have expectations or preconceived beliefs that influence their interpretation of data conclusions. The expectations or preconceived beliefs may be explicit or implicit. Furthermore, unlike the *Researcher Bias* subcode, this subcode does not specify if the "interpreters" are the audience, statisticians, or data scientists. If the interpreters are the statisticians or data scientists, then this may be subcoded as *Researcher Bias* instead. However, in other homework assignments, Jacky also refers to the audience as people who are reading a data analysis report. If Jacky meant interpreters as the audience, then this is the only subcode that talks about potential biases from the audience perspective.

The *Well-Intended* subcode emerged in the Class 6 homework. The Class 6 homework was: "One of your colleagues states that 'technology is politically neutral, therefore data is politically neutral.' Do you agree or disagree? Explain." This subcode was used to describe statements that suggest that statisticians, data scientists, or researchers are well-intended and do not aim to use data in a racist way. For example, a student stated: "I agree because **I don't believe that whoever creates some sort of technology, creates it to be racist**. I believe that it was **made to be neutral**." This response is talking about the intent of how data was meant to be used. This does not necessarily suggest that the student does not think that data is the result of a social process or that there may be potential biases in data.

The *Algorithmic Bias* subcode also emerged in the Class 6 homework. This subcode was used to describe statements that refer to algorithmic bias that encodes systemic biases (racism, sexism, other forms of discrimination) and creates differentiated outcomes. For example:

> I will hardly ever agree that something is politically neutral in the data world… **Humans are inherently bias and because they are the ones creating these machines and learning systems, their creations are also biased.** (Caden)
>
> We as people have been programmed to believe that technology is politically neutral and therefore data is politically neutral but that is not true. We have found that technology is made up of **algorithms and they can be flawed or easily manipulated and reflect inequalities not neutrality.** (Elenai)
>
> I disagree with it fully… **machine(s) learns from the humans' writing so if there is a bias that the humans have written it then it will leak into the code.** Like the University that had the machine-learned what was, it had a bias to spot White faces better and that was due to the lack of other faces that the machine was shown as examples…. so **we should be able to teach the machine better if we give other examples**. (Jaime)

All three statements refer to human biases that are encoded into algorithms, whether implicit or explicit. One interpretation of this algorithmic bias is that algorithms are a social construct that are created by humans and encode human biases.

The *Agency* subcode was the third subcode that emerged in the Class 6 homework. This subcode describes responses that mention ways to mitigate algorithmic bias. It is important to note that the Class 6 homework did not necessarily ask for agency. Particularly, when referring to the facial recognition example, Jaime mentioned how algorithmics may be improved through better sampling (e.g., a more diverse training set for the facial recognition algorithm). In other words, Jaime recognized that data and algorithms may encode human biases, but Jaime also showed a sense of agency by providing examples of how algorithmic bias can be addressed and how algorithms can be improved.

The remaining of the subcodes appeared during the Class 12 question in the survey. The Class 12 question was the same as the Class 1 question: "A student says, 'numbers speak for themselves.' Do you agree or disagree? Please explain." It is worth noting that the responses to the Class 12 question were relatively shorter than the responses for the previous questions. This may be attributed to the question being part of a larger survey (rather than a short homework assignment like Class 3 and Class 6), because the question was collected near the end of the class, or for another unknown reason.

The *Question Framing* describes statements that suggest that the statisticians' or data scientists' research questions (e.g., anti-deficit or deficit) may influence their engagement with the data and lead to potential biases. This is similar to the subcode on researcher bias, but specifically considers research questions. For example, Elenai stated that "Numbers no longer speak for themselves. It is the **formation of deficit and anti-deficit statements or questions that create the numbers**." One interpretation of Elenai's statement is that biases can often be rooted as early as the when the statistician or data scientists are forming statistical questions, where deficit or anti-deficit questions may lead to different types of engagement with data throughout the PPDAC cycle.

The *Context* subcode describes statements where students suggest that the context provides meaning to the data and may help identify any potential biases. This may occur at any point of the statistical investigation cycle. For example, Jacky stated that "No. Numbers are just numbers but in order to interpret, **we must look at the bigger picture**. This **might show if there is anything wrong with the data or how to interpret the data**." One interpretation of this is that the context provides meaning to the data. However, Jacky extends this by suggesting that the context might help highlight any concerns with the data and interpretation. In other words, if we don't consider the context, we might misinterpret or misuse the data. It is worth noting that this subcode is similar to the *Context - Historical and Political* subcode, but does not explicitly state anything about the historical and political nature of the data. In fact, Jacky's response to the Class 3 homework was coded as *Context - Historical and Political*, so it may be possible that when she refers to the "bigger picture" she is also referring to the historical and political context although it is not explicitly stated here.

In what follows, I present the focusing interactions, features of the task, and nature of mathematical activity for the CoFs related to: (a) *Context - Historical and Political*, (b) *Algorithmic Bias* and *Agency*, and (c) *Question Framing*. The CoFs related to *Context* and *General Bias* were not included in this analysis because they only appeared once or students did not give broad consent. Furthermore, the students who provided the statements coded as *Context* and *General Bias* had provided similar but more specific responses in previous assignments (*Context - Historical and Political* and *Researcher Bias*, respectively). Thus, instead of suggesting that the CoF shifted to a more general CoF, it is possible that the students noticed the same CoF but provided less detail in their responses to the Class 12 assignment.

Analytical Passes for the Centers of Focus That Emerged in The Class

The next three analytical passes entailed reading through the contact summary forms to help identify instances that may have helped highlight particular Centers of Focus, then read through the field notes to identify specific time frames in the class, and finally watched the classroom video of whole class conversations to identify specific interactions. In the videos, I particularly looked for or described the: (a) focusing interactions, (b) features of the task, (c) nature of the mathematical activity. I also add a final analytical pass connecting four analytical passes to the design features.

Center of Focus: Context - Historical and Political

Analytical Pass 2: **Focusing Interactions**. The first focusing interaction I present is for the *Context - Historical and Political* Center of Focus. This CoF appeared during the Class 3 homework, so I searched the classroom data on and prior to Class 3 for focusing interactions related to this CoF. There was one focusing interaction that was related to the *Context - Historical and Political* CoF that occurred at the end of Class 3 before the homework was assigned. An overview of Class 3 is shown in Figure 7.3.

In the first half of the class session, students worked on a short simulation activity in breakout rooms with the goal of preparing a two to three-minute presentation on any findings. This activity was open-ended to allow students to create their own statistical questions, select which data to use, and create visuals for their presentation. After the activity, we debriefed the activity and made connections to the PPDAC cycle (which they had read about for homework). The second part of the lesson introduced CRT and asks students to provide examples of the implications of the five tenets of CRT in education for both mathematics teaching and learning as well as statistics and data science (Figure 7.4). The goal of the activity was to discuss "what Critical Race Theory is and talk about what it means for math learning and what it means for statistics and data science" (said by the instructor while presenting the activity).

The focusing interaction occurred during the debrief of the classroom activity, where students provided implications of the five tenets of CRT for statistics and data science. During the debrief, Group 3 was summarizing what they said about the implications of the centrality of experiential knowledge to statistics and data science. They said that statisticians and data scientists should use "experiences and people's stories and **history to contextualize and add**

<table>
<tr><th colspan="2">Class 3: Statistical and Data Investigation, Introduction to Critical Race Theory (CRT)</th></tr>
<tr><th colspan="2">Lesson Summary</th></tr>
<tr><td colspan="2">In the first part of today’s lesson, students will work in groups on a short simulation activity. In this activity, students will work in groups to organize data that is randomly simulated (each group will have a different sample) and then prepare a 2–3-minute presentation about their findings. Note: this is an open-ended and vague task on purpose. After this activity, we will have a class discussion about how this relates to different parts of the PPDAC cycle (which they read about for homework), highlighting specific moments where students engaged with different parts of the cycle.
In the second part of today’s lesson, we build on the first part by learning about Critical Race Theory and its implications for the PPDAC cycle. Students will continue to read about CRT and QuantCrit for HW.</td></tr>
<tr><th>Statistical Pedagogical Goals</th><th>Social Justice Pedagogical Goals</th></tr>
<tr><td>Understand the different components of the statistical problem-solving process and how they may be used to formulate statistical questions, collect and consider data, analyze data, and interpret results</td><td>Diversity 8: DI.9-12.8
I respectfully express curiosity about the history and lived experiences of others and exchange ideas and beliefs in an open-minded way

Justice 12: JU.9-12.12
I can recognize, describe and distinguish unfairness and injustice at different levels of society.</td></tr>
<tr><th colspan="2">Homework</th></tr>
<tr><td colspan="2">1. Readings
Read the following two papers (preferably in this order)
● Ladson-Billings, G., & Tate IV, W. (1995). Toward a Critical Race Theory of Education. Teachers College Record, 97(1), 47-68. Link provided
● Crawford, C. E., Demack, S., Gillborn, D., & Warmington, P. (2018). Quants and crits: Using numbers for social justice (Or, How Not to be lied to with statistics). Understanding critical race research methods and methodologies: Lessons from the field, 125-137. Link provided
Reflection Questions:
a. Is there anything that you want clarity on? (e.g., words, terms, ideas):
b. In your own words, how would you define Critical Race Theory?:
c. How is QuantCrit related to Critical Race Theory?:How can data science be used to advance social justice?:
d. How can QuantCrit be applied to the Problem, Planning, Data, Analysis, and Conclusion (PPDAC) cycle?:
● Problem:
● Planning and Data:
● Analysis:
● Conclusion:
e. A friend tells you that "data can't be racist, numbers don't lie." Do you agree or disagree with this statement? Explain.
2. Project
Write two to three paragraphs about:
a. What is the goal of this research project? What is your research question?
b. What is the context? What do we need to know about this context?
c. Why is this context important?</td></tr>
</table>

Figure 7.3: Lesson summary, pedagogical goals, and homework of Class 3

Questions my group had about CRT:

-

CRT Tenet, description, and implications for mathematics teaching and learning

CRT Tenets	Description	Implication for mathematics teaching and learning	Implications for statistics and data science
1. The centrality and intersectionality of race and racism	(provided)		
2. The challenge of dominant ideology	(provided)		
3. The centrality of experiential knowledge	(provided)		
4. The interdisciplinary perspective	(provided)		
5. The commitment to social justice	(provided)		

Figure 7.4: Information gap for Critical Race Theory tenets description, implications for mathematics teaching and learning, and implications for statistics and data science

meaning to data, which I think is very powerful." One student (Margarita) reacted with a thumbs up emoji, which may have prompted two other students to react with a thumbs up as well. In fact, although other groups also had emoji reactions when they presented, this statement had the most reactions. It is worth noting that Jacky, the student who provided a response to the homework that was coded as *Context - Historical and Political*, was not in Group 3 but was one of the students that reacted with a thumbs up.

After the group was done sharing, I summarized their contributions, brought up other points that they had mentioned in my conversation with them in the breakout room, and made connections with another group:

> There was one group that was getting at the tensions between reinforcing stereotypes with data, which was really cool to see. So, for example, what happens when data says something that you don't necessarily agree with? How do you deal with that? **And I like how we talked about, kind of, situated the data within the larger historical political context** and talking about the sampling methods and biases that, that go with that.

There were at least two MLRs that highlighted student contributions in this interaction: (a) Information Gap, and (b) Collect and Display. First, prior to the debrief, students engaged in an Information Gap routine where students were asked to complete the row for the tenet in Figure 7.4 that corresponded with their breakout room (a form of a jigsaw activity). After about ten minutes, each group shared their responses. Students were expected to focus on their tenet during the breakout room, then focus on the other four tenets as the respective groups presented in the whole class conversation, and told to "please ask follow-up or clarifying questions after each presentation, if you have any" (said by the instructor). The second MLR was the Collect and Display routine. Particularly, as groups shared their responses to their assigned tenet in Figure 7.4, I shared my screen with a blank template of the activity and took notes of what students were saying. Combined, the Information Gap activity may have helped guide students' attention to the new information that was being visibly collected and annotated using the Collect and Display.

A possible third MLR was the Compare and Connect Routine where I made connections with what Group 3 had said about reinforcing stereotypes and a comment that Group 1 had mentioned in a small group about being cautious about using data to support deficit narratives. However, this MLR was not directly connected with the *Context - Historical and Political* Center of Focus. It is also possible that the readings prior to this class (about Critical Race Theory and Quantitative Critical Race Theory) may have guided students' attention to notice the historical and political context of data.

Analytical Pass 3: Features of the Task. The task leading to the classroom debrief was an information gap activity where students were asked to compete complete one row of the activity shown in Figure 7.4. Each row contained a short description of a tenet of CRT in education and students were expected to provide implications for teaching mathematics and statistics as well as implications for engaging with statistics and data. Since there were five breakout rooms, each breakout room worked on one of the tenets, then continued on the following tenet if there was time. Students worked in groups for about ten minutes. Students were also told that they would be expected to prepare a short presentation to the class. During the debrief, I shared my screen with a blank template of the activity and summarized group and student contributions on the blank template as a form of a Collect and Display. I also encouraged students to ask each other follow-up questions, and summarized the group contributions and made connections with other groups when possible. There were also cases when students asked to share their screen to show information.

An affordance of the information gap and collect and display activities were that they encouraged students to work with others in the class to listen carefully to acquire information about the tenets that they were not assigned. As a result, students' perspectives on data neutrality may have shifted as students presented new information, especially as the student contributions were being displayed for the whole class. A possible constraint is that there were many different centers of focus in terms of the implications of the tenets of CRT in education for statistics and data science. Furthermore, this was at the end of the class, so there was a time constraint that limited the class's time to substantially unpack the ideas shared by each group (at least not as intended). This was especially important with Group 3 because I wanted to emphasize the role of the problem context as well as situating the problem context within the larger social, political, cultural, and historical context. However, I did not spend as much time to unpack this because I was concerned that we would not have time for the other groups to present.

In future iterations, I would aim to have this activity towards the beginning or middle of the class meeting. This would allow more time for the students and instructor to unpack and build on particular ideas or goals of the activity, such as discussing of how data is situated in a larger social, historical, and political context. Furthermore, I also noticed that there were not as many follow-up or clarifying questions (only two students asked follow-up or clarifying questions) during the whole class conversation as anticipated. This may have been attributed to the time constraints, because this was towards the beginning of the term (students were engaging with each other more towards the middle and end of the term), or for another reason. To help prompt more student-student engagement, I would also include a form of Gallery Walk where students visit other groups, share ideas, ask questions, make connections, or add other comments to their work. This may provide a different mode of communication that may not be perceived as having a high of a "cost" as contributing to whole class conversations.

Although not in the focusing interaction presented here, the Class 6 homework reading may have also been a pivotal task in focusing students' attention to the *Context - Historical and Political* CoF. Particularly, students were asked to read an article about Critical Race Theory in Education (Ladson-Billings & Tate IV, 1995) as well as one on Quantitative Critical Race Theory (Crawford et al., 2018) before responding to the homework question about data neutrality. Both articles talk about the role of the historical and political context in racial injustices. Crawford et al. (2018) extend CRT by providing implications from the five tenets of CRT in education with statistics and quantitative analysis, similar to the class activity. These implications include considering the role of the historical and political context in racial injustices. Thus, it may be possible that both readings may have guided students' attention to the *Context - Historical and Political* CoF.

Analytical Pass 4: Nature of the Mathematical Activity. The fourth analytical pass entailed identifying any possible classroom norms that may have influenced participation dynamics and, consequently, what students noticed in the activity. Some classroom norms that may have influenced the participation dynamics are: (a) setting up classroom norms for discussing social justice, (b) using the Information Gap and Collect and Display routines to establish the roles of students and teacher, and (c) interacting with presenters.

First, in terms of class rooms, students were asked to respond to the following question during their Class 1 homework: "List anything that the instructor or your fellow students can do to support your engagement with these [social justice] conversations." The goal of this activity was for students to communicate possible ways that I can support social justice conversations and how we can foster a brave space in the course. Then, before the break in Class 3, students were asked to help "establish classroom norms when talking about social justice topics" (said by the instructor). Particularly, each grouped worked on the activity shown in Figure 7.5. This was similar to the Class 1 homework, but added the first question about hurtful words and behaviors and was done in groups instead of individually.

In groups, discuss and respond to the questions below. I will review the responses during the class and share anonymous response to the whole class
1. Please describe any specific words or behaviors that are hurtful when used.
2. List anything that the instructor or your fellow students can do to support your engagement with conversations about social justice

Figure 7.5: Establishing classroom norms for the course

During the break, I looked through and copied all the responses from the Class 1 homework and the Class 3 activity onto a Google Doc with the question from Figure 7.5. All responses were anonymized. Initially, I was going to summarize the student contributions but found that all the contributions were short and likely not hurtful. After the break, students reviewed the responses and we made comparisons and connections as a class. This document was referred to throughout the class.

Although the classroom norms were not explicitly mentioned in the information gap or collect and display activities, I believe that this activity was important in setting the nature of mathematical activity for the this CoF because it provided explicit guidelines for how the classroom community was expected to interact and communicate to establish a safe, brave, and confidential space. This was especially important in this course since it may have been the first time that many students talked about race and racism in the context of mathematics, statistics, or data science.

Second, the information gap activity encouraged students to rely on each other to obtain information about the tenets that they were not assigned. Similarly, each group may have been positioned as an expert in their tenet and were expected to share their understandings with the rest of the class. This may have helped create a classroom culture where students were co-constructing knowledge together as a class rather than supporting a banking method of education (Freire, 1988).

Finally, during the collect and display debrief, students were expected to interact with the presenters. Particularly, I designed the activity so that students would unmute and ask follow-up or clarifying questions. Of the two students that asked follow-up or clarifying questions in this debrief, one unmuted themselves to ask a question and one asked a question in the chat. This may have set up the culture for future activities where students asked more follow-up or clarifying questions during whole class conversations.

I also noticed the beginnings of two new ways of interacting that I did not anticipate and may be attributed to the virtual Zoom platform: using the chat and emoji reactions. The chat and emoji reactions were often used to support or agree with a contribution. These reactions may have helped students notice particular whole class contributions. For example, students might have noticed when a contribution received a lot of emoji reactions and, as a result, remembered that contribution as something that was considered valuable to the classroom community.

Analytical Pass 5: **Connections to the Design Features**. I end by summarizing the first four analytical passes in terms of the design features. Table 7.6 shows a summary of the design features, descriptions, and evidence about the CoF related to *Context - Historical and Political*. A goal of the lesson was to talk about the implications of Critical Race Theory for engaging with statistics and data science, including focusing on the larger structural contexts of oppression (Design Feature 1: Reflect on structures of social injustices). To help set up these conversations, students compared, connected, and summarized anonymous student contributions (Design Feature 3: Communicate, Anonymous Contributions) about classroom norms that may have helped create a safe, brave, and confidential learning environment. After, students were assigned different CRT tenets to summarize and provide implications of these tenets for engaging with statistics and data science (Design Feature 3: Communicate, Information Gap MLR). Then, students presented what they learned to the class as I shared my screen and collected student contributions (Design Feature 3: Communicate, Collect and Display MLR). These activities may have helped establish classroom norms where students are expected to learn from each other and interact with each other in creative ways (e.g., emoji reactions, chat). Further, the idea that had the highest number of emoji reactions was when a student mentioned that data should be contextualized in people's experiences as well as history, which was summarized by the instructor as situating data within a larger historical and political context and reaffirmed in the homework readings (Design Feature 1: Reflect on structures of social injustices).

Table 7.6: Focusing interactions for the Context - Historical and Political CoF in relation to the design features

Design Feature	Description	Class Evidence (Class 3)
1) Reflect on structures of social injustices	Opportunities for students to reflect on the social, political, cultural, and historical contexts of oppression in the context of statistics and data science	In breakout rooms, students discussed the implications of Critical Race Theory to statistics and data science For homework, students read about Critical Race Theory and Quantitative Critical Race Theory
2) Deepen and revise thinking	Opportunities for students to create, reflect, and recreate knowledge across different time scales (e.g., within a class, across two or three classes, across the entire course term).	NA
3) Communicate a) Mathematical Language Routings b) Anonymous contributions c) Journals	Opportunities for students to engage in dialogue with each other and with the professor	**Mathematical Language Routines:** Information gap (MLR 4) - students were assigned different tenets (corresponding to the breakout room they were in) Collect and Display (MLR 2) - debrief breakout room activity and display student responses on a shared classroom artifact **Anonymous Contributions:** Setting up classroom norms **Journals**: "A student says, 'numbers speak for themselves.' Do you agree or disagree? Please explain"
4) Engage with relevant contexts	Incorporating generative themes into the classroom and helping students identify generative themes for their projects	NA
5) Engage with all phases of the statistical investigation cycle	opportunities for students to engage with and discuss how race and racism are embedded into the PPDAC statistical investigation cycle	NA
6) Design and implement a statistical study throughout the course	Course project on a social justice topic of their choosing	NA

Situating data within a larger historical and political context also appeared during Jacky's Class 3 homework (Design Feature 3: Communicate, Journals), which was identified as a Center of Focus. Notably, Jacky was not in the group that talked about the larger historical or political context during the whole class debrief. One interpretation of why Jacky noted this CoF is that the focusing interactions, task features, and nature of the mathematical activity may have helped her notice how data should be contextualized in a historical and political context and, therefore, influenced her response to the homework assignment. Additionally, there may have been discussions in breakout rooms, with her project group members, or other interactions outside of whole class conversations that may have guided her attention to this CoF.

Center of Focus: Algorithmic Bias and Agency

Analytical Pass 2: **Focusing Interactions**. Next, I present the analytical passes for *Algorithmic Bias* and *Agency* CoF. I present both of these CoFs together because they emerged in the Class 6 homework and may be related to the same focusing interaction. I searched the data corpus on and prior to Class 6 for interactions related to these CoFs but there were no interactions during the whole class conversation that were about *Algorithmic Bias* or *Agency*. However, it is possible that the Class 6 homework may have helped focus students' attention on properties of data neutrality related to the Algorithmic Bias and Agency CoFs.

Prior to Class 6, we had discussed statistical questions (Class 4) and descriptive statistics and visualizations in R (Class 5). During Class 6, we reviewed sampling, sampling biases, and probabilistic and nonprobabilistic sampling methods. For the homework, students were asked to watch a video from Vox titled *Are We Automating Racism?* (Fong, 2021). An overarching theme of the YouTube video is that data is not neutral and is the result of a social process. Throughout the video, the host discusses the role of the problem and data (sampling), analysis (machine learning), and conclusion (applying the algorithm) phases in algorithmic biases. The video ends by talking about possible ways to try to mitigate algorithmic bias.

Although there is no evidence from whole class conversations on or prior to Class 6 that the YouTube video may have served as a focusing interaction or discussions that guided students' attention to the *Algorithmic Bias* or *Agency* CoFs, it is evident that the YouTube video played a significant role in focusing students' attention to algorithmic bias and agency for at least three reasons. First, this is the first time that the term "algorithmic bias" was mentioned in any of the course materials and students referred to the term "algorithmic bias" and to examples from the YouTube video in their homework responses. Second, in Class 7, at least five students out of the 12 that were present mentioned that they enjoyed the video in the beginning of Class 7 when we were checking in. Finally, during the post-survey, students were asked:

> One of the learning objectives of this class was learning about how race and racism is interwoven into data science. Were there any specific lessons, activities, etc. that helped you learn about how race and racism are interwoven into data science? Please explain.

Of the nine responses, three students specifically referred to the YouTube video or "video from class." Thus, while there were no focusing interactions or discourse practices during the whole class conversation that guided students' attention to the *Algorithmic Bias* or *Agency* CoFs, it is possible that the task (watching the YouTube video and responding to the homework questions) may have guided students' attention to the *Algorithmic Bias* or *Agency* CoFs.

Analytical Pass 3: Features of the Task. The primary task related to the *Algorithmic Bias* CoF was the YouTube video on algorithmic bias. Since there were no focusing interactions during the whole class conversation related to the *Algorithmic Bias* CoF, I provide specific instances in the video that mention algorithmic bias and how they were related to the day's lesson and homework questions. Below is a description of the video from the YouTube page:

> Many of us assume that tech is neutral, and we have turned to tech as a way to root out racism, sexism, or other "isms" plaguing human decision-making. But as data-driven systems become a bigger and bigger part of our lives, we also notice more and more when they fail, and, more importantly, that they don't fail on everyone equally. *Glad You Asked* host Joss Fong wants to know: Why do we think tech is neutral? How do algorithms become biased? And how can we fix these algorithms before they cause harm? (Fong, 2021)

I chose to assign this video for homework for four reasons: (a) the YouTube video was related to the statistical pedagogical goals about sampling and randomization, (b) the YouTube video presented real-world examples of the intersection of social justice and data science that relate to the social justice pedagogical goals, (c) the YouTube video provided students opportunities to reflect on social injustices that were outside of reading articles (in alignment Design Feature 1), and (d) the YouTube video was shorter than the *Coded Bias* (Kantayya, 2020) movie that inspired parts of this course. The movie was not assigned in the course, but it was mentioned and highly encouraged for students to watch. Although I anticipated that this video may help students see some of the potential racial biases that are encoded into algorithmics, admittedly I did not realize how popular it was among the class.

I present an excerpt of the video in Table 7.7. I selected this excerpt because it is representative of the topics discussed in the YouTube video, has connections with the topics discussed in the class, and is a relatively short excerpt (about three and a half minutes). Table 7.7 shows the timestamp, utterances, and connections to the four reasons for why I assigned that video and how it relates to the *Algorithmic Bias* and *Agency* COFs. In this portion of the YouTube video (Fong, 2021), the host is using the model shown in Figure 7.6 to discuss "how machines become biased" (13:52). Prior to this, the host reviewed an example where facial

Table 7.7: Sample excerpt from the *Are We Automating Racism*? YouTube video

Time	Utterance	Comment
13:54	When someone collects data into a training data set, they can be motivated by things like convenience and cost and end up with data that lacks diversity. **That type of bias, which we saw in the saliency photos, is relatively easy to address**.	**Agency**: Addressing biases may be easy to address in machine learning algorithms.
14:08	If you **include more images representing racial minorities**, you can probably **improve the model's performance on those groups.**	**Sampling and Randomization**: More images that represent racial minorities (and, ultimately, of a random sample) to improve the model performance
14:14	But **sometimes human subjectivity is embedded right into the data itself.** Take crime data for example. Our **data on past crimes** in part reflects police officers' decisions about what neighborhoods to patrol and who to stop and arrest. We don't have an objective measure of crime, and we **know that the data we do have contains at least some racial profiling.** But it's still being used to train crime prediction tools.	**Algorithmic Bias**: Human subjectivity may be encoded in our sampling and randomization processes. For example, we use historical policing data, but also know that that data contains racial profiling.
14:39	And then there's the question of **how the data is structured** over here. Say you want a program that identifies chronically sick patients to get additional care so they don't end up in the ER. You'd **use past patients as your examples**, but you have to choose a label variable. **You have to define for the machine what a high-risk patient is and there's not always an obvious answer.** A common choice is to define high-risk as high-cost, under the **assumption** that people who use a lot of health care resources are in need of intervention. Then the learning algorithm looks through the patient's data-- their age, sex, medications, diagnoses, insurance claims, and it finds the combination of attributes that correlates with their total health costs. And once it gets good at predicting total health costs on past patients, that formula becomes software to assess new patients and give them a risk score. **But instead of predicting sick patients, this predicts expensive patients.** Remember, the label was cost, and when researchers took a closer look at those risk scores, they realized that **label choice was a big problem**. But by then, the algorithm had already been used on millions of Americans	**Data and Sampling**: Statisticians and data scientists have to make decisions about what to collect and how they define a phenomenon of interest, as well as reflect on any assumptions during that decision making process. This directly affects the interpretation of the machine learning algorithms and what they are predicting.

Table 7.7: Sample excerpt from the *Are We Automating Racism*? YouTube video, Continued

Time	Utterance	Comment
16:36	And so what happened is in producing these risk scores and using spending, they failed to recognize that on average **Black people incur fewer costs for a variety of reasons, including institutional racism, including lack of access to high-quality insurance,** and a whole host of other factors. But **not because they're less sick.** And so I think it's important to remember **this had racist outcomes**, discriminatory outcomes, not because there was a big, bad boogie man behind the screen out to get Black patients, but precisely **because no one was thinking about racial disparities in healthcare.** No one thought it would matter. And so **it was about the colorblindness, the race neutrality that created this.**	**Algorithmic Bias**: Colorblindness and race neutrality led to racist outcomes
17:24	The good news is that now the researchers who exposed this and who brought this to light are **working with the company that produced this algorithm to have a better proxy**. So instead of spending, it'll actually be people's actual physical conditions and the rate at which they get sick, et cetera, that is harder to figure out, it's a harder kind of proxy to calculate, but it's more accurate	**Agency**: Once identifying racist outcomes, statisticians, data scientists, and other can work together to make machine learning algorithms more accurate and reflective of the desired process

Figure 7.6: Image from the *Are We Automating Racism?* YouTube video

recognition algorithms were better at identifying faces of light complexion than of darker complexion (similar to the example that would be later discussed in the class).

The first example shown in Table 7.7 discusses algorithmic bias in terms of sampling and randomization. Particularly, the host notes that the algorithmic bias from the facial recognition example occurred because the majority of the images used to create the machine learning algorithm (the training data set) oversampled White people. In fact, in another portion of the video, she referred to a data set of 264 photos containing faces from the MIT1003 Saliency Dataset (Judd et al., 2012). Of the 264 photos, about 67% were faces of people identified as White, 11% were of people identified as Other Non-White, 8% were of people identified as East Asian, 7% were of cats, 4% were of people identified as Black, and 4% were of statues. In other words, White faces were slightly overrepresented in the sample (about 63% of the population in the USA in 2012 and 67% of the data) whereas Black faces were underrepresented in the sample (about 14% of the population in the USA 2012 and 4% of the data). Turning to the *Agency* CoF, the host also notes that one way to improve an algorithm like this is to "include more images representing racial minorities." In other words, include a more diverse sample. This was an activity that we did using similar data in Class 12 and Class 13.

That the host also makes connections with a more obviously political example with racist outcomes, particularly about the algorithmic bias in predictive policing and predicting sick patients. In both cases, the host talks about how sampling and how we define constructs in data (e.g. needing medical attention) may lead to racist outcomes, whether intentional or not. The host shows what agency might look like in this situation by bringing awareness to how statisticians, data scientists, and others (e.g., the health insurance company) may work together to make machine learning algorithms more accurate and reflective of the desired process (e.g., better sampling). Interestingly, the agency largely belongs to the data scientists, computer programmers, and other people building the algorithm (rather than people victimized by the racist outcomes of the algorithm). Nonetheless, people using the algorithm may still bring awareness to some of the algorithmic biases that could spark change.

The homework questions for this video (shown in Figure 7.7), further guided students' attention to the *Algorithmic Biases* CoF by asking them to describe the machine learning process that is part of algorithms (Question a), define algorithmic bias (Question b), and state how social biases are reinforced in algorithms (Question d). In terms of the *Agency* CoF, a core component of the course project was to identify action steps that the students could take to advance social justice in the topic they chose. Similarly, the last question in the homework may have focused students' attention to the *Agency* CoF because it asked students to identify ways to "help mitigate algorithmic bias" (Question e).

> Watch this YouTube video and answer the following questions. When possible, provide examples from the video to support your claim.
>
> a. Around 6:45, the video introduces "**how machine learning works and what can go wrong**." Assume that you are teaching a high school data science course and one student asks you "how does machine learning work?" How would you respond?
> b. In the video, they mentioned "**algorithmic bias.**" What do they mean by this term and how might this be related to sampling? How might it be related to QuantCrit?
> c. What is the role of humans in machine learning? Provide at least two to three examples to support your claim.
> d. In her book Race After Technology, Ruha Benjamin states that "Ultimately the danger of the New Jim Code positioning is that existing **social biases are reinforced** – yes. But new methods of social control are produced as well. Does this mean that every form of technological prediction or personalization has racist effects? Not necessarily. It means that, whenever we hear the promises of tech being extolled, our antennae should pop up to question what all that hype of "better, faster, fairer" might be hiding and making us ignore. And, when bias and inequity come to light, "lack of intention" to harm is not a viable alibi. One cannot reap the reward when things go right but downplay responsibility when they go wrong."
>   i. Assume that you are teaching a high school data science course and want to discuss this with your students. How would you explain this to them?
> e. What can we do to help mitigate **algorithmic bias**?
> f. One of your colleagues states that "technology is politically neutral, therefore data is politically neutral." Do you agree or disagree? Explain

Figure 7.7: Reflection questions for the YouTube video on Algorithmic bias

One possible affordance of the homework was that it provided a video, instead of an article, where students are able to visually see the role of algorithmic bias and agency in the machine learning process. This is especially important for machine learning since it is often portrayed as a "black box" or technologically neutral process. For example, in Figure 7.6, the author visualized how algorithmic often starts with the world that we live in (Problem phase of the PPDAC cycle), which may influence how we sample and define data (Data and Planning phases of the PPDAC cycle) and the learning algorithms and predictive models that we create (Analysis phase of the PPDAC cycle). Understanding the sampling and randomization process may have been necessary for students to understand what algorithmic bias looks like in machine learning and how they can engage in forms of agency to combat that bias.

A second possible affordance and related the first design feature about reflecting on the structures of injustices is that videos provided another mode for students to interact with the content outside of course lectures or scholarly readings. Drawing on my own experiences, I found that having breaks from lectures or scholarly readings often reenergized me. I included this video for that reason and to provide a different way of engaging with the material.

A third possible affordance of this homework is that the video includes examples of algorithmic bias and what agency may look like in the context of statistics and data science. For example, the host talks about how Twitter crops pictures, how Twitter users taught a Twitter bot how to be racist, biases in hand sensors for soap and water in the bathroom, beauty filters that lighten skin tone to make people look "hotter" or make noses thinner, facial recognition biases in policing data, and biases in predicting sick people. In fact, they replicated the study about Twitter cropping photos by posting pictures similar to the one shown in Figure 7.8 to see if Twitter's algorithmic automatically cropped the pictures on the dark skinned or light skinned faces. They noticed that dark skinned people appeared in the crop 131 times (36%) and light skinned people appeared in the crop 229 times (64%). Thus, one interpretation of this task is that the uses of multiple, reproducible, and explicit examples of algorithmic bias may have helped guide students' attention to what algorithmic bias was, how it can be detected, and what agency may look like in the context of statistics and data science.

In future iterations of the task and knowing how popular this activity was for the students, I would include this video during a whole class lesson after Class 6. For example, after learning about sampling and randomization in Class 6, we may watch the full video (about 23 minutes) or excerpts from the video in the beginning of Class 7. Then, in breakout rooms, we may discuss some of the connections with sampling and randomization, discuss what algorithmic bias and how biases are encoded into the machine learning process, and revisit the role of race and racism in the PPDAC cycle. In doing so, we may be able to have focusing interactions that highlight different CoFs across students and breakout rooms rather than limiting the interactions to individual homework assignments. This combination may be further supported through the use of MLRs in class discussions.

Analytical Pass 4: Nature of the Mathematical Activity. Since the focusing interaction and task occurred individually as students worked on the homework, it is unclear if there were any classroom dynamics or norms that guided students' attention to the *Algorithmic Bias* or *Agency* CoFs besides general expectations for the homework. For example, students were expected to complete the entire homework and provide specific examples when necessary (as stated in the prompt). This may be why nearly every student (13 out of 14) referenced one of the examples of algorithmic bias that were mentioned in the video. Furthermore, the focus on race and racism in the context of statistics and data sciences may have further guided students'

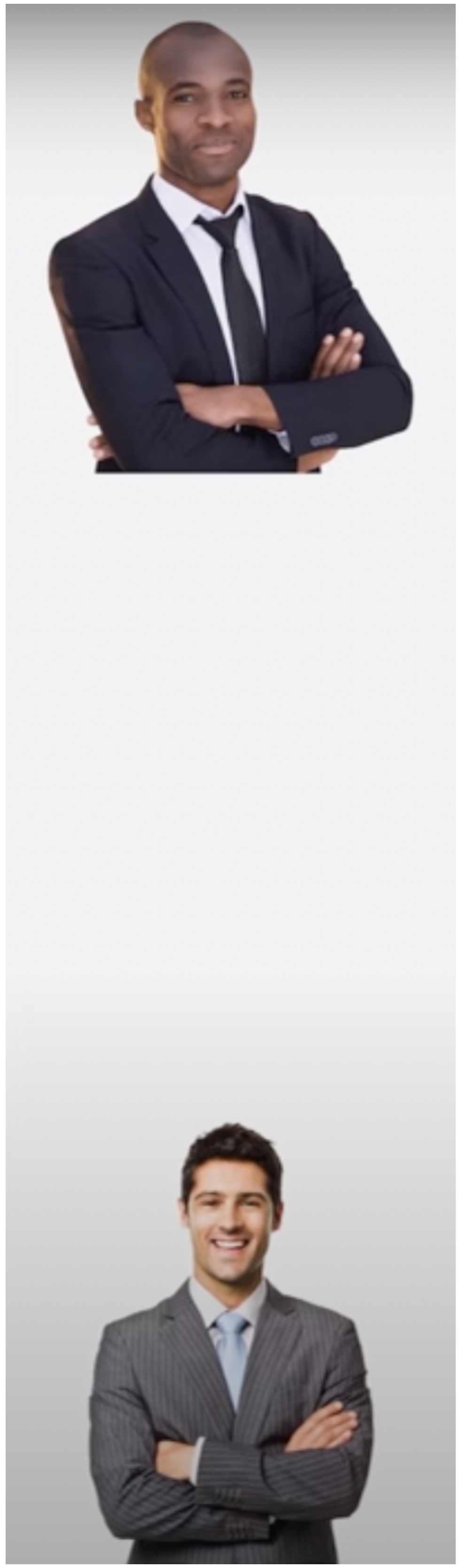


Figure 7.8: Sample photo used to identify biases in how Twitter crops photos (2:03)

attention to the role of race and racism in the machine learning process, which may be why six out of 14 of the students mention a form of racial bias or racist outcome.

Analytical Pass 5: **Connections to the Design Features**. Table 7.8 shows a summary of the design features, descriptions, and evidence about the CoF related to *Algorithmic Bias* and *Agency*. The CoFs related to *Algorithmic Bias* and *Agency* first appeared in the Class 6 homework, but there were no interactions during whole class conversations related to these CoFs prior to Class 6 (although there were some after Class 6). As part of a journal reflection in their Class 6 homework (Design Feature 3: Communicate, journals), students watched a video that talked about the role of social, political, cultural, and historical contexts on algorithmic bias (Design Feature 1: Reflect on the structures of social justice). In this sense, the task was watching the YouTube video and responding to the reflection questions in the homework, and the focusing interaction was between the student, YouTube video, and homework. This may have guided students' attention to particular aspects of how data is not neutral.

For example, Caden stated that "[h]umans are inherently bias and because they are the ones creating these machines and learning systems, their creations are also biased" (Caden). Similarly, Elenai noted that algorithms "reflect inequalities not neutrality" and Jaime stated that the "machine learns from the humans' writing so if there is a bias that the humans have written it then it will leak into the code." In all three of these student responses, students are referring to the socialization of data, where data exists in the world we live in (the Problem phase of the PPDAC cycle) which shapes what data is collected, how, and what the data actually measures (the Planning and Data phase of the PPDAC cycle), the algorithmics we create (the Analysis phase of the PPDAC cycle?), and the stories that are told with data (the Conclusion phase of the PPDAC cycle).

Table 7.8: Focusing interactions for the Algorithmic Bias CoF in relation to the design features

Design Feature	Description	Class Evidence (Class 6)
1) Reflect on structures of social injustices	Opportunities for students to reflect on the social, political, cultural, and historical contexts of oppression in the context of statistics and data science	*Are We Automating Racism?* YouTube video (Fong, 2021)
2) Deepen and revise thinking	Opportunities for students to create, reflect, and recreate knowledge across different time scales (e.g., within a class, across two or three classes, across the entire course term).	NA
3) Communicate a) Mathematical Language Routings b) Anonymous contributions c) Journals	Opportunities for students to engage in dialogue with each other and with the professor	**Mathematical Language Routines**: NA **Anonymous Contributions**: NA **Journals**: "One of your colleagues states that 'technology is politically neutral, therefore data is politically neutral.' Do you agree or disagree? Explain"
4) Engage with relevant contexts	Incorporating generative themes into the classroom and helping students identify generative themes for their projects	Algorithmic Bias: Racist outcomes related to facial recognition algorithms, healthcare, twitter, etc.
5) Engage with all phases of the statistical investigation cycle	opportunities for students to engage with and discuss how race and racism are embedded into the PPDAC statistical investigation cycle	NA
6) Design and implement a statistical study throughout the course	Course project on a social justice topic of their choosing	NA

Jaime also added that "we should be able to teach the machine better if we give other examples." One interpretation of Jaime's full statement is that everyone carries implicit biases, and therefore we run the risk of encoding these biases into algorithmics that may lead to racist outcomes. However, Ruha Benjamin notes that these racist outcomes occur "not because there was a big, bad boogie man behind the screen out to get Black patients, but precisely because no one was thinking about racial disparities in healthcare" (Fong, 2021). Thus, Jaime may recognize that racism may be encoded into algorithmics, but he may also draw from the class lesson on sampling and randomization as well as the video examples to guide his agency by identifying ways to better the algorithmics.

Center of Focus: Question Framing

Analytical Pass 2: **Focusing Interactions**. The next focusing interaction I present is for the *Question Framing* CoF. This CoF refers to statements that refer to how research or statistical questions (e.g., anti-deficit or deficit) may influence our engagement with the data and lead to potential biases. This is similar to the *Researcher Bias* CoF, but specifically focuses on how implicit biases are manifested in research or statistical questions. This CoF appeared during the Class 12 homework, so I searched the interactions on and prior to Class 12 for interactions related to this CoF and the design features. There was one main focusing interaction related to the *Question Framing* CoF that occurred during Class 4. An overview of Class 4 is shown in Figure 7.9. Prior to the class, students learned about statistical questions and drafted statistical questions for their course project. In the first part of the lesson, we used the R programming language to talk about different types of variation (variability within a group, variability within and across groups as well as covariation, and variability in model fitting). We then looked at different research questions and determined if the questions were statistical and, if so, what type of variation each question was considering. In the second part of the lesson, we talked about deficit and anti-deficit framings of research questions. We ended the class by revisiting the

QuantCrit implications for the course project, focusing on the “guiding questions and tips” as we critically engage with the PPDAC cycle (Figure 4.8).

The focusing interaction came from an activity where students looked at sample deficit and anti-deficit research questions from Harper (2010) and answered the questions shown in Figure 7.10. Students were also asked to read the article for homework. The purpose of this activity was for students to compare different research questions about a similar topic to (a) help create a classroom definition of what an anti-deficit research question is, and (b) help students create an anti-deficit research question for their own project. Similar to the other breakout activities in the course, students were asked to be prepared to share during a whole class debrief. There were four breakout rooms, and each breakout room was randomly selected to summarize their conversations during the whole class debrief. As students talked, I summarized student responses, shown in Figure 7.11. A transcript of the whole class debrief is shown in Table 7.9.

The general conversation was about how deficit framed research questions may carry assumptions about particular groups or types people whereas anti-deficit questions focus more on social structures and the surrounding context of the data. For example, Madelyn noted that the anti-deficit questions shown in Figure 7.10 may have “insinuated things about People of Color that, you know, weren’t, you know, true.” Similarly, Jacky referred to the third question from Figure 7.10 (Why are their grades and other indicators of academic achievement disproportionately lower than those of their White and Asian American counterparts?, Harper, 2010, p. 69) and stated that the question carries assumptions about academic achievement across groups of students “before we even have data.” In other words, research questions may enforce a model-minority myth and engage in gap-gazing discourses before sampling, analyzing, and

Class 4: Statistical Questions. Reviewing QuantCrit and its implications for this class

Lesson Summary

In the first part of today's lesson, we talk about three types of variation using R: (a) variability within a group, (b) variability within and across groups as well as covariation, and (c) variability in model fitting. Students will then look at different questions (some statistical, some not although not called "statistical questions yet") and describe the different types of variation that each question is considering. After showing which ones are statistical questions, we will come up with a class definition of what statistical questions are.

The second part of today's lesson will build on Class 3 and our understanding of statistical questions. We will talk about deficit and anti-deficit framing of questions by looking at some examples by Harper (2010) and then talk about Harper's (2010) and QuantCrit's implication for the course project. We will end by creating a class set of guiding questions to consider for this project related to: 1. What is the goal of this research project? 2. What is your research question? 3. What is the context? What do we need to know about this context? and 4. Why is this context important? Note: These are the same questions they had for homework.

Statistical Pedagogical Goals	Social Justice Pedagogical Goals
Understand how to read data, find descriptive statistics of data, and make basic visualizations of data using the R programming language **I.a. Formulate statistical investigative questions** Formulate multivariable statistical investigative questions and determine how data can be collected and analyzed to provide an answer **IV.f. Interpret results** Use multivariate thinking to understand how variables impact one another	**Justice 13: JU.9-12.13** I can explain the short and long-term impact of biased words and behaviors and unjust practices, laws and institutions that limit the rights and freedoms of people based on their identity groups **Action 20 AC.9-12.20** I will join with diverse people to plan and carry out collective action against exclusion, prejudice and discrimination, and we will be thoughtful and creative in our actions in order to achieve our goals.

Homework

1. Reading
Harper, S. R. (2010). An anti-deficit achievement framework for research on students of color in STEM. New Directions for Institutional Research, 2010(148), 63-74. Link provided

This paper discusses how we can write anti-deficit research questions. Answer the following questions:

d. The authors imply that the way we ask questions has an important role in research. Do you agree? Why or why not?:
e. Your student is worried that their research question might carry some deficit-oriented framing. What recommendations would you provide for students to rewrite their question to use an anti-deficit framing?:
f. What, if any, is the relationship between race, racism, and statistics?:

2. Pre-Assessment
Due date

Figure 7.9: Lesson summary, pedagogical goals, and homework of Class 4

With your group, discuss the following task. Be prepared to share as a group :

1. What do the deficit-oriented questions have in common?
2. What do the anti-deficit reframings have in common?
3. If a student asked you how you could transform a deficit-oriented question to an anti-deficit question, what would you say?

Note: These are not all statistical questions, but use this activity to focus on how you can identify deficit-oriented questions and how you can reframe them using anti-deficit language.

Table 6.4. Sample Reframed Research Questions for Students of Color in STEM

Deficit-Oriented Questions	*Anti-Deficit Reframing*
Why do so few pursue STEM majors?	What stimulates and sustains students' interest in attaining degrees in STEM fields?
Why are they so underprepared for college-level mathematics and science courses?	How do STEM achievers from low-resource high schools transcend academic underpreparedness and previous educational disadvantage?
Why are their grades and other indicators of academic achievement disproportionately lower than those of their White and Asian American counterparts?	What enables students of color in STEM to make the dean's list, compete for prestigious fellowships and research opportunities, and earn high GPAs?
Why do so many change their majors to non-STEM fields?	What compels students of color to persist in STEM fields, despite academic challenge and the underrepresentation of same-race peers and faculty?
Why do so few continue on to graduate degree programs in STEM?	What are common aspects of students' pathways from high school completion through doctoral degree attainment in STEM fields?

Figure 7.10: Deficit and anti-deficit research questions using examples from Harper et al. (2010, p. 69)

Implications for education research

1. What do the deficit-oriented questions have in common?

- Really negative. The nature how they were written was negative +1 +1
- Implied things about POC that weren't true, but were assumed +1
- More about a person that we're asking a Q about +1 +1
- Third Q: Assumption that other POC perform different than there counterparts (model-minority)
- They all start with "why". Negative questions

2. What do the anti-deficit reframings have in common?

- More about the surroundings (structures, systems)
- Observed on how and what are things that are impacting. Emphasis on social structures

3. If a student asked you how you could transform a deficit-oriented question to an anti-deficit question, what would you say?

- "why..." to how to we prevent, asking questions about surroundings, structures, systems, etc.

Figure 7.11: Collect and Display activity during the debrief of comparing and connecting deficit and anti-deficit research questions

Table 7.9: Compare and Connect whole class debrief transcript about deficit and anti-deficit framing of research questions in Class 4

Speaker	Group	Utterance
Instructor		So first question is, "what do the deficit-oriented questions have in common?" ((uses random number generator)) Let's go with group two, which is ((student names))
Madelyn	2	We noticed that the deficit-oriented questions were really, really negative like just the nature of how they're written was really, **it like insinuated things about People of Color that, you know, weren't, you know, true**. But like you know **may have been assumed from the data that is in this research**.
Instructor		Great observation. Is there anyone in the group that wants to add something? (8s) Cool. So then I'll go on to ((uses random number generator)) let's say group three. That is Eric and Elenai, if you want to add something to the first question, maybe something else that you notice about the deficit-oriented questions?
Eric	3	I don't know if Elenai you want to go or you want me to go. We both kind of do the same thing, I guess, I already spoke so. We talked about how **one seems to be more oriented in like the person, like a person** ((deficit framing))**, and the other one about its surroundings** ((anti-deficit)). So one just seems to be talking about like, the like person we're, we're asking the question about and the other one's like, well, **what are the surroundings like affecting that, this outcome.**
Instructor		More about the surroundings. ((reading as I typed))
Eric	3	Yeah
Instructor		And so, a lot of times, some of the words that you see in a lot of papers for surroundings you're talking about like structures or systems, which I think is kind of similar to what you mean by surroundings. Thank you. And then we'll move on to ((uses random number generator))let's say number six, which I think was literally no one, let me redo that. ((Random number generator generates three three times in a row)) Three again, oh. Well, maybe this isn't that random. Group number one. Anything about what you notice about the deficit reframing anti deficit reframing? Or comments-
Jacky	1	So what I said about the anti-deficit reframing so, is that they observed how and what our outward things that affect an individual and the performance either negatively or positively, so they were **focusing more on like social structures and how they affect their behavior like that**
Instructor		Social structures. Is there anything else that you or your group want that to add to question one or question three?

Table 7.9: Compare and Connect whole class debrief transcript about deficit and anti-deficit framing of research questions in Class 4, Continued

Speaker	Group	Utterance
Jacky	1	Yeah so I wanted to add how like I believe it was the third question and the deficit oriented questions where it said, like, why are their grades and other indicators of academic achievement this appropriately lower than those of their white and Asian counterparts, and I believe, like it **has that assumption that, like other People of Color** are actually, like, perform, like perform much, much like worse than their white and Asian American counterparts, **which kind of like assumes that model minority kind of a mindset before we even have data**.
Comment		((omitting [time] discussion about model-minority))
Instructor		Yeah, thank you for bringing that up and so. So this gets that I think some similarities between all of these is that kind of the **underlying assumptions in what the question is asking but also kind of what it's ((the question)) implying**. ((5s)) And its interesting how this third question ((Why are their grades and other indicators of academic achievement disproportionately lower than those of their White and Asian American counterparts? Harper 2010, p. 69)) was asked before they even analyzed the data, right? And so typically we asked a question, then talk about what that means here. **But they kind of are already talking about what their assumptions are in the question.** ((6s)) And then I think, let me see if I can run ((random number generator)) again and see if we're actually going to a different group. Well, no. Let's go maybe with the room that has not said anything ((student names)). Is there anything that you would want to add to question one, two, or three?
Ellie	4	Sure, for Question 1, I, it's pretty standard, I guess, I just said that they all start with "why" and that it's also kind of negative, like the other group said all of the questions are just really negative.
Robert	4	yeah, going along with the starting with "why" the number three, if you just replace "Why" with like "how do we prevent it" in most cases. And then it would it would change it from a deficit and type (anti) deficit.
Instructor		Great. That's super helpful. Was there an example that you were thinking of?

Note: Comments are in ((text)) and bold is added for emphasis

There were at least two Mathematical Language Routines that helped highlight student contributions in this focusing interaction: (a) Compare and Connect, and (b) Collect and Display. In this activity, the goal of using the Compare and Connect MLR during the breakout room activity was to encourage students to highlight the differences between deficit and anti-deficit research questions. In doing so, students were expected to notice how they can identify a deficit research question, what makes an ani-deficit research question anti-deficit, and how they can reframe deficit questions to anti-deficit questions. The Collect and Display routine was used during the whole class debrief to highlight specific student contributions. For the most part, I tried my best to use the words that students were using (following Zwiers et al., 2017). In this Collect and Display, I also added "+1" to comments that students agreed with for the first question. For example, Madelyn first noted that the framing of the deficit questions was really negative, which was also supported by Jacky and Ellie. I forgot to repeat this process for the second and third question in Figure 7.11, but there were also opportunities to make connections there. Combined, the Compare and Connect routine may have guided students' attention to similarities and differences across groups which were visibly collected, annotated, and discussed using the Collect and Display routine.

Analytical Pass 3: Features of the Task. The Compare and Connect task was designed to highlight differences between the deficit and anti-deficit research questions. Thus, an affordance of the Compare and Connect activity was that it made the *Question Framing* Center of Focus explicit and the center of the whole class discussion. In that sense, the task was enacted as planned. Furthermore, similar to the *Center of Focus: Context - Historical and Political* CoF, the Collect and Display helped highlight student contributions as their contributions were displayed in the shared screen and through the "+1" that were added to second student contributions.

One constraint of this task is that the anti-deficit research questions provided by Harper (2010) were not statistical questions, or questions that addressed a type of variation. This may have influenced students' final statistical questions for their course project. Particularly, before Class 4, students were asked to write a draft of their statistical questions for their course project and explain what type of variation their question asked. Nine out of 12 of the students wrote questions that were identified as statistical questions, and one of the students initially wrote a question that was identified as an anti-deficit question. After learning about deficit and anti-deficit research questions in Class 4, students were asked "Reflect on today's lesson and the statistical question that you drafted for your course project. Is your statistical question an anti-deficit question? If so, why. If not, rewrite the statistical question so that it is also an anti-deficit question." This was briefly discussed as a class. Four of the nine students that initially wrote statistical questions wrote questions that were identified as anti-deficit but not statistical. While there were more students that wrote anti-deficit questions after Class 4, the decrease in statistical questions may be reflective of how the second half of Class 4 included questions that were not statistical questions.

In future iterations, I may provide deficit and anti-deficit statistical questions. However, I also appreciate that Harper (2010) explicitly outlines examples of deficit and anti-deficit questions and how to reframe deficit questions to anti-deficit questions. Thus, another alternative may be asking students to Compare and Connect the deficit and anti-deficit questions (how it was enacted) but also ask students to write *statistical anti-deficit questions* as part of the activity and why the rewritten questions are statistical and anti-deficit.

Analytical Pass 4: Nature of the Mathematical Activity. The final pass entailed identifying any possible classroom norms that may have influenced participation dynamics and, consequently, what students noticed in the activity. As with the *Context - Historical and Political* CoF that emerged in Class 3, this activity was designed to be centered around student contributions. Particularly, students were expected to debrief their conversations from their breakout room during the whole class conversation. In the contact summary form, I also wrote that there was less wait time between me asking groups to present and the groups presenting. Although more evidence is needed to support this reflection (e.g.., counting the wait time for the previous whole class debriefs), this may indicate the development of a norm around students preparing for the whole class after a breakout room activity.

In terms of students interacting with each other, no one asked follow-up or clarifying questions after each group presented (similar to the interaction from Class 3). However, students were reiterating other students' comments about the difference between deficit and anti-deficit questions. As the instructor, I noted those connections by adding a "+1" to the contributions shown in Figure 7.11as the students were sharing. This may have helped students notice specific contributions by classmate and make connections as well. This was the only time in the course that I used the "+1" to make connections. In future iterations, I would be interested in continuing to use that as a visual way to build connections and bring attention to relations among specific student contributions.

Analytical Pass 5: **Connections to the Design Features**. I end by summarizing the first four analytical passes in terms of the design features. Table 7.10 shows a summary of the design features, descriptions, and evidence related to the *Question Framing* CoF. First, a goal of the lesson was to understand the short and long-term impact of biased words and assumptions of

Table 7.10: Focusing interactions for the Question Framing CoF in relation to the design features

Design Feature	Description	Class Evidence (Class 12)
1) Structures of Social Injustices	Opportunities for students to reflect on the social, political, cultural, and historical contexts of oppression in the context of statistics and data science	Deficit and Anti-Deficit research question framings from Harper (2010)
2) (Re)creating Knowledge	Opportunities for students to create, reflect, and recreate knowledge across different time scales (e.g., within a class, across two or three classes, across the entire course term).	**Journals**: "Reflect on today's lesson and the statistical question that you drafted for your course project. Is your statistical question an anti-deficit question? If so, why. If not, rewrite the statistical question so that it is also an anti-deficit question."
3) Dialogue a) Mathematical Language Routings b) Anonymous contributions c) Journals	Opportunities for students to engage in dialogue with each other and with the professor	**Mathematical Language Routines:** Compare and Connect (MLR 7) - identify differences between deficit and anti-deficit questions Collect and Display (MLR 2) - debrief breakout room activity and display student responses on a shared classroom artifact **Anonymous Contributions:** NA **Journals:** "A student says, 'numbers speak for themselves.' Do you agree or disagree? Please explain."
4) Generative Themes	Incorporating generative themes into the classroom and helping students identify generative themes for their projects	NA
5) Race and Racism in the PPDAC Cycle	opportunities for students to engage with and discuss how race and racism are embedded into the PPDAC statistical investigation cycle	**Problem**: Statistical Questions and anti-deficit research questions, usually first written during the problem phase
6) Course Project	Course project on a social justice topic of their choosing	Drafting statistical questions for the project

statistical questions (Design Feature 1: Reflect on the structures of social injustice). This is specific to the problem phase of the PPDAC cycle, where statistical questions are often first drafted (Design Feature 5: Engage with all phases of the statistical investigation cycle). As part of the activity, students were introduced to deficit and anti-deficit framings of research questions (Design Feature 1: Reflect on the structures of social injustice), then asked to compare, contrast, and identify characteristics of deficit and anti-deficit research questions in breakout rooms (Design Feature 3: Communicate, Compare and Contrast MLR). The activity was debriefed during a whole class conversation where I shared my screen and took notes on the student and group contributions (Design Feature 3: Communicate, Collect and Display MLR), noting similarities across groups when possible. For homework, students were also asked to reflect on their first draft of statistical questions and, if possible, use the first draft to write anti-deficit statistical questions for their course project (Design Feature 2: Deepen and revise knowledge, and Design Feature 6: Design and implement a statistical study throughout the course).

Summary of Focusing Phenomenon Analysis

I drew on elements from Lobato et al.'s (2003, 2013) focusing framework to analyze how discourse practices, tasks, the nature of mathematical activities, and other elements of the writing any reports of the data. In terms of the anti-deficit research questions, Eric notes that the anti-deficit research questions focus on the "surroundings like affecting that, this outcome." Similarly, Jacky brings awareness to the role of the problem context, specifically noting that the anti-deficit research questions attend to the "social structures and how they affect" possible outcomes learning environment may have influenced what students noticed. I specifically focus on students' views on data neutrality and Centers of Focus related to those views. As the class progressed, the majority of the students shifted towards viewing data as a socialized and racialized object which were coordinated with specific lessons and activities. For example, students referred to the lesson where we discussed anti-deficit research questions and how our framing of questions may create some explicit or implicit bias that are embedded throughout the statistical investigation cycle. Furthermore, the video may have helped students understand how race and racism are embedded in the statistical investigation cycle. Notably, the focusing interactions that occurred during whole class discourse entailed highlighting (operated visibly on external phenomena) student contributions using a Collect and Display MLR paired with another MLR (e.g., Compare and Connect or Information Gap).

Chapter 8: Discussion

The purpose of this dissertation was to study the potential that using a social-justice oriented approach to teaching content courses may have on pre-service mathematics teachers' understanding of both statistics and social justice. In doing so, I aimed to challenge a perceived dichotomy between content (e.g., mathematics, statistics, or data science) and social justice. Specifically, I argue that social justice is necessary and part of data science if we view data science as a holistic process that includes all phases of the Problem-Plan-Data-Analysis-Conclusion cycle.

This dissertation included designing and teaching a data science for social justice class to a group of students during the summer 2021 term. There were four result chapters: (a) the first was a description of six design features that aimed to foster students' critical statistical and data scientific consciousness through praxis, (b) the second was a quantitative analysis of pre- and post-assessments, (c) the third was a qualitative analysis of pre- and post-task-based interviews, and (d) the fourth was a qualitative analysis using elements of Lobato et al.'s (2003, 2013) focusing framework to coordinate how aspects of the classroom environment may have guided students' attention towards understanding race and racism in the context of data science. Combined, the results illustrate different aspects of learning in one class designed to introduce PSMTs to the intersectionality between statistics and data science with social and racial justice.

In this discussion, I expand on the intersectionality between statistics and data science and social and racial justice as well as avenues for future work given that intersectionality. I focus on: (a) the role of gap-gazing versus praxis; (b) anti-deficit statistical questions, statistical studies, and experiential knowledge; and (c) hesitancy to talk about race and racism in mathematics and statistics classes. When necessary, I reference data presented in the results chapters, my own experiences, or other sources of data to help guide the discussion of results from this study or to provide insights for future work.

Gap-Gazing versus Praxis

The first point I discuss is the role of gap-gazing in praxis. Freire (1988) defined praxis as a complementary and cyclical relationship between reflection (understanding the social, cultural, historical, and political understandings of social injustices) and action (individual or collective action taken to advance social justices). I also draw on QuantCrit (Castillo & Gillborn, 2022; Crawford et al., 2018; Covarrubias, 2011; Covarrubias et al., 2018; Garcia et al., 2018; Gillborn et al., 2018; Pérez Huber et al., 2018; Sablan, 2019) to foreground the role of race and racism in social justices in our national context.

Furthermore, Gutiérrez (2011) describes gap-gazing (e.g., research on the "achievement gap") as research that places an emphasis on discrepancies between different communities that raise issues about achievement (the dominant axis of equity) "with little concern for how students are constructed in the process, what additional skills are needed to negotiate the discursive spaces of education, and/or how power relations play out in learning" (Gutiérrez, 2017, p. 21-22). From personal experiences and from a statistician's perspective, gap-gazing research might be more appealing for some in educational research because it may be portrayed as straightforward and a relatively simple analysis (e.g., a t-test between two groups, regression to identify any differences in slopes or intercept) when compared to qualitative approaches that aim to transform systems of structures of inequalities.

Relating gap-gazing and praxis, gap-gazing may help identify and bring awareness to educational injustices. Gutiérrez (2017) states that gap-gazing may be "a first step to identify who is not being served well by the school system, but they [students, teachers, or researchers] recognize the limitations of defining equity around such things as 'closing the achievement gap'" (p. 21). That is, gap-gazing may not account for the critical axis of equity (identity and power; Gutiérrez, 2017).

Relating this to the design features of the study, gap-gazing was intentionally included in the course and task-based interviews to foreground these tensions and phenomena. For example, we discussed how if thoughtful analyses of inequities are not situated within a sociopolitical context (DF1: reflect on structures of injustices), the analyses run the risk of reinforcing deficit narratives (Bartell, 2013; Giroux, 2001; Gonzalez, 2009) and sparking or leading to stereotype threat (Brantlinger 2013; Rubel et al. 2016). This may be reflected by comments in the pre-interview where students engaged with gap-gazing research but noticed how "weird" it felt to analyze data from that lens or stated "but like now what?" once they finished the analysis.

However, situating the gaps (i.e., inequalities in achievement that are often observed with educational datasets) within the larger sociopolitical context may have some potential for developing critical consciousness. Particularly, achievement gaps may be situated in a larger context that accounts for the sociopolitical and racialized experiences that give rise to educational inequities and inequalities. For example, in the post-interview, Elenai noted how saying that schools have lower test scores because they have higher proportions of Students of Color "is not right because there's so much more…it's so much more than the student" and states how educational equity is related to the "quality of education, redlining, … and all the other messed up stuff." Elenai's example may be illustrative of how gap-gazing may be helpful for bringing awareness to social and racial injustices, but also extends gap-gazing to situate the "gaps" in a larger sociopolitical context. This helps shift responsibility or blame away from individuals or communities and accounts for the larger structures at play (the reflection component of praxis). If the data analysis is paired with ways to challenge and advance social and racial justice (the action component of praxis), it is possible that a thoughtful analyses of inequities may have the potential to develop into statistical and data scientific critical consciousness. That is, a thoughtful analyses of inequities may be helpful (or sometimes necessary) for identifying and understanding social and racial justice, but gap-gazing by itself is insufficient for advancing social and racial justice.

Turning to statistics and data science courses, teacher preparation, and teaching for social justice, it is important that statistics and data science courses go beyond applying formulas and algorithms and also situate the data within the larger sociopolitical context. One of the biggest concerns with teaching for social justice is the time commitment in class dedicated to providing prior knowledge about the social justice problem context (Bartell, 2013; Gutstein, 2006). Additionally, Gutstein (2006) stated that the real-world projects and related conversions accounted for 15% to 20% of the total class time in his class focused on using mathematics for social justice. On one hand, this may appear to be a significant portion of the class. However, if we take the perspective that statistics and data are numbers situated in a context and that courses should include the all the phases of the PPDAC statistical investigation cycle (Wild & Pfannkuch, 1999), then the 15% to 20% of the total class time discussing the relevant context is appropriate and, in fact, likely central to teaching statistics and data science with fidelity to the discipline! Thus, I argue that statistics and data science lessons do not only entail learning about algorithms, formulas, procedures, or computer programming. Rather, lessons on the sociopolitical context of the data, experiential knowledge related to the data, and other social and racial justice topics are also statistics and data science lessons.

Additionally, it is important for schools and departments to consider ethics when discussing data, especially given the recent increase in data science programs and majors. For example, data scientists may have to consider consent for data collection, tracking, commercialization of individuals' data, and other ethical concerns with using data. Although I think that race and racism may be discussed in all data science courses, data science ethics courses may also provide a place for students to discuss the role of race and racism in data science.

Anti-Deficit Statistical Questions, Statistical Studies, and Experiential Knowledge

The second discussion point is about statistical questions. During the TE, we first discussed statistical questions in Class 4, where we defined statistical questions as questions that "motivate a need to collect data that vary" (from the whole class lecture). This variation includes variation within a group, across groups or covariation, and variation in model fitting. Statistical questions are different from mathematical questions because mathematical questions usually are answered using single value or binary (e.g., yes or no) and will result in the same answer if you ask the same question again. For example, asking "how tall am I?" is mathematical because there is one answer that will not change whereas asking "how tall is everyone in the class?" will vary across classes.

At the end of Class 4, we also read about anti-deficit reframing of research questions (Harper, 2010). Notably, Harper's (2010) anti-deficit research questions were not necessarily statistical questions. This reading raised questions in a breakout room about how to reframe some of the examples of research questions from Harper (2010) into anti-deficit *statistical* questions. One of the students in the breakout room suggested that the questions may be statistical depending on the design of the study. For example, the students in the breakout room referred to a sample anti-deficit question from Harper (2010, p. 69): “What stimulates and sustains students’ interest in attaining degrees in STEM fields?” This question may be answered using qualitative and/or quantitative methods. Qualitatively, this may entail individual or focus group interviews. Quantitatively, this may be answered using a survey where the question is specified so that students are able to select or rank programs, organizations, or other communities and resources that stimulate and sustain students’ interest in attaining STEM degrees. However, the students in the breakout room also noted that the quantitative approach might not include all communities or resources, which might be addressed by a fill-in-the-blank response. This idea then raised new questions–including how fill-in-the-blank answers may raise new issues when cleaning and organizing data.

Additionally, and returning to the framing theoretical perspectives from Critical Race Theory (Crenshaw et al., 1995; Ladson-Billings, 2009; Ladson-Billings & Tate, 1995; Solórzano & Yosso, 2002) and QuantCrit (Castillo & Gillborn, 2022; Crawford et al., 2018; Covarrubias, 2011; Covarrubias et al., 2018; Garcia et al., 2018; Gillborn et al., 2018; Pérez Huber et al., 2018; Sablan, 2019), I wonder how much important information we lose when we aim to quantify experiences. I especially think of the role of experiential knowledge and how experiential knowledge may not be captured using a survey or other quantitative approaches. For example, what is the role of experiential knowledge in statistics and data science? From the pre- and post-task-based interviews, experiential knowledge guided the data that was selected (e.g., race or ethnicity and free or reduced priced meals because they are traditionally associated with educational equity) and situated the data within a larger sociopolitical context (e.g., drawing on their own experiences of education equity). However, this is mostly from the statistician’s or data scientist’s perspective and does not necessarily draw on the experiential knowledge of the participants (people whose data were collected to make the dataset). Thus, if the experiential knowledge of People of Color is appropriate, legitimate, and necessary to understand racial inequities, then do quantitative studies need to be supplemented with qualitative studies? Or are there other ways for quantitative studies to capture the participants' experiential knowledge? Relating to this study and teaching data science for social justice, I think that it may be worth discussing these questions with the class. It may also be possible to discuss the role of experiential knowledge across the entire statistical investigation cycle (similar to how we discussed race and racism through the entire statistical investigation cycle).

Hesitancy to Talk about Race and Racism in a Mathematics Class

The third point in this discussion is about students being hesitant to talk about race and racism in the TE. This point is guided by some of the conversations I had with students in small groups or outside of formal class time as well as notes that I took immediately after those interactions (shared with consent). In particular, I talked to three students who expressed that they were seeking permission to talk about race and racism in the course. All three students had mentioned that they had learned about educational equity and Critical Race Theory in a previous course, but did not know how much of that knowledge they could bring up in a course that was listed under the mathematics and statistics department. All students were asked the following questions in the pre- and post-survey and responded strongly agree to both of them:

> **Question 1 (non-math college class):** Please indicate how much you agree or disagree with the following statements: - As a student in a non-math college class, I am comfortable discussing educational equity issues.
> **Question 2 (math college class):** Please indicate how much you agree or disagree with the following statements: - As a student in a math college class, I am comfortable discussing educational equity issues.

Notably, these questions asked about educational equity issues, not necessarily race and racism in education. Nonetheless, the questions might have helped provide an insight into students' comfort level with talking about race and racism in the context of education. Future iterations of this survey might consider rewording the question or including another set of questions that specifies "I am comfortable discussing race and racism."

I first noticed these tensions in a breakout room with Jacky and Caden during Class 4. Both students mentioned that they appreciated that we were talking about race and racism in a mathematics class, but were not sure about how we would engage with those conversations. I related to the students because, as a student, I also did not know what it was like to talk about race and racism in the context of mathematics, statistics, or data science.

"Angry BIPOC" Stereotype and Positioning

I followed up with Caden during Class 14 (the second to last week of the course). In a contact summary form, I wrote that Caden

> really liked that we actually talked about race and racism throughout the entire course... **They were very comfortable talking about this and is involved in activism-related work outside of school**... Their homework is also was always fire. I wish that they would've participated more during whole class conversations, but **they said that they didn't want to be portrayed as an angry [BIPOC student] and didn't want to educate others**.

Central to Caden's response was their identity, particularly as a BIPOC student.

There were three points that resonated with me. First, Caden had expressed that they were involved in activism work outside of their classes. This may imply that they felt comfortable talking about social justice topics. This is reflected in their response to the two survey questions, where they reported that they strongly agreed with feeling comfortable talking about educational equity in mathematics and non-mathematics classrooms. From an instructor perspective, someone might think of how Caden could have drawn from their professional and experiential knowledge to contribute to the classroom community and dialogue.

Second, Caden raises tensions about what their role (especially as a BIPOC student) was in these conversations. While they may have been able to contribute a lot to the classroom community and dialogue, Caden notes that they did not want to be portrayed as the "angry BIPOC student," a feeling I resonate with as well since this is a stereotype that I have often been labeled. This stereotype positions BIPOC people, especially Black people, as hostile, aggressive, or illogical. Caden noted that I did not put them in that position and they appreciated how I made them feel comfortable, but they have had previous experiences that were negative and led to them being called an "angry BIPOC student." These stereotypes were heightened during the time and political climate of this study, specifically among the Black Lives Matter movement. It is possible that Caden was cautious of these stereotypes and, as a result, was cautious about how they participated in the class because they were worried about how they would be positioned.

The third point was about the responsibility and cultural taxation that we place on students when asking them to talk about race and racism. As mentioned above, Caden may have been comfortable talking about race and racism, could draw from their experiential knowledge as a BIPOC student, and had experience with activism work. However, that does not mean that BIPOC students have a responsibility to educate others. From my experiences, I think of the traumas that are resurfaced when discussing how my experiences as a student were racialized and being tokenized or essentialized.

Turning to Critical Race Theory and QuantCrit, researchers and course designers may aim to draw on students' experiential knowledge to understand social and racial injustices. However, Caden brings awareness to the positioning and responsibility of BIPOC students when discussing educational equity, race, and racism. In particular, Caden notes how their previous conversations with similar conversations were not always positive (e.g., may have led to them being called an "angry BIPOC student"). Further, they refer to the cultural taxation and responsibility that we place on people when discussing topics related to race and racism. Thus, I want to specify that leveraging experiential knowledge should be a practice to place expertise in the hands of students and invite them to share their experiences in order to shed light on social and racial injustices; the use of experiential knowledge should not be a burden that instructors place on BIPOC students for the sake of educating others.

Race and Racism in the Mathematics and Statistics Classes

There were two other students that expressed hesitancy to talk about race and racism in a course that was listed in a mathematics and statistics department: Jacky and Elenai. Both students identified as Women of Color and mentioned at different times throughout the study that they had learned about Critical Race Theory in a previous class and had talked about race and racism in educational contexts.

In a Class 3 breakout room when we introduced Critical Race Theory, Jacky mentioned that she had learned about Critical Race Theory and was comfortable talking about race and racism. I followed up with Jacky in a group project check in during Class 15. In a contact summary form, I wrote that

> Jacky sounds like she was hesitant to contribute to whole class discussions because she didn't know how much she could talk about race and racism. I think she said "**I wasn't sure if I was allowed to take it there in a math class,**" or something super similar to that. How was this different in other classes? Bill asked this before the class started too. **Like would she feel more comfortable if this was in an education class? Sounds like some situated cognition stuff.**

Similar to Caden, Jacky may have been someone that felt very comfortable talking about race and racism in contexts outside of the mathematics and statistics classes. However, Jacky noted that she did not know if she was "allowed to take it there (discussing race and racism) in a math class." One interpretation of this is that talking about race and racism is situated in the context, where students may be less likely to talk about race and racism in a class that is listed under the mathematics and statistics department when compared to other departments (e.g., education). It is also interesting that Jacky specifically said "allowed," possibly noting that there was a sense of permission that Jacky was seeking (by her peers, the instructor, or broader mathematics) to discuss race and racism in mathematics and statistics classes or related to her identity as a Woman of Color. In future iterations of this study, it may be worth gathering more evidence for why students may seek permission to discuss race and racism in mathematics and statistics classes.

Elenai may have provided further insight into why students may be hesitant to "take it there in a math class." Particularly, in the end of the post-interview after the task, Elenai said that

> I was pretty nervous to talk about all this stuff because **I didn't know how we were going to talk about racism in this class.** (6s pause) Like I thought that data is just data, but I like how you showed us **different ways that race (and racism) is in data**, yea. Like the video that you showed us about how the face app thing is racist, but like we can fix it if we get a more, a better diverse sample, **so I feel like I understand all of this better**….Are you going to teach another class? I feel like I want to do more of this because it makes more sense now

One interpretation of Elenai's statement is that she was hesitant to talk about race and racism in the context of statistics and data science because she thought that "data is just data," or that data is apolitical and neutral. That is, she may have not known or considered how race and racism was embedded throughout the statistical investigation cycle prior to this class. However, it is possible that, throughout this course, she gained an understanding of both statistics and data science as well as how race and racism are embedded in statistics and data science. Thus, it is possible that she may have felt like she was able to "understand all of this (statistics, data science, and their intersectionality with race and racism) better" by the end of the class because we had talked about specific examples (e.g., the algorithmic bias video that showed how the "face app thing is racist").

Turning to the broader mathematics education and teaching for social justice landscape, Jacky and Elenai's stories resonated with my own experiences as a student in a mathematics department. In particular, Jacky may have highlighted how there is a perceived dichotomy between social justice and mathematics, where race and racism are often not talked about in courses listed under mathematics departments. Elenai added that she did not know how we were going to talk about race and racism in this class, possibly because she had not experienced a mathematics class where race and racism were discussed or because she was unfamiliar with the content discussed in the course. Thus, Elenai and Jacky may counter the narrative that mathematics and social justice are mutually exclusive. Rather, Elenai and Jacky highlighted how students may be interested in talking about social justice in mathematics and statistics classes, but there is a need for a shift in the mathematics culture from one where there is a perceived dichotomy between mathematics and social justice to one where (a) race and racism are not considered taboo or separate from mathematics and (b) social justice is embedded in the curriculum (e.g., race and racism are discussed alongside the mathematics content) so that students can build explicit connections between the mathematics and social justice content.

Thinking about future research, one possible avenue for future iterations of this study would be to look at how students navigate statistics and data science spaces with intersectional identities. For example, Carlone and Johnson's (2007) present a model of science identity to make sense of the experiences of Women in Color in undergraduate and graduate science disciplines. They note that someone's identity consists of their performance, recognition, and competence. So, when students are worried about the high social cost of being wrong, it could be related to their perceived performance, recognition, and competence or how others perceive them.

In summary, this discussion raises further questions about teaching data science for social justice to pre-service teachers. In particular, a thoughtful analysis of inequities that is intended to identify social injustices may be an entry point into using data for praxis (i.e., the reflection component of praxis). However, this starting point may be expanded to also help us understand the social injustices (e.g., by situating them within the larger sociopolitical landscape) as well as provide action items to advance social justice (the action component of praxis). Furthermore, experiential knowledge should play a role throughout the entire statistical and data scientific process. For example, statisticians and data scientists may draw on their experiential knowledge to guide their research questions, identify data of interest, reflect on their positionality, and add meaning to the data. Finally, students noted a hesitation to talk about race and racism in mathematical contexts (including statistics and data science) for at least two reasons. First, we are challenging a dominant culture where mathematics classrooms are portrayed as apolitical and, as a result, we are creating our own culture around what it means to talk about, thinking about, and reflect on race and racism in mathematical contexts. Building this culture takes trust, community building, and time. Second, some hesitation may come from cultural taxation, especially for BIPOC students. This raises questions around whose responsibility is it to educate others about race and racism, what type of communities (and counterspaces) are needed to engage in conversations about race and racism, and who needs to be educated about race and racism?

Chapter 9: Conclusion

In this chapter, I summarizing the findings for each research question and subquestion. I also include possible limitations, especially given the virtual restrictions due to the COVID pandemic as well as avenues for future research.

Research Question 1: Design Features

The first research question was about the design features that were used in the teaching experiment as well as how they were implemented in the teaching experiment. In particular, the research question was:

Research Question 1: Design Features
a. What design features support students' understandings of race and racism in the context of statistics and data science?
b. How were the design features enacted in the curriculum?

There were six design features that were centered around creating opportunities for students to develop critical statistical and data scientific consciousness in statistics and data science classrooms, with a focus on pre-service mathematics teachers. The six design features and a brief description is shown below in Figure 9.1. The six design feathers were motivated by Freire's (1998) notion of praxis, where three were related to reflection (understanding the social, cultural, historical, and political understandings of social injustices) and three were related to action (individual or collective action taken to advance social justices). I also drew on QuantCrit (Castillo & Gillborn, 2022; Crawford et al., 2018; Covarrubias, 2011; Covarrubias et al., 2018; Garcia et al., 2018; Gillborn et al., 2018; Pérez Huber et al., 2018; Sablan, 2019) to foreground the centrality of race and racism in data, especially in our national context. Additionally, although the design features are presented across two categories (reflection and action), the design features are complementary (shown by the overlap of the green and blue lines) and occurred as part of a larger journey towards critical statistical and data scientific consciousness (shown by the three dots on the left and right).

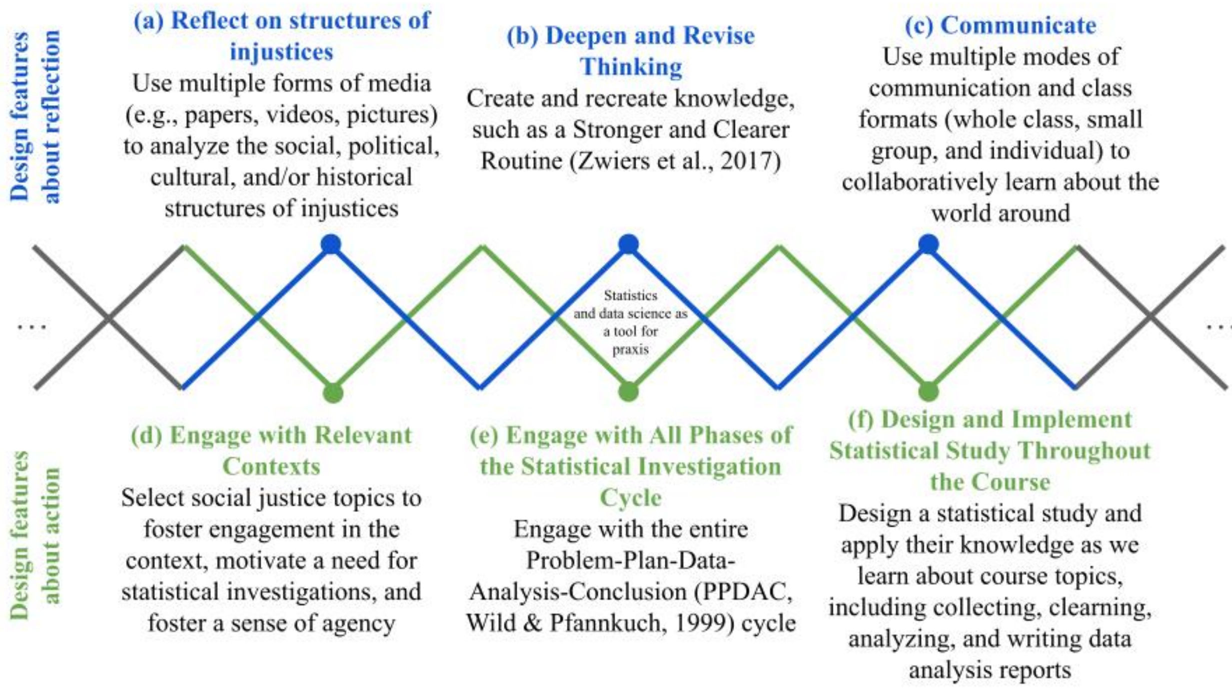


Figure 9.1: Design features about the opportunities for students that were incorporated into the curriculum

Design Features Related to Reflection

The three design features related to reflection were about providing students opportunities to: (a) reflect on structures of injustices, (b) deepen and revise thinking, and (c) communicate with each other. The first design feature (DF1: reflect on structures of injustices) aims to encourage students to account for larger systemic or structural causes of injustices. Part of the motivation for this design feature was to minimize gap-gazing or avoid reifying deficit narratives. This was enacted by using multiple forms of media (e.g., scholarly papers, videos, teacher-oriented papers) that helped students make connections between statistics and data science with social and racial justice.

The second design feature (DF2: deepen and revise thinking) builds on Freire's assumption that developing critical consciousness is a cyclical process that entails learning, relearning (learning something again, possibly clarifying or strengthening previous knowledge), and unlearning (modifying or editing knowledge) as needed. This is especially important in mathematics and social justice disciplines where there may be a perceived high social cost for sharing incomplete ideas that, as a result, may push students away from contributing to whole-class conversations. In doing so, a goal of this design feature is to provide a learning model that normalizes growth and involves changing ideas over time. This included opportunities for students to deepen and revise thinking at the individual level (e.g., asking students to react to a prompt, participate in a lesson related to the prompt, react to the same prompt again, and reflect on any changes), at the group level (e.g., editing survey questions used for the group project after instructor and peer feedback), and at the classroom level (e.g., revisiting classroom artifacts to strengthen or add meanings).

The third design feature (DF3: communicate) builds on Freire's notion of dialogue and Vygotsky's sociocultural theory that suggests language is a mediator for learning. In particular, this design feature aims to provide opportunities for students to engage in dialogue with themselves and the others to learn about their relationship with the world around them. At the individual level, this included using journals for students to reflect on their learning and growth, particularly about how statistics and data science can be used to advance social and racial justice. At the whole-class level, this included using Mathematical Language Routines (e.g., Collect and Display, Information Gap; Zwiers et al., 2017) to facilitate whole-class discussions as well as displaying anonymous contributions that may have encouraged more authentic student contributions.

Design Features Related to Action

The other three design features aim to provide students opportunities to advance social and racial justice using statistics and data science. For example, the fourth design feature (DF4: engage with relevant contexts) aims to provide students opportunities to engage with problem contexts that are: (a) relevant to them, (b) motivate a need for analyzing data, and (c) provide avenues for social change. Initially, I planned on using educational contexts throughout the course because I assumed that the majority of the students would find that relevant since most were pre-service mathematics teachers. However, after getting to know the students and seeing how they engaged with other contexts, I made the decision to focus on the facial recognition context.

This brings awareness to how what may be considered relevant to one student, group, class, or community may not be relevant to another. Thus, it is possible that every iteration of this study may need to consider different contexts. One possible consideration is using contexts that may have a clearer path for advancing social and racial justice. For example, it is likely easier to sample new data to decrease facial recognition misclassifications than it is to advance educational equity. Additionally, course designers may want to consider data that is easier to quantify. For instance, facial recognition misclassification is easily quantifiable whereas educational equity is more abstract.

The fifth design feature (DF5: engage with all phases of the statistical investigation cycle) aims to provide students opportunities to engage with the entire statistical investigation cycle. This shifts away from a traditional view of statistics as a field that focuses on applying formulas and algorithms to simplified datasets (Bargagliotti & Franklin, 2015; Franklin, 2013; Garfield & Ben-Zvi, 2008) that focus on the analysis phase to a holistic view of statistics that includes the Problem-Plan-Data-Analysis-Conclusion phases of the statistical investigation cycle (Wild & Pfannkuch, 1999). This motivated the sequencing of the course. For example, we started the course by talking about Critical Race Theory and QuantCrit (as the landscape for the context, problem phase), then moved into statistical questions (problem phase), sampling and randomization (plan and data phase), and study designs (plan and data phase) before moving into regression and other types of analysis (analysis phase) and project presentations (conclusion phase).

The last design feature (DF6: design and implement statistical study throughout the course) provides opportunities for students to concurrently apply what they learn in the course on their own study. The students selected their context (as long as it was related to social justice) and decided if they wanted to work on the project by themselves or in groups. In the course, I followed elements of Chance's (1997) guidelines for statistics course projects: (a) integrating the project into the course, (b) providing students with timely feedback, and (c) clear expectations and guidelines. In terms of integrating the project into the course, students worked on the project throughout the course as we discussed new topics. For example, they wrote an initial draft of their research questions after we learned about statistical questions, drafted survey questions and got instructor and peer feedback while we learned about survey questions, and drafted a methods chapter for their final data analysis report after the study design unit.

Future Work

First, I am interested in continuing to explore what problem contexts are most appropriate or beneficial for teaching statistics and data science for social justice. For example, as mentioned above, the educational equity problem context did not appear to be as engaging for the students as the facial recognition problem context. This may be partially attributed to how the facial recognition data had a clearer path towards advancing social and racial justice and because misclassification was less abstract than educational equity. A future experimental study could further explore this by teaching one course using the educational equity problem context and another course with the facial recognition problem context, then compare student learning across both courses (e.g., using analyses similar to Research Questions 2 to 4). Perhaps a simpler study would be designing a task-based interview with both contexts, randomly assigning a group of students to one of the two contexts, and analyzing how students engage with both contexts. These types of studies may help develop theoretical contributions about what social justice problem contexts are most relevant in statistics and data science courses.

A second possible avenue for future research is using these design features in an in-person course. In particular, the course in this dissertation was taught virtually, which provided different challenges and advantages than teaching the class in person. For example, from my experiences teaching both in-person and virtually, facilitating whole-class conversations was more challenging in a virtual setting than an in-person setting. On the other hand, the chat and emoji reactions opened an avenue for students to engage with each other that may not be possible in an in-person setting. Thus, while the design features were designed for both an in-person and virtual setting, I wonder what modifications would be needed for an in-person setting.

Finally, a third possible avenue for future research may be related to the sequencing of lessons. In particular, I noticed that students were more engaged in conversations about race and racism after they had learned the traditional statistics and data science content. That is, when the social and racial justice content was taught as a layer that was placed on top of the statistics and data science content. For example, I noticed that students were able to identify how the machine learning process was racialized after we talked about sampling, randomization, and training and testing datasets. Admittedly, as a course designer, I struggled with presenting the social and racial justice content as a layer that was added to the statistics and data science content because I feared that it would perpetuate a dichotomy between both disciplines. However, further research may help explain why this phenomena occurred or how the social and racial justice content can be more embedded with the statistics and data science content. For example, this may include some form of experimental design where one class learns the social justice and racial justice content before the statistics and data science content, a second class is the opposite, and a third class presents the material in a more embedded way. I could collect similar data to this dissertation (e.g., pre-post assessments, interviews, classroom data), but the student learning and how students conceptualized race and racism in the context of statistics and data science across the different classrooms.

Research Question 2: Statistical and Data Scientific Content Knowledge

The second research question was about students' statistical and data scientific content knowledge as measured by curriculum-aligned pre- and post-assessments for the study design and regression units. The research question was:

> **Research Question 2:** Statistical and Data Scientific Content Knowledge
> a. What was the effect of the teaching experiment (TE) on statistical content knowledge as measured by the student response patterns on curriculum-aligned assessments?
> b. How did the response patterns by question type (e.g., conceptual or procedural, study design and regression) vary across the TE?

This research question included a quantitative analysis for the difference in centers (paired t-test, effect size, and normalized gains for a difference in means, paired Wilcoxon signed-rank test for a difference in medians) for both the study design and regression unit. Appropriate modifications for the small sample size were taken when appropriate and available. The purpose of using the different statistical measures was to provide multiple sources of evidence rather than relying on only one statistical measure.

In terms of comparing across both units, the regression unit had a larger increase (average of about 53% in the pre-assessment, about 75% in the post-assessment) than the study design unit (average of about 67% in the pre-assessment, about 79% in the post-assessment). Additionally, both units had a larger increase in the free response questions than the multiple-choice questions. This may be partially attributed to the students having relatively high scores on the multiple-choice questions in the pre-assessment that may have resulted in a ceiling effect. For example, in the study design unit, the average for the multiple-choice questions in the pre-assessment was 73% and for the post-assessment it was 83%. The average for the free response questions was 56% in the pre-assessment and 72% in the post-assessment.

As expected, students also provided more details in the free response questions from the post-assessment when compared to the pre-assessment in both units. For example, in the study design unit, students provided more details about how they would implement randomization in a study (e.g., flipping a coin, assigning to groups) and noted the importance of randomization and experimental designs for inferring cause-and-effect relationships. Similarly, in the regression unit, students were more likely to describe correlations using all four features of a correlation (strength, linearity, direction, and context) and used language like "residuals" that may be directly tied to some of the classroom discussions.

Limitations and Future Work

One of the main limitations of this research question is the small sample size (14 students total). As a result, the findings may not be generalizable to other settings. Nonetheless, the statistical methods (mainly, using different statistical measures for difference in centers) may be useful for future studies. Future iterations of this study may consider having a larger class size or multiple classes to sample from a larger population.

Additionally, in the original design of this study, I intended to compare the pre- and post-assessments across two settings: (a) the course presented in this study that used a social justice-oriented curriculum, and (b) a statistics class that did not use a social justice-oriented curriculum. This analysis included propensity score matching that would allow for comparisons across treatments in observational studies by pairing students across the settings given relevant data (e.g., scores on previous statistics courses, number of statistics courses taken prior to this course, comfort with computer programming, other demographics). Although this is a positivist approach to research, I was motivated by my own experiences using social justice-oriented pedagogy, particularly instances where I was told that students did not learn as much of the content in a social justice-oriented class than a class that is not social justice-oriented. One possible avenue for future research could be a form of this study to show that student learning occurs in both settings, either experimental where students are assigned to one of the two classes or observational where students select the class.

However, pushing back on the critique of teaching for social justice that suggests that students do not learn as much of the content, it is also possible to reframe what is considered content in statistics and social justice. In particular, if we view statistics and data science as the entire PPDAC statistical investigation cycle (Wild and Pfannkuch, 1999) and if the problem context is central to statistics (Burrill & Biehler, 2011; Chance, 2002; Franklin et al., 2007; Lee & Tran, 2015; Visnovska & Cobb, 2019), then the problem context is also considered content. Thus, under this perspective, knowledge about social and racial justice in statistics and data science courses is, or should be, considered content and should be part of the curriculum.

Research Question 3: Statistical and Data Scientific Practices

The third research question was about documenting how participants engaged with the statistical investigation cycle and the practices that emerged in the interview, focusing on the intersection between statistical and data scientific practices with critical practices. The research question was:

> **Research Question 3:** Statistical and Data Scientific Practices
>
> a. How do students' engagement with the statistical investigation cycle evolve through the course of a TE that uses a social justice-oriented approach to teaching statistics?
>
> b. How do students' statistical practices evolve through the course of a TE that uses a social justice-oriented approach to teaching statistics?

Data came from pre- and post-task-based interviews with four students. Students used Common Online Data Analysis Platform (CODAP) to explore data about local schools, focusing on identifying three schools to visit and providing an explanation for how and why they chose those schools.

There were three practices that were evident in the post-interview but not the pre-interview: (a) political context, (b) assets instead of deficits, (c) and action items. The political context practice was about considering the sociopolitical and/or racialized nature of data (Gillborn et al., 2018; Weiland et al., 2017), focusing on systemic structures of social inequities. The practice about focusing on assets instead of deficits was about highlighting the strengths of communities or schools. Finally, the practices about action items occurred when students provide avenues to advance social and/or racial justice.

Furthermore, only one student engaged in a form of an exploratory data analysis in the pre-interview whereas all students engaged with a form of an exploratory data analysis in the post-interview. Additionally, two students had multiple questions that built on each other, similar to a snowball effect where new questions emerge as they analyze the data.

In terms of the role of race and racism in the PPDAC cycle, most students focused on identifying social injustices in the pre-interviews. However, in the post-interviews, all four of the students suggested that they wanted to avoid painting a static picture of educational equities. For example, they stated that educational equity cannot only be completely captured by standardized assessments and, instead, considered ways to account for the larger social, cultural, political, and historical contexts of educational equity in this statistical investigation (e.g., using data about funding, asking for new data, providing recommendations for future studies). Thus, it is possible that students showed a shift in their critical statistical and data scientific consciousness from one that focused on identifying social injustices in the pre-interview to one where they engaged with more dimensions of praxis in the post-interview that included stating the structures of injustices and provided avenues to address those social injustices.

Finally, there were also some changes in the role of experiential knowledge across the pre- and post-interviews. For example, experiential knowledge mainly came up in the pre-interviews when students were deciding which data to select from the analysis, where they often focused on data that is traditionally associated with equity (e.g., race or ethnicity, free or reduced priced meals). While this was common in the post-interview, students also drew on their experiential knowledge about the context (educational equity) to situated the context in a larger sociopolitical context in the post-interview. For example, students noted how standardized assessments should not be the only measure of educational equity and provide examples of what else could be collected (e.g., students' disposition towards mathematics).

Limitations and Future Work

One possible limitation of the task-based interviews was that the data that was discussed during the interviews was collected and provided to students. This may have limited which phases of the PPDAC cycle that students engaged with, particularly in terms of the planning and data phases since the data was already collected. Additionally, drawing on my own experiences, providing data removed a significant part of data science: collecting data from multiple sources, merging data, and cleaning data. Although I did manipulate the data (in the course) to include some of these common parts (e.g., making binary data 0 or 1 instead of a character to see if they treated the variables as categorical or numeric, adding typos), for a future iteration, it may have been worthwhile to have students search for, merge, and clean the data themselves.

Related to this point, future iterations of this study may consider analyzing the data analysis reports that students submitted as part of their final course project. In these reports, students were expected to include any challenges and approaches to collecting and cleaning data. For example, in one of the groups, all three students used the same survey but distributed them using different platforms (e.g., Google Survey and Qualtrics). As a result, they mentioned how they had to reformat their excel sheets and merge them in R. Additionally, other groups discussed how some of their survey questions were not clear (e.g., asked for distance from a place but responses included blocks, miles, and meters) and the decisions they made to clean and organize the data (e.g., convert meters to miles, remove responses that were in blocks).

Students also noted how they considered race and racism when engaging with the statistical investigation cycle in their final data analysis report. In some cases, their final data analysis report provided more or additional examples of how they considered race and racism (a possible limitation of the task-based interview) than the interviews. For example, some students provided more details about the sociopolitical context of the social justice topic that they chose (e.g., drawing on their own experiential knowledge, referencing research that they had read about in other classes), likely because they were able to choose a topic that they were interested in instead of being assigned a problem context (like in the task-based interview). Similarly, they discussed how they created questions for their survey, the different iterations of their survey, and the intentionality behind their survey questions (e.g., asking participants "What is your gender identity?" instead of asking them to select from a binary list).

Finally, future iterations of this study would benefit from double coding. This is especially important since Gould et al. (2017) used a similar analysis but had low interrater reliability (the extent to which two or more coders agreed). Similar to Gould et al. (2017), I would recommend having two researchers code the interviews individually, then have a third researcher make decisions when there is a disagreement between the other two researchers.

Research Question 4: Focusing Phenomenon

The fourth and final research question drew on elements from Lobato et al.'s (2013) focusing phenomena framework to analyze how discourse practices, tasks, the nature of mathematical activities, and other classroom data may have influenced how students attended to race and racism in the context of statistics and data science. In particular, the research question was:

Research Question 4: Focusing Phenomenon

a. How do elements of the TE contribute to the students' understanding of race and racism in the context of statistics and data science?

There were four phases to the analysis: (a) identifying Centers of Focus, (b) identifying focusing interactions, (c) describing the features of the task, including affordances and constraints of the task, and (d) describing the ways in which classroom participation is organized and regulated by classroom norms. I also add a fifth analytical pass that makes connections with the design features.

In this dissertation, I presented Centers of Focus that emerged during the class and appeared with more than one student. This resulted in the Centers of Focus related to: (a) Context - Historical and Political, (b) Algorithmic Bias, (c) Agency, and (d) Question Framing. The Center of Focus related to *Context - Historical and Political* describes statements that suggest that the social, cultural, historical, or political contexts may help understand or create biases at any point of the statistical investigation cycle. *Algorithmic Bias* included statements that refer to how computers may encode systemic biases (racism, sexism, other forms of discrimination) that create differentiated outcomes. *Agency* was about showing possible ways to mitigate algorithmic bias. Finally, *Question Framing* was about statements that suggested that the statisticians' or data scientists' research questions (e.g., anti-deficit or deficit) may influence their engagement with the data and lead to potential biases.

Notably, the majority of the students viewed data as neutral at the time when the first prompt was collected (six out of the 14 coded responses). However, almost all of the students viewed data as not neutral by the end of the class (ten out of the 14 coded responses) and provided specific examples of how data may not be neutral. For example, students referenced the class session where we talked about anti-deficit research questions to illustrate that researchers may have implicit biases that are manifested in their research questions and carried out in their analysis. Similarly, students referred to a video that talked about algorithmic bias (Fong, 2021) to provide examples of racial biases that may be encoded through sampling.

Limitations and Future Work

It is also important to note two of the modifications for this study included possible limitations and related avenues for future work: (a) using student work to identify potential Centers of Focus, and (b) focusing on discourse practices that highlighted (operated visibly on external phenomena). First, I used student responses to four question prompts about data neutrality to help identify potential Centers of Focus (instead of using interviews or classroom interactions). Part of the motivation for this approach was because I was not able to record small group interactions (i.e., breakout rooms) since the course was taught over Zoom. Additionally, I only had four student interviews whereas most of the students gave consent to use and analyze classroom data, including the four prompts about data neutrality. Furthermore, the student responses to the four prompts sometimes revealed more evidence about student learning as it relates to this course than the task-based interviews did. For example, students often referred to specific moments from the class (e.g., debriefing an activity, videos, readings) that were easily coordinated with their learning. However, future iterations of this study may consider coordinating findings from the task-based interview aspects of the classroom data. In particular, it may be worth identifying moments in the classroom data that may have guided students' attention to the three practices from Chapter 6 (political context, assets instead of deficits, and action items).

The second major modification was focusing on discourse practices that highlighted student contributions. As mentioned above, there were not many instances of quantitative dialogue or renaming (the two other codes used by Lobato et al., 2013). This may be because I searched for quantitative dialogue instead of statistical or data scientific dialogue. Future iterations of this study might consider keeping track of statistical and data scientific dialogue, roughly defined as talk related to variation, the context, data wrangling (i.e., cleaning data), or other attributes that differentiate statistics and data science from mathematics. Furthermore, I extended practices that highlighted student contributions by coordinating them with the Mathematical Language Routines that were present in that interaction. The purpose of this was to add more details about how and why highlighting took place (e.g., using an information gap that guided students' attention to features that were then collected and displayed to the entire class).

Finally, I provide another avenue for potential future work. The analysis in this chapter had a backwards approach in that I began by identifying Centers of Focus, then looked at classroom data prior to when the first Center of Focus appeared. In future studies, it may be worth continuing to track those Centers of Focus after they are first identified as well. For example, once the Center of Focus related to algorithmic bias was identified, how did it evolve over time? What moments in the classroom data helped other students notice that Center of Focus in future classes? What role did the Center of Focus take in the class (e.g., was it taken up as shared knowledge, was it used to justify arguments, or was there more information needed before the Center of Focus was taken up by other students)?

Concluding Remarks

The widespread availability of data and the emerging field of data science has brought attention to how we teach statistics and data science (Bargagliotti et al., 2020; Franklin et al., 2007) and prepare the next generation of statistics and data science teachers (Franklin et al., 2013). In fact, I would say that it is almost impossible to make it through a day without encountering data or data-informed decisions. To realize the full potential of statistics and data science, I agree with other researchers that have also called for a need to use data to guide conversations about race and racism in data science courses (Philip et al., 2016, 2017), especially given the recent coverage of the Black Lives Matter movement, climate change, and public health.

As a result, in this dissertation, I designed and analyzed data from a course that taught data science for social and racial justice to PSMTs. I drew on elements of Teaching Mathematics for Social Justice (TMSJ; Gutstein, 2006), Quantitative Critical Race Theory (QuantCrit; Castillo & Gillborn, 2022; Crawford et al., 2018; Covarrubias et al., 2018; Garcia et al., 2018; Gillborn et al., 2018), and research about Habits of Mind (Cuoco et al., 1996) to study the potential of using a social justice-oriented approach to teaching data science for preservice mathematics teachers. The study described a credit-bearing course taken by 14 students that was taught virtually during the Summer 2021 term at a four-year public university in the US-Mexico borderlands of Southern California. Data included pre- and post-curriculum aligned assessments, pre- and post-task-based interviews, and classroom data (e.g., student work, whole-class recordings, field notes).

There were four research questions used to highlight the intersectionality between statistics, data science, and social and racial justice. First, there was a qualitative description of the features used to design the course centered around Freire's (1998) notion of critical consciousness and praxis with illustrations of how the design features were enacted in the course. Second, there was a quantitative analysis of pre- and post-curriculum- aligned assessments that aimed to measure the students' statistical and data scientific content knowledge. Third, there was a qualitative analysis of pre- and post-task-based interviews that aimed to capture students' critical statistical and data scientific practices, providing illustrations of what engaging with statistics and data science in a critical way may entail. Finally, elements of a focusing phenomenon framework were used to coordinate how aspects of the classroom environment (e.g., design features, tasks, tools, and the teacher) directed students' attention towards understandings of race and racism in the context of data science.

Turning to my personal motivation and personality, I hope that this dissertation helps illustrate some of the intersectionality between race and racism with statistics and data science, where statistics and data science can be used to identify, understand, and challenge social and racial injustices. More importantly, I hope that all students (especially BIPOC students) are able to see that they belong and, in fact, are needed in statistics and data science.

References

Anderson, G. L. (2002). Reflecting on research for doctoral students in education. *Educational Researcher*, 31(7), 22–25.

Anderson, G. L., & Herr, K. (1999). The new paradigm wars: Is there room for rigorous practitioner knowledge in schools and universities? *Educational Researcher*, 28(5), 12–40.

Anzaldúa, G. (2012). *Borderlands: La Frontera: The New Mestiza*. 25th anniversary, fourth edition. San Francisco: Aunt Lute Books.

Anzaldúa, G. (2015). *Light in the Dark/Luz en lo Oscuro*. Duke University Press; JSTOR.

Apple, Michael W. (1996). Power, meaning and identity: Critical sociology of education in the United States. *British Journal of Sociology of Education*, 17(2), 125–144.

Austin, P. C. (2010). Statistical Criteria for Selecting the Optimal Number of Untreated Subjects Matched to Each Treated Subject When Using Many-to-One Matching on the Propensity Score. *American Journal of Epidemiology*, 172(9), 1092–1097.

Austin, P. C., & Stuart, E. A. (2015). Optimal full matching for survival outcomes: a method that merits more widespread use. *Statistics in medicine*, *34*(30), 3949-3967.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special. *Journal of Teacher Education*, 59(5), 389–407.

Bargagliotti, A., & Franklin, C. (2015). The statistical education of teachers: Preparing teachers to teach statistics. *Chance*, 28(3), 19–27.

Bargagliotti, A., Franklin, C., Arnold, P., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). Pre-K-12 guidelines for assessment and instruction in statistics education II (GAISE II) (Second edition). American Statistical Association.

Bartell, Tonya G, & Meyer, M. R. (2008). Connecting Research to Teaching: Addressing the Equity Principle in the Mathematics Classroom. *The Mathematics Teacher*, 101(8), 604–608.

Bartell, Tonya Gau. (2011). Caring, race, culture, and power: A research synthesis toward supporting mathematics teachers in caring with awareness. *Journal of Urban Mathematics Education*, 4(1), 50–74.

Bartell, Tonya Gau. (2013). Learning to Teach Mathematics for Social Justice: Negotiating Social Justice and Mathematical Goals. *Journal for Research in Mathematics Education*, 44(1), 129–163.

Barton, A. C., Tan, E., & Rivet, A. (2008). Creating Hybrid Spaces for Engaging School Science Among Urban Middle School Girls. *American Educational Research Journal,* 45(1), 68–103.

Batanero, C., Burrill, G., & Reading, C. (2011). Overview: challenges for teaching statistics in school mathematics and preparing mathematics teachers. *Teaching statistics in school mathematics-challenges for teaching and teacher education: a joint ICMI/IASE study*, 407-418.

Benjamin, R. (2019). Race After Technology: Abolitionist Tools for the New Jim Code. *Social Forces*.

Bernard, H. R. (1988). *Research methods in cultural anthropology.* Sage Publications

Berry, R. Q., Conway IV, B. M., Lawler, B. R., & Staley, J. W. (2020). *High School Mathematics Lessons to Explore, Understand, and Respond to Social Injustice*. Corwin Press.

Bond, G., & Chernoff, E. (2015). Mathematics and Social Justice: A Symbiotic Pedagogy. Journal of Urban Mathematics Education, 8(1), 24–30.

Brantlinger, A. (2013). Between politics and equations: Teaching critical mathematics in a remedial secondary classroom. *American Educational Research Journal*, 50(5), 1050–1080.

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42.

Buolamwini, J. A. (2017). *Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers* (Doctoral dissertation, Massachusetts Institute of Technology).

Burrill, G., & Biehler, R. (2011). Fundamental Statistical Ideas in the School Curriculum and in Training Teachers. In C. Batanero, G. Burrill, & C. Reading (Eds.), *Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education* (Vol. 14, pp. 57–69). Springer Netherlands.

Blake, C. L., & Merz, C. J. (1998). *UCI repository of machine learning databases*. University of California. [www http://www.ics.uci.edu/~mlearn/MLRepository.html]

California State Board of Education (2022). *Mathematics framework for California public schools, kindergarten through grade twelve*. Sacramento, CA: California Department of Education.

Casey, S., Ross, A., Maddox, S., & Wilson, M. (2019). MODULE(S2): Statistics for secondary mathematics teaching. *Mathematics of doing, understanding, learning, and educating for secondary schools.* Washington, D.C.: APLU.

Chance, B. L. (1997). Experiences with Authentic Assessment Techniques in an Introductory Statistics Course. *Journal of Statistics Education*, 5(3).

Chance, B. L. (2002). Components of Statistical Thinking and Implications for Instruction and Assessment. *Journal of Statistics Education*, 10(3), 3.

Chance, B., Wong, J., & Tintle, N. (2016). Student Performance in Curricula Centered on Simulation-Based Inference: A Preliminary Report. *Journal of Statistics Education*, 24(3), 114–126.

Civil, M. (2014). Why Should Mathematics Educators Learn from and about Latina/o Students' In-School and Out-of-School Experiences? *Journal of Urban Mathematics Education*, *7*(2).

Civil, M. (2016). STEM learning research through a funds of knowledge lens. *Cultural Studies of Science Education*, *11*(1), 41-59.

Civil, M., & Planas, N. (2010). Latino/a immigrant parents' voices in mathematics education. *Immigration, diversity, and education*, 130-150.

Clarke, S. N. (2015). The right to speak. In L. B. Resnick, C. S. C. Asterhan, & S. Clarke (Eds.), *Socializing intelligence through academic talk and dialogue* (pp. 167–180). Washington, DC: AERA.

Cobb, G. W., & Moore, D. S. (1997). Mathematics, Statistics, and Teaching. *The American Mathematical Monthly*, 104(9), 801–823.

Cobb, P., McClain, K., & Gravemeijer, K. (2003). Learning about statistical covariation. *Cognition and Instruction,* 21(1), 1–78.

Cobb, P., Confrey, J., diSessa, A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, 32(1), 9- 13

Cochran, W. G., & Rubin, D. B. (1973). Controlling Bias in Observational Studies: A Review. *Sankhyā*: *The Indian Journal of Statistics*, Series A (1961-2002), 35(4), 417–446. JSTOR.

Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Taylor & Francis.

Corbin, J. M., & Strauss, A. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology,* 13(1), 3–21.

Crawford, C., Demack, S., Gillborn, D., & Warmington, P. (2018). Quants and crits: Using numbers for social justice (Or, How Not to be lied to with statistics). In J. DeCuir-Gunby, T. Chapman, & P. Schutz (Eds.), *Understanding critical race research methods and methodologies: Lessons from the field* (pp. 219–227). Routledge.

Crenshaw, K., Gotanda, N., Peller, G., & Thomas, K. (1995). *Critical race theory. The Key Writings That Formed the Movement*. New York, 276–291.

Crenshaw, K. W. (1988). Race, reform, and retrenchment: Transformation and legitimation in antidiscrimination law. *Harvard Law Review,* 1331–1387.

Crocker, L. M., & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. Holt, Rinehart, and Winston.

Cuoco, A., Paul Goldenberg, E., & Mark, J. (1996). Habits of mind: An organizing principle for mathematics curricula. *The Journal of Mathematical Behavior*, 15(4), 375–402.

D'Ambrosio, U., & D'Ambrosio, B. S. (2013). The role of ethnomathematics in curricular leadership in mathematics education. *Journal of Mathematics Education at Teachers College*, *4*(1).

DelMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students’ conceptual understanding after a first course in statistics. *Statistics Education Research Journal,* 6, 28–58.

Design-Based Research Collective. (2003). Design-Based Research: An Emerging Paradigm for Educational Inquiry. *Educational Researcher,* 32(1), 5–8.

Du Bois, W. E. B. (1899). *The Philadelphia Negro: A Social Study*. University of Pennsylvania Press.

Ernest, P. (2002). Empowerment in mathematics education. *Philosophy of mathematics education journal*, *15*(1), 1-16.

Esmonde, I. (2014). “Nobody’s Rich and Nobody’s Poor … It Sounds Good, but It’s Actually Not”: Affluent Students Learning Mathematics and Social Justice. *Journal of the Learning Sciences*, 23(3), 348–391.

Fan, X., & Nowell, D. L. (2011). Using propensity score matching in educational research. *Gifted Child Quarterly*, *55*(1), 74-79.

Fernández, J. S., & Gamero, A. M. (2018). Latinx/Chicanx students on the path to conocimiento: Critical reflexivity journals as tools for healing and resistance in the Trump Era. *Association of Mexican American Educators Journal*, 12(3), 13-32.

Fong, J. (2021). Are we automating racism? [Video]. Vox. https://www.vox.com/videos/2021/3/31/22348722/ai-bias-racial-machine-learning

Forman, E. A. (1996). Learning mathematics as participation in classroom practice: Implications of sociocultural theory for educational reform. *Theories of Mathematical Learning*, *115*, 130.

Frankenstein, M. (1983). Critical Mathematics Education: An Application of Paulo Freire's Epistemology. *Journal of Education*, 165(4), 315–339.

Frankenstein, M. (1989). *Relearning mathematics: A different third R--radical math(s)*. Free Association Books.

Frankenstein, M. (1994). Understanding the politics of mathematical knowledge as an integral part of becoming critically numerate. *Radical Statistics*, 56, 22–40.

Frankenstein, M. (2001). Reading the world with math: Goals for a critical mathematical literacy curriculum. *The Australian Association of Mathematics Teachers* Inc., 53.

Franklin, C. (2013). Common core state standards and the future of teacher preparation in statistics. *The Mathematics Educator*, *22*(2).

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: A pre-K-12 curriculum framework.* American Statistical Association.

Franklin, C., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M., Schaeffer, R., Friel, S., Godbold, L., & Hartlaub, B. (2005). *A curriculum framework for preK-12 statistics education*. American Statistical Association Board of Directors for Endorsement.

Freire, P. (1998). *Pedagogy of the Oppressed*. Continuum.

Freire, P., & Macedo, D. (1987). *Literacy: Reading the word and the world.* Praeger.

Garcia, N. M., López, N., & Vélez, V. N. (2018). QuantCrit: Rectifying quantitative methods through critical race theory. *Race Ethnicity and Education,* 21(2), 149–157.

Garfield, J., & Ben-Zvi, D. (2008). Preparing school teachers to develop students' statistical reasoning. Teaching Statistics in School Mathematics-Challenges for Teaching and Teacher Education: *A Joint ICMI/IASE Study: The 18th ICMI Study*, 299–310.

Garii, B., & Appova, A. (2013). Crossing the great divide: Teacher candidates, mathematics, and social justice. *Teaching and Teacher Education*, 34, 198–213.

Garii, B., & Rule, A. C. (2009). Integrating social justice with mathematics and science: An analysis of student teacher lessons. *Teaching and Teacher Education*, 25(3), 490–499.

Gillborn, D., Warmington, P., & Demack, S. (2018). QuantCrit: Education, policy, 'Big Data' and principles for a critical race theory of statistics. *Race Ethnicity and Education*, 21(2), 158–179.

Giroux, H. A. (2001). *Theory and Resistance in Education: Towards a Pedagogy for the Opposition*. Greenwood Publishing Group.

Goldenberg, E. P. (1996). “Habits of Mind” as an Organizer for the Curriculum. *Journal of Education*, 178(1), 13–34.

Gonzalez, L. (2009). Teaching mathematics for social justice: Reflections on a community of practice for urban high school mathematics teachers. *Journal of Urban Mathematics* Education, 2(1), 22-51.

Gould, R., Machado, S., Ong, C., Johnson, T., Molyneux, J., Nolen, S., ... & Zanontian, L. (2016). Teaching data science to secondary students: The mobilize introduction to data science curriculum. *Iase-Web. Org*.

Gould, R. (2017). Data literacy is statistical literacy. *Statistics Education Research Journal,* 15(1), 22–25.

Gould, R., Bargagliotti, A., & Johnson, T. (2017). An analysis of secondary teachers’ reasoning with participatory sensing data. *Statistics Education Research Journal,* 15(2), 305–334.

Griswold, D., Rack, M., Hooper, J., Hayes, J., Lindemer, P. (2018). *Statistics*. United States: CPM Educational Program.

Grossman, P. L. (1990). *The Making of a Teacher: Teacher Knowledge and Teacher Education.* Teachers College Press, Teachers College, Columbia University.

Groth, R. E. (2007). Research commentary: Toward a conceptualization of statistical knowledge for teaching. *Journal for Research in Mathematics Education*, 38(5), 427–437.

Groth, R. E. (2012). The Role of Writing Prompts in a Statistical Knowledge for Teaching Course. *Mathematics Teacher Educator*, 1(1), 23–40.

Groth, R. E. (2013). A day in the life of statistical knowledge for teaching course. *Teaching Statistics,* 35(1), 37–42.

Groth, R. E. (2014). Using Work Samples from the National Assessment of Educational Progress (NAEP) to Design Tasks that Assess Statistical Knowledge for Teaching. *Journal of Statistics Education*, 22(3).

Gutiérrez, R. (2008). A" gap-gazing" fetish in mathematics education? Problematizing research on the achievement gap. *Journal for Research in Mathematics Education,* 357–364.

Gutiérrez, R. (2009). Embracing the inherent tensions in teaching mathematics from an equity stance. *Democracy and Education*, 18(3), 9–16.

Gutiérrez, R. (2013). The sociopolitical turn in mathematics education. *Journal for Research in Mathematics Education*, 44(1), 37–68.

Gutiérrez, R. (2016). Strategies for creative insubordination in mathematics teaching. *Teaching for Excellence and Equity in Mathematics*, 7(1), 52–60

Gutiérrez, R. (2017). Political conocimiento for teaching mathematics: Why teachers need it and how to develop it. In S. Kastberg, A. M. Tyminski, A. Lischka, & W. Sanchez (Eds.), *Building Support for Scholarly Practices in Mathematics Methods* (pp. 11–38). Charlotte, NC: Information Age

Gutstein, E. (2006). *Reading and writing the world with mathematics: Toward a pedagogy for social justice.* Taylor & Francis.

Gutstein, E. (2012). Connecting community, critical, and classical knowledge in teaching mathematics for social justice. In *Alternative forms of knowing (in) mathematics* (pp. 299–311). Brill Sense.

Gutstein, E. (2012). Mathematics as a weapon in the struggle. In *Opening the Cage* (pp. 23–48). Springer.

Haines, B. (2015). Conceptualizing a framework for advanced placement statistics teaching knowledge. *Journal of Statistics Education*, 23(3).

Hake, R. (2002). Relationship of Individual Student Normalized Learning Gains in Mechanics with Gender, High-School Physics, and Pretest Scores on Mathematics and Spatial Visualization.

Hedges, L. V. (1981). Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *Journal of Educational Statistics*, 6(2), 107–128.

Hill, H. C., Ball, D. L., & Schilling, S. G. (2008). Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of students. *Journal for Research in Mathematics Education*, 39(4), 372–400.

Ho DE, Imai K, King G, Stuart EA (2011). "MatchIt: Nonparametric Preprocessing for Parametric Causal Inference." *Journal of Statistical Software*, **42**(8), 1–28.

Jacobbe, T., Case, C., Whitaker, D., & Foti, S. (2014). Establishing the validity of the locus assessments through an evidenced-centered design approach. In *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9), Flagstaff, Arizona, USA* (pp. 1-6)

Jong, C., & Jackson, C. (2016). Teaching mathematics for social justice: Examining preservice teachers' conceptions. *Journal of Mathematics Education at Teachers College*, 7(1).

Judd, T., Durand, F., & Torralba, A. (2012). A benchmark of computational models of saliency to predict human fixations.

Kanim, S., & Cid, X. C. (2020). Demographics of physics education research. *Physical Review Physics Education Research*, 16(2), 020106.

Kantayya, S. (Director). (2020). *Coded Bias* [Film]. 7th Empire Media.

Kokka, K. (2015). Addressing Dilemmas of Social Justice Mathematics Instruction through Collaboration of Students, Educators, and Researchers. *Educational Considerations,* 43(1), 9.

Kokka, K. (2019). Healing-Informed Social Justice Mathematics: Promoting Students' Sociopolitical Consciousness and Well-Being in Mathematics Class. *Journal of Urban Education,* 54(9), 1179–1209.

Kokka, K. (2020). Social Justice Pedagogy for Whom? Developing Privileged Students' Critical Mathematics Consciousness. *The Urban Review*, *52*, 778-803.

Ladson-Billings, G. (1995). Toward a Theory of Culturally Relevant Pedagogy. 32(3), 27.

Ladson-Billings, G. (2006a). Culturally relevant pedagogy 2.0: Aka the remix. *Harvard Educational Review,* 84(1), 74–84.

Ladson-Billings, G. (2006b). From the Achievement Gap to the Education Debt: Understanding Achievement in U.S. Schools. *Educational Researcher*, 35(7), 3–12.

Ladson-Billings, G. (2009). Race still matters: Critical race theory in education. In M. W. Apple, W. Au, & A. Gandin (Eds.), *The Routledge International Handbook of Critical Education* (pp. 110–122). Routledge.

Ladson-Billings, G. & Tate, W. (1995) Toward a critical race theory of education. *Teachers College Record,* 97, 47–68

Larnell, D. G. V., Bullock, D. E. C., & Jett, D. C. C. (2016). Rethinking Teaching and Learning Mathematics for Social Justice from a Critical Race Perspective. *Journal of Education,* 196(1), 19–29.

Lave, J., & Wenger, E. (1991). Situated learning: Legitimate peripheral participation. Cambridge university press.

Lee, E., Menkart, D., Okazawa-Rey, M., & Teaching for Change (Organization). (2002). *Beyond Heroes and Holidays: A Practical Guide to K-12 Anti-racist, Multicultural Education and Staff Development.* Teaching for Change.

Lee, H., Kersaint, G., Harper, S. R., & Driskell, S. O. (2014). Teachers' Use of Transnumeration in Solving Statistical Tasks with Dynamic Statistical Software. *Statistics Education Research Journal*, 13(1), 25–52.

Lee, H., & Tran, D. (2015). Statistical Habits of Mind. *Teaching statistics through data investigations MOOC-Ed, Friday Institute for Educational Innovation*: NC State University, NC State University, Raleigh, NC.

Lesser, L. M. (2007). Critical values and transforming data: Teaching statistics with social justice. *Journal of Statistics Education,* 15(1).

Levasseur, K., & Cuoco, A. (2003). Mathematical Habits of Mind. In H. Schoen & R. Charles (Eds.), *Teaching Mathematics through Problem Solving: Grades, 6-12*. (pp. 27–23). National Council of Teachers of Mathematics.

Lobato, J., Ellis, A., &; Muñoz, R. (2003). How “focusing phenomena” in the instructional environment support individual students’ generalizations. *Mathematical Thinking and Learning,* 5(1), 1–36.

Lobato, J., Hohensee, C., & Rhodehamel, B. (2013). Students' Mathematical Noticing. *Journal for Research in Mathematics Education*, 44(5), 809-850.

MacKay, R.J. & Oldford, W. (1994). *Srut 231 Course Notes Full 1994.* Waterloo: University of Waterloo.

Makar, K. (2004). *Developing Statistical Inquiry: Prospective Secondary Math and Science Teachers’ Investigations of Equity and Fairness through Analysis of Accountability Data* (Doctoral Dissertation). University of Texas at Austin.

Makar, K., & Fielding-Wells, J. (2011). Teaching teachers to teach statistical investigations. In *Teaching statistics in school mathematics-challenges for teaching and teacher education* (pp. 347–358). Springer.

Martin, D. B. (2007). Beyond missionaries or cannibals: Who should teach mathematics to African American children? *The High School Journal*, 91(1), 6–28.

Martin, D. B. (2009). Researching race in mathematics education. *Teachers College Record,* 111(2), 295–338.

McKernan, J. (1991). Action inquiry: Studied enactment. Forms of Curriculum Inquiry, 309–326.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative Data Analysis: An Expanded Sourcebook.* SAGE Publications.

Miles, M. B., Huberman, A. M., & Saldana, J. (2020). *Qualitative Data Analysis: A Methods Sourcebook (4th ed.)*. SAGE Publications.

Milner, I. (2007). Race, Culture, and Researcher Positionality: Working Through Dangers Seen, Unseen, and Unforeseen. *Educational Researcher,* 36(7), 388–400.

Moll, L. C., Amanti, C., Neff, D., & Gonzalez, N. (1992). *Funds of knowledge for teaching: Using a qualitative approach to connect homes and classrooms. Theory into Practice,* 31(2), 132–141.

National Council of Teachers of Mathematics. (2000). *Principles and Standards for School Mathematics*. National Council of Teachers of Teachers of Mathematics.

National Governors Association Center for Best Practices & Council of Chief State School Officers. (2010). *Common Core State Standards for English language arts and literacy in history/social studies, science, and technical subjects.* Washington, DC: Authors.

National Research Council. (2001). *Adding It Up: Helping Children Learn Mathematics.* National Academies Press.

O'Neil, C. (2017). *Weapons of math destruction*. Penguin Books.

Patton, M. (1990). *Qualitative Evaluation and Research Methods*. Sage.

Pedhazur, E. J. & Schmelkin, L. (1991). *Measurement, design, and analysis: An integrated approach. Hillsdale*, NJ: Lawrence Erlbaum Associates.

Pelaez, K. (2018). *Latent Class Analysis and Random Forest Ensemble to Identify At-Risk Students in Higher Education* (Masters thesis, San Diego State University).

Pfannkuch, M., & Rubick, A. (2002). An exploration of students' statistical thinking with given data. *Statistics Education Research Journal,* 1(2), 4–21.

Pfannkuch, M., & Wild, C. (2004). Towards an understanding of statistical thinking. In *The challenge of developing statistical literacy, reasoning and thinking* (pp. 17–46). Springer.

Philip, T. M., Olivares-Pasillas, M. C., & Rocha, J. (2016). Becoming racially literate about data and data-literate about race: Data visualizations in the classroom as a site of racial-ideological micro-contestations. *Cognition and Instruction*, *34*(4), 361-388.

Philip, T., Rocha, J., & Olivares-Pasillas, M. (2017). Supporting Teachers of Color as They Negotiate Classroom Pedagogies of Race: A Study of a Teacher's Struggle with "Friendly-Fire" Racism. *Teacher Education Quarterly, 44*(1), 59-79.

Prediger, S., Gravemeijer, K., & Confrey, J. (2015). Design research with a focus on learning processes: An overview on achievements and challenges. *ZDM*, 47(6), 877–891.

R Core Team (2017) *R: A Language and Environment for Statistical Computing*. https://www.R-project.org/

Reed, R. J., & Oppong, N. (2005). Looking Critically at Teachers' Attention to Equity in their Classrooms. *The Mathematics Educator,* 1, 2–15.

Rosenbaum, P. R. (2020). Modern Algorithms for Matching in Observational Studies. A*nnual Review of Statistics and Its Application,* 7(1), 143–176.

Rossman, A. J. (1994). Televisions, physicians, and life expectancy. *Journal of Statistics Education*, 2(2).

Rossman, A. J., Chance, B. L., & Lock, R. H. (2001). Workshop statistics: Discovery with data and Fathom. *Springer Science & Business Media.*

Rousseau, C., & Tate, W. F. (2003). No time like the present: Reflecting on equity in school mathematics. *Theory into Practice,* 42(3), 210–216.

Rubel, L. H., Lim, V. Y., Hall-Wieckert, M., & Sullivan, M. (2016). Teaching mathematics for spatial justice: An investigation of the lottery. *Cognition and Instruction*, *34*(1), 1-26.

Rubin, D. B. (1973). Matching to Remove Bias in Observational Studies. *Biometrics*, 29(1), 159–183. JSTOR.

Rubin, D. B. (1976). Multivariate Matching Methods That are Equal Percent Bias Reducing, I: Some Examples. *Biometrics*, 32(1), 109–120. JSTOR.

Rubin, D. B. (1979). Using Multivariate Matched Sampling and Regression Adjustment to Control Bias in Observational Studies. *Journal of the American Statistical Association,* 74(366a), 318–328.

Schorr, R. Y., & Goldin, G. A. (2008). Students' expression of affect in an inner-city simcalc classroom. *Educational Studies in Mathematics*, 68(2), 131–148.

Sentance, S., & Waite, J. (2017). PRIMM: Exploring Pedagogical Approaches for Teaching Text-Based Programming in School. *Proceedings of the 12th Workshop on Primary and Secondary Computing Education,* 113–114.

Shah, N. (2017). Race, Ideology, and Academic Ability: A Relational Analysis of Racial Narratives in Mathematics. *Teachers College Record*, 119, 42.

Shulman, L. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher,* 15(2), 4–14.

Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–23.

Sleeter, C. E., & McLaren, P. (1995). *Multicultural Education, Critical Pedagogy, and the Politics of Difference.* State University of New York Press.

Solorzano, D. G., & Delgado Bernal, D. (2001). Examining transformational resistance through a critical race and LatCrit theory framework: Chicana and Chicano students in an urban context. *Urban Education*, 36(3), 308-342.

Solórzano, D. G., & Yosso, T. J. (2002). Critical race methodology: Counter-storytelling as an analytical framework for education research. *Qualitative Inquiry,* 8(1), 23-44.

Staats, C., Capatosta, K., Tenney, L., & Mamo, S. (2017). Race and ethnicity: Views from inside the unconscious mind. *State of the Science: Implicit Bias Review.*

Stinson, D. W., & Wager, A. (2012). A sojourn into the empowering uncertainties of teaching and learning mathematics for social change. *Teaching mathematics for social justice: Conversations with educators,* 1(1), 3-18.

Strauss, A., & Corbin, J. (1994). Grounded theory methodology: An overview. In *Handbook of qualitative research.* (pp. 273–285). Sage Publications, Inc.

Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science: A Review Journal of the Institute of Mathematical Statistics*, 25(1), 1–21. PubMed.

Tate, W. F. (1995). Returning to the root: A culturally relevant approach to mathematics pedagogy. *Theory Into Practice*, 34(3), 166–173.

Thanheiser, E., Harper, F. K., Jackson, C., Jessup, N., Kalinec-Craig, C., Yeh, C., & Sugimoto, A. (2020). New Working Group: Teaching mathematics for social justice in the context of University Mathematics Content and Methods courses. *Mathematics Education Across Cultures: Proceedings of the 42nd Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, 185–188.

Thanheiser, E., Rosencrans, B., Ellis, B., Sugimoto, A., Kulow, T., & Robinson, M. (2020). Modeling to understand the world around us and our place in it: If the world were a village. *Mathematics Education Across Cultures: Proceedings of the 42nd Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education,* 1527–1534.

Torchiano, M (2020). *effsize: Efficient Effect Size Computation*. R package version 0.8.1, https://CRAN.R-project.org/package=effsize.

Turner, E. E., & Font Strawhun, B. T. (2005). With math, it's like you have more defense. *Rethinking mathematics: Teaching social justice by the numbers*, 81-87.

Vahey, P., Roy, G. J., & Fueyo, V. (2013). Sustainable Use of Dynamic Representational Environments: Toward a District-Wide Adoption of SimCalc-Based Materials. In S. J. Hegedus & J. Roschelle (Eds.), *The SimCalc Vision and Contributions: Democratizing Access to Important Mathematics* (pp. 183–202). Springer Netherlands.

Visnovska, J., & Cobb, P. (2019). Supporting shifts in teachers' views of a classroom statistical activity: Problem context in teaching statistics. *Mathematical Thinking and Learning,* 21(4), 285–305.

Walters, C. D. (2017). *The development of mathematical knowledge for teaching for quantitative reasoning using video-based instruction* (Doctoral Dissertation). University of California, San Diego, and San Diego State University

Weiland, T. (2017). Problematizing statistical literacy: An intersection of critical and statistical literacies. *Educational Studies in Mathematics,* 96(1), 33–47.

Weingarten Learning Resources Center (2017, January 9). *Writing Strategies: What's Your Positionality?* The Weingarten Blog. https://weingartenlrc.wordpress.com/2017/01/09/research-writing-whats-your-positionality/

Whitaker, D., Foti, S., & Jacobbe, T. (2015). The Levels of Conceptual Understanding in Statistics (LOCUS) Project: Results of the Pilot Study. Numeracy: Advancing Education in Quantitative Literacy, 8(2).

Wild, C. J., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review*, 67(3), 223–248.

Woodard, V., & Lee, H. (2021). How Students Use Statistical Computing in Problem Solving. *Journal of Statistics Education*, 1–18.

Yosso, T. J. (2005). Whose culture has capital? A critical race theory discussion of community cultural wealth. *Race Ethnicity and Education*, 8(1), 69–91.

Zahner, W., Calleros, E. D., & Pelaez, K. (2021a). *Designing learning environments to promote academic literacy in mathematics in multilingual secondary mathematics classrooms*. ZDM–Mathematics Education, 53(2), 359-373.

Zahner, W., Pelaez, K., & Calleros, E. D. (2021b). *Principles for Curriculum Design and Pedagogy in Multilingual Secondary Mathematics Classrooms*. In Multilingual Education Yearbook 2021 (pp. 235-255). Springer, Cham.

Zhao, Z. (2004). Using Matching to Estimate Treatment Effects: Data Requirements, Matching Metrics, and Monte Carlo Evidence. *The Review of Economics and Statistics*, 86(1), 91–107.

Zwiers, J., Dieckmann, J., Rutherford-Quach, S., Daro, V., Skarin, R., Weiss, S., & Malamut, J. (2017). *Principles for the design of mathematics curricula: Promoting language and content development*. Stanford University UL/Scale

Appendices

Appendix 1: Sample Pre- and Post-Assessment Questions for the Study Design Unit

Study Design: Multiple Choice

Source: LOCUS (link)
In a survey of students from middle schools in a large city in the United States, the most popular type of music was hip-hop. Would it be appropriate to conclude that hip-hop is the most popular type of music for all middle school students in the United States?

(a) No, because hip-hop is not my favorite type of music.
(b) No, because middle school students from across the country should be surveyed. Answer
(c) Yes, because the opinions of other middle school students would be similar.
(d) Yes, because hip-hop is my favorite type of music.

Source: LOCUS (link)
A middle school class found that there was a relationship between the number of seeds in a pumpkin and the number of ribs on the pumpkin. The more ribs the pumpkin had, the more seeds it had. Can it be concluded that the more ribs a cantaloupe has, the more seeds it has?

(a) No, because cantaloupes are usually much smaller than pumpkins.
(b) No, because data have only been collected on pumpkins and not on cantaloupes. Answer
(c) Yes, because both pumpkins and cantaloupes are fruits and all fruits have similar properties.
(d) Yes, because both pumpkins and cantaloupes are round and seeds inside would grow the same way.

Source: LOCUS (link)
An advertisement makes the claim: "Lighter shoes make you run faster." Of the following, which is the best way to investigate this claim?

(a) Choose the records of the top twenty runners who are wearing the lighter shoes and compare their times to run 400 meters before and after they began wearing the shoes.
(b) Choose twenty runners and select ten at random to wear lighter shoes and have the other ten wear heavier shoes to run 400 meters and compare their times. Answer
(c) Choose twenty runners at random and have the women wear the lighter shoes and the men wear the heavier shoes to run 400 meters and compare their times.
(d) Choose to observe the results of 400-meter races for the next year and see how many winners are wearing the lighter shoes

Source: LOCUS (link)
A seventh grade class of twenty-seven students wants to estimate the proportion of eligible voters in their school district who intend to vote in the upcoming school board election. They decide to base their estimate on 270 eligible voters. Which of the following plans would allow the class to generalize from the sample to the population of all eligible voters?

(a) Have each of the 27 students randomly select 10 neighbors to participate in the survey.
(b) Mail surveys to all eligible voters and take the first 270 who respond.

(c) Mail surveys to 270 randomly selected eligible voters and follow-up with those who do not respond. Answer
(d) Survey 270 people visiting a local grocery store on the Saturday before the election.

Source: LOCUS (link)
A study was conducted to investigate whether washing with soap and water or using hand sanitizer removes more bacteria from a person's hands. Volunteers were recruited from a high school and randomly assigned to a group that washed their hands with soap and water or to a group that used hand sanitizer. When they were finished, each volunteer pressed his or her hands into specially prepared petri dishes. After several days, the number of bacteria colonies was counted on each petri dish. Which of the following statements best describes the random assignment in this study?

(a) The random assignment was important because it tends to create groups that are similar with respect to other variables that might affect bacteria growth. Answer
(b) The random assignment was important so that these results could be applied to all high school students.
(c) Including random assignment was incorrect because students should be divided into the two groups based on their usual method of cleaning their hands.
(d) The random assignment was unnecessary because using volunteers makes the study worthless.

Source: LOCUS (link)
Each student in a class selected a random sample of 25 marbles from a large jar of red and white marbles and then determined the proportion of red marbles in his or her sample. The proportion in one student's sample was 0.28. The two people sitting beside that student got sample proportions of 0.36 and 0.24. Of the following, which gives the best explanation for the differences in the sample proportions?

(a) Sample proportions will generally differ from one random sample to another. Answer
(b) Only one of the students knew the true proportion of red marbles.
(c) Two of the three students obtained bad samples.
(d) Two of the three students miscalculated the percentages.

Source: LOCUS (link)
Joe and Tom attended a rally on a Thursday night to protest the removal of vending machines from their school. They both wondered what percentage of students from their school actually attended the rally. Joe decided to get an estimate the next day (Friday) by selecting a random sample of 25 students and asking each one if he or she was at the rally Thursday night. Six of the students (24 percent) said they attended the rally. On Monday, Tom selected a random sample of 50 students and asked each one the same question. Fifteen of the students (30 percent) said they had attended the rally.
Which of the following is the most likely reason that Joe and Tom came up with different percentages?

(a) A larger sample has more students and is more likely to have a higher percentage than the smaller sample.
(b) If they had both taken their samples on Friday the percentages would be the same.
(c) One of them did something wrong when selecting his random sample of students.

(d) The difference between the sample percentages could happen by chance with random sampling. Answer

Source: LOCUS (link)
The President of a large university with 30,000 students wants to investigate student support for an increase in tuition (the cost to enroll in classes). The President requests a sample of 200 students. Which of the following methods of sample selection would be best?

(a) Select 200 students at random from the list of students currently enrolled at the university. Answer
(b) Select 200 students at random from those in the campus bookstore on the first day of class.
(c) Select 50 students at random from the list of the 10,000 students living in dorms on campus.
(d) Select 50 students at random from each of the first four football games of the season.

Source: LOCUS (link)
Rebecca wants to know how many books students in her school read over summer vacation. She attends a large school, and doesn't have time to ask every student. Which of the following would best allow her to make generalizations about all students in her school?

(a) Select 40 students from her school at random Answer
(b) Select all of the students in her English class
(c) Select the first 40 students that she sees after school
(d) Select 40 of the students in the library at random

Source: LOCUS (link)
Each member of a random sample of 1,000 adult males from the United States was asked a number of questions, including questions about height and annual income. When the responses were analyzed, it was determined that taller men had greater incomes than shorter men, on average, and the difference was statistically significant. Which of the following conclusions would be most appropriate based on these results?

(a) The study establishes that being tall causes men to have greater incomes, on average, and this conclusion can be generalized to all men in the United States
(b) The study establishes that being tall causes men to have greater incomes, on average, but this result only applies to the men in the sample
(c) The study establishes that taller men tend to have greater incomes, on average, than shorter men, and this conclusion can be generalized to all men in the United States Answer
(d) The study establishes that taller men tend to have greater incomes, on average, than shorter men, but this result only applies to the men in the sample

Source: LOCUS (link)
To investigate a possible association between chocolate consumption and depression, a researcher had 20 volunteers who suffer from depression and 20 volunteers who do not suffer from depression keep food journals for one month. He then used the information in the journals to determine the amount of chocolate consumed for each person. The resulting data were then used to compare chocolate consumption for the two groups. This is an example of

(a) A census
(b) A random sample
(c) An experiment
(d) An observational study Answer

Source: LOCUS (link)
Which of the following plans would be the best way to collect data to determine if listening to music while studying for a social studies exam has an effect on exam performance?

(a) Randomly select a sample of students enrolled in a social studies class. After an exam in that class, ask each of these students the following two questions: (1) Did you listen to music while studying for the exam? and (2) What score did you get on the exam?
(b) Survey every student enrolled in a social studies class. After an exam in that class, ask each of these students the following two questions: (1) Did you listen to music while studying for the exam? and (2) What score did you get on the exam?
(c) Use a group of student volunteers who are enrolled in a social studies class. Randomly assign each student to either the music group or the no music group. Students in the music group will listen to music while studying for an upcoming exam and those in the no music group will not listen to music while studying for this exam. Record the exam score for each student. Answer
(d) Use a group of students who are all enrolled in a social studies class. Have each student volunteer for either the music group or the no music group. Students in the music group will listen to music while studying for an upcoming exam and those in the no music group will not listen to music while studying for this exam. Record the exam score for each student.

Source: LOCUS (link)
A student wants to estimate the mean number of books that have been read by all students at his school over the summer. On Monday morning, he will survey the first 35 students who enter the library. Is this the best way to select a sample for this purpose?

(a) No. The student should survey students entering the library on more than one day of the week.
(b) No. The student should take a random sample of students entering the library instead.
(c) No. The student should take a random sample of students from all students, not just those entering the library. Answer
(d) Yes. Selecting a sample in this way will not introduce the possibility of bias.

Source: LOCUS (link)
Marvin would like to answer the following question:
In Houston, do restaurants with drive-thru service have more health code violations, on average, than restaurants without drive-thru service?
Which of the following study designs would help Marvin answer his question?

(a) Select restaurants at random and use random assignment to divide them into two groups (with and without drive-thru service). Use health department records to determine the number of violations over the past 12 months.

(b) Select a random sample of restaurants with drive-through service and a random sample of restaurants without drive-through service. Use health department records to determine the number of violations over the past 12 months. Answer
(c) Ask restaurant owners to volunteer to complete a survey. Collect data on their responses to these questions: (1) Does your restaurant have drive-thru service? (2) How many health code violations have you had over the past 12 months?
(d) Select restaurants at random and divide them into two groups (with and without health code violations). Count the number of restaurants in each group that offered drive-thru service over the past 12 months.

Study Design: Free Response

Source: LOCUS (link)
A department store manager wants to know which of two advertisements is more effective in increasing sales among people who have a credit card with the store. A sample of 100 people will be selected from the 5,300 people who have a credit card with the store. Each person in the sample will be called and read one of the two advertisements. It will then be determined if the credit card holder makes a purchase at the department store within two weeks of receiving the call.
(a) Describe the method you would use to determine which credit card holders should be included in the sample. Provide enough detail so that someone else would be able to carry out your method.
(b) For each person in the sample, the department store manager will flip a coin. If it lands heads up, advertisement A will be read. If it lands tails up, advertisement B will be read. Why would the manager use this method to decide which advertisement is read to each person?

Source: LOCUS (link)
A farmer conducted an experiment to find out whether a new type of fertilizer would increase the size of tomatoes grown on his farm. The farmer randomly assigned 10 tomato plants to receive the new fertilizer and 10 tomato plants to receive the old fertilizer. All other growing conditions were the same for the 20 plants. At the end of the experiment, the mean weight of tomatoes grown with the new fertilizer was 0.4 ounce heavier than the mean weight of the tomatoes grown with the old fertilizer.
(a) Describe one method that the farmer could have used to randomly assign the 20 plants into groups of 10 each.
(b) Based on the results, the farmer is convinced that the new fertilizer produces heavier tomatoes on average. Briefly explain to the farmer why simply comparing the two means is not enough to provide convincing evidence that the new fertilizer produces heavier tomatoes.

Source: LOCUS (link)
A public library is currently open from 9 a.m. to 5 p.m. on Saturdays. The director is considering whether or not to keep the library open until 8 p.m. on Saturdays. A library employee develops a one-question survey. The question is, Would you use the library between the hours of 5 p.m. and 8 p.m. on Saturdays? The survey was administered using two different methods.

In Method 1, 100 individuals were selected at random from a list of people who have library cards at that library.
(a) To what population can the results of Method 1 be generalized?
In Method 2, the survey was given to all 25 individuals who were in the library at 4:45 p.m. on a particular Saturday. The results of the two surveys are summarized in the table below.

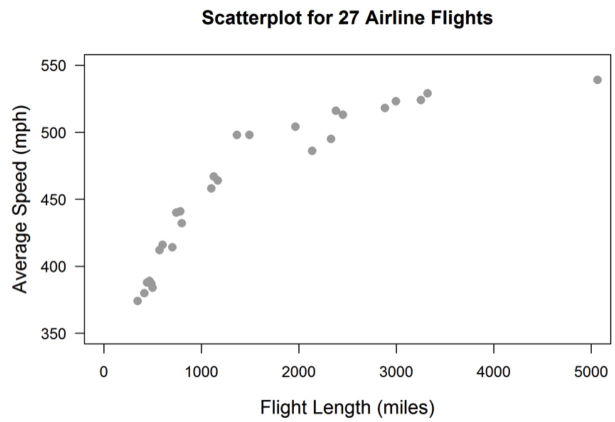
Method	Yes	No
1	30	70
2	15	10

(b) Create a graphical display that allows you to compare the results of the two surveys.
(c) Why do you think the two methods produced such different results?

Appendix 2: Sample Pre- and Post-Assessment Questions for the Regression Unit

Regression: Multiple Choice

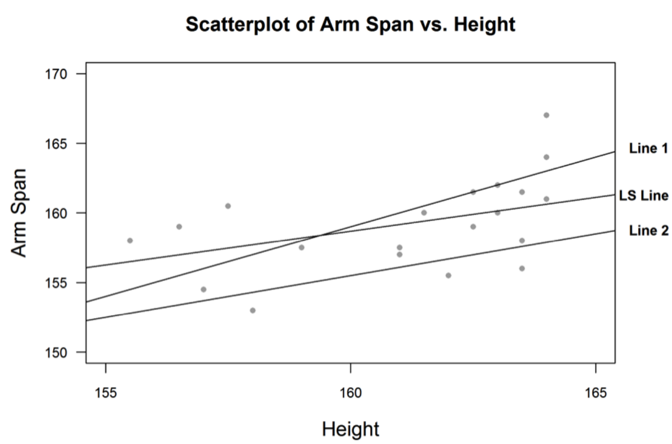
Source: LOCUS (link)
The average mean speed in miles per hour and length of flight in miles were recorded for 27 airline flights. The scatterplot of these data is shown below.



Which of the following best describes the relationship between flight length and average speed?

(a) There is a linear relationship between flight length and speed.
(b) There is a non-linear relationship between flight length and speed. Answer
(c) There is no apparent relationship between flight length and speed.
(d) There is a relationship because speed tends to decrease as flight length increases.

Source: LOCUS (link)
The scatterplot below shows the relationship between height and arm span for a group of students. The least squares line (labeled LS line) and two other lines have been added to the scatterplot. Which of the following statements do you agree with?



(a) Compared to the other lines, Line 1 has the smallest sum of squared residuals.
(b) The sum of squared residuals for Line 1 is greater than the sum of squared residuals for Line 2.=

(c) Compared to the other lines, the least squares line has the smallest sum of squared residuals. Answer
(d) The sum of squared residuals for the least squares line is greater than the sum of squared residuals for Line 2.

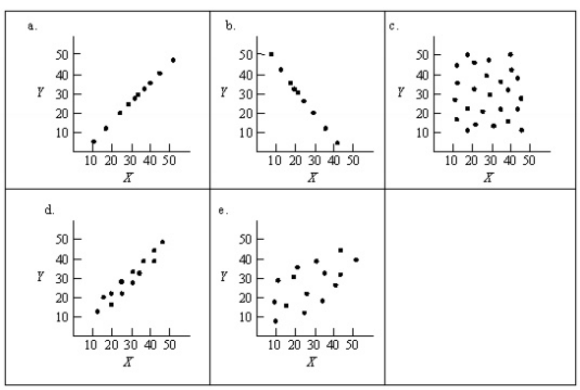
Source: ARTISTT
A college statistics class conducted a survey of how students spend their money. They gathered data from a large sample of students who estimated how much money they typically spent each week in different categories (e.g., food, entertainment, etc.). Jack wanted to predict how much students spend on leisure activities based on how much they spend on necessities. He calculated a regression equation, using spending on necessities as the explanatory variable (x) and leisure spending as the response variable (y). Read and evaluate this statement:

Jack should use the regression equation to make a prediction only if the scatterplot of x and y indicates a reasonably linear relationship between "necessities spending" and "leisure spending."

a. Agree, a regression equation is useful only when there appears to be a fairly linear relationship between x and y. Answer
b. Disagree, you can make a good prediction whether or not there is a linear relationship between the two variables.
c. Disagree, you should always use a regression line to make predictions.

Source: ARTISTT

Consider the five scatterplots that are shown below:



Select the scatterplot that shows a correlation of zero

a. a
b. b
c. c Answer
d. d
e. e

Select the scatterplot that shows a correlation of about 0.60

a. a
b. b
c. c
d. d

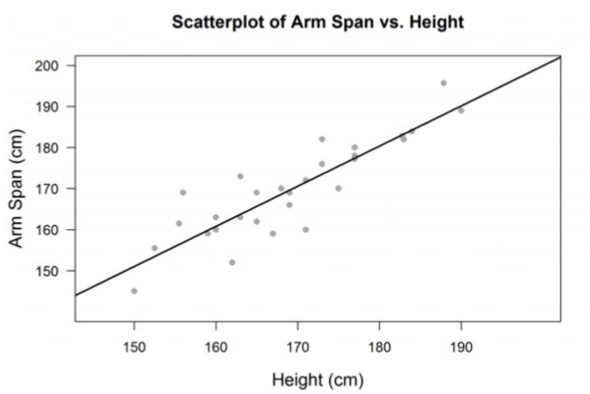
e. e Answer

Select the scatterplot that shows the strongest relationship between the X and Y variables?

a. a
b. b
c. a and b Answer
d. a and c
e. a, b, and d

Regression: Free Response

Source: LOCUS (link)
The heights (in centimeters) and arm spans (in centimeters) of 31 students were measured. The association between x (height) and y (arm span) is shown in the scatterplot below. The equation of the least-squares regression line for this association is also given [LOCUS].



estimated arm span = 4.5 + 0.977height

(a) If Mike is 5 cm taller than George, what is the expected difference in their arm spans? Show your work.
(b) Jane is 158 cm tall and has an arm span of 154 cm. Rhonda is 163 cm tall and has an arm span of 165 cm. Does the least-squares regression line give a more accurate predicted value for Jane or Rhonda? Explain.
(c) Doug is 210 cm tall. Would you use this least-squares regression line to predict his arm span? Explain.

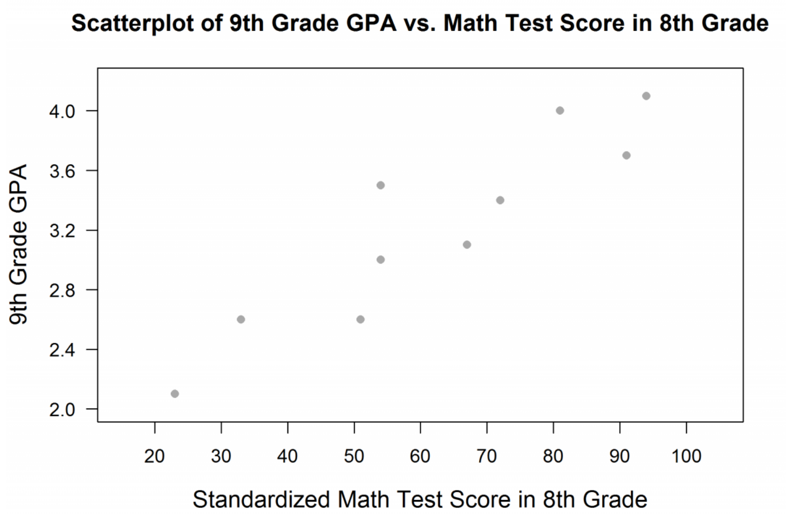
Source: LOCUS (link)
A study was carried out to investigate whether there is a relationship between the percent of hearing loss and the volume at which people typically listen to music. Ten high school students agreed to participate in a study. Each was given a music player with headphones and was asked to listen to music for 10 minutes. The students were told to adjust the volume to a comfortable setting. After 10 minutes, the volume setting, which ranges from 1 to 10, was observed for each student. Each student then took a hearing test, and a measure of hearing loss (in percent) was recorded. The data are shown in the table below.

Volume Setting (x)	**Hearing Loss (y)**

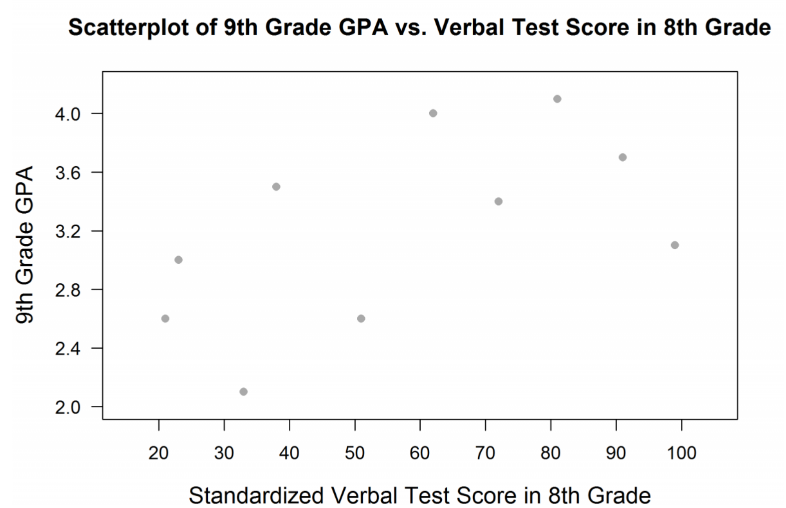
8	23
10	24
1	11
4	9
5	15
8	19
3	14
1	5
2	7
8	15

(a) Construct an appropriate graphical display that allows you to investigate the relationship between volume setting and hearing loss.
(b) Based on the graphical display, describe the relationship between volume setting and hearing loss.
(c) From this study, is it reasonable to conclude that listening to music at a high volume causes hearing loss? Explain why or why not

Source: LOCUS (link)
A random sample of 10 high school students was selected to investigate the relationship between standardized test scores in 8th grade and GPA (grade point average) in 9th grade.
The scatterplot below shows the relationship between standardized math test scores in 8th grade and GPA (grade point average) in 9th grade.



(a) Based on scatterplot, describe the relationship between standardized math test scores in 8th grade and GPA (grade point average) in 9th grade.
For the data on standardized math test score in 8th grade and GPA in 9th grade, the value of the correlation coefficient is r = 0.92. The scatterplot below shows the relationship between standardized verbal test scores in 8th grade and GPA (grade point average) in 9th grade.



(b) For the data on standardized verbal test scores in 8th grade and GPA in 9th grade, will the value of the correlation coefficient be greater than, less than, or about the same as r = 0.92? Explain.
(c) If you want to predict 9th grade GPA, which variable would you use as a predictor— 8th grade standardized math test score or 8th grade standardized verbal test score? Explain.

Source: LOCUS (link)
The student council members at a large middle school have been asked to recommend an activity to be added to physical education classes next year. They decide to survey 100 students and ask them to choose their favorite among the following activities: kickball, tennis, yoga, or dance.

(a) What question should be asked on the survey? Write the question as it would appear on the survey.
(b) Describe the process you would use to select a sample of 100 students to answer your question.
(c) Create a table or graph summarizing possible responses from the survey. The table or graph should be reasonable for this situation.
(d) What activity should the student council recommend be added to physical education classes next year? Justify your choice based on your answer to part (c).

Appendix 3: Task-Based Interview

Excerpt from: **Does teacher collaboration improve student learning?**
By: A joint project of the Spencer Foundation and Public Agenda
https://files.eric.ed.gov/fulltext/ED591332.pdf

A growing body of research shows that when teachers work more collaboratively, student outcomes can improve, teachers can be more satisfied in their jobs and teacher turnover can decrease. A focus on advancing teaching and learning by fostering collaboration stands in contrast to a focus on improving and assessing teachers solely as individuals. How can teachers, principals, superintendents and school boards begin to understand what collaboration might mean for their schools, districts and students?

Moreover, no educational practice is used in isolation. A school that encourages a collaborative approach to induction may also be characterized by more collegial relationships between teachers and principals or by greater coordination of curricula across grade levels. Therefore, it can be hard, though not necessarily impossible, for researchers to isolate any one approach to collaboration from the broader context and character of a school.

For example, an induction program for teachers new to a school might consist merely of a single week's orientation at the beginning of the school year, followed by infrequent and unstructured meetings with a colleague. But a more comprehensive induction program might continue over multiple years and incorporate frequent peer mentoring, regular collaborative planning, quarterly feedback following observations of instruction and repeated opportunities to observe master teachers' instruction.

In Japan, the process typically works as follows: A group of teachers reviews a curriculum and works collaboratively to identify goals for student learning and to design a lesson. They conduct a live classroom lesson led by one teacher and observed by the rest, who collect data and make observations on teaching and learning during the lesson. Teachers then meet to discuss and reflect on the data to evaluate the lesson on whether and how it achieved the student learning goals. Finally, this reflection is documented and carried forward in an iterative process to continue to refine the lesson and teaching methods. In addition to or instead of these steps, teachers may observe a highly accomplished teacher talk through the planning of a lesson and then observe that teacher teach it and reflect on it. New teachers may be asked to do the same, with guidance from peers and from more accomplished teachers.

Reading Questions

1. The article implies that teacher collaboration may increase student learning. Why?
2. What types of collaborations might you be interested in? Please give at least three examples of types of collaborations and explain why.
3. What other factors may influence student learning? Please give at least three examples and explain why.
4. If you wanted to confirm this study, what types of data would you collect to measure student learning? Please give at least three examples of things you might collect and explain why.

Activity

Background: The superintendent of the Dolores Huerta Unified School District (DHUSD) recently read the same article. They are interested in learning more about the great mathematics teaching and learning and their middle schools. To highlight some of this awesome work, the superintendent is interested in visiting three schools. Since DHUSD knows that you are familiar with statistical investigations, they are hoping that you can help them design, explore, and analyze data for this project.

Task: DHUSD would like you to make one or two recommendations that would address the following questions. The only requirement is that you use at least one linear regression model.

Question 1: Which three schools should we visit and why?

Question 2: What are predictors of a school's mathematics learning?

Question 3: What evidence do you have to support your questions? Include any necessary plots and analysis.

District Information: DHUSD is one of the largest districts in the state. The rising bio and tech industry, proximity to an international border, and racial and ethnic diversity has shaped the history of the city.

Data:

The data, variable names, and type of data are summarized below. The data can be found here.

Variable Name	Description	Data Type
CDSCode	School code assigned from the government	Text
School	Name of the middle school	Text
Charter	Marks if the schools is charter or not charter	Binary: 0 - Traditional, 1 - Charter
PercentFRPM	Proportion of students who qualify for free or reduced priced meals (FRPM)	Numeric
Race	Proportion of students classified as the given race	Numeric
TotalEnrollment	Total enrollment at the middle school	Numeric
MeanScore_Eng6 / 7 / 8 Mean.Scale.Score_Math6 / 7 / 8	Average score of all students on a 6th, 7th, or 8th grade English and Math standardized assessment	Numeric
PercMetAbove_Eg6/ 7 / 8 PercMetAbove_Math6/ 7 / 8	Percent of students that met or exceeding the 6th, 7th, or 8th grade English / Math standards on a standardized assessment	Numeric

Appendix 4: Contact Summary Form

Contact Summary Form - Interviews

Contact Date: ___ Time: ___ Interviewee: ___

What main practices did you encounter in this interview?

Cluster	Code	Definition	Findings
Transnumeration			
	Forming data	Creating meaningful representations of the data	
	Changing data	Altering visuals to enhance meaning	
Variability			
	Anticipating and Looking for Variation	Describing variability within a group, variability within and across groups, covariability, and variability in model fitting	
	Generalizability	Considering the generalizability of data	
Interpretations			
	Relevance of Data	Considering how well data measures an attribute of interest in a statistical task	
	Sociopolitical Nature of Data	Considering the sociopolitical nature of data	
Implications			
	Implications of Data	Considering the problem context when providing data-based conclusions	
	Sociopolitical Implications of Data	Considering the sociopolitical context when providing data-based arguments	

Any new practices?

New Code	Definition	Findings

Was there anything that struck you as salient, interesting, or important during the interview?

What new (or remaining) target questions do you have in considering the next interview?

Appendix 5: Pre- and Post-Survey

The survey includes five main sections: (a) introduction, (b) teaching experiences and beliefs, (c) views on equity, (d) experience with mathematics for social justice, and (e) conclusion. The survey will be administered through Qualtrics twice, once in at the end of the first class and once at the end. Questions are identified as pre-survey only (PRE), post-survey only (POST), and both pre- and post-survey (BOTH).

Intro

Q1. How many statistics courses have you taken before this? (PRE)

○ 1 ○ 2 ○ 3 ○ 4+

Q2. What is(are) your declared or planned major(s)? (PRE)
Q3. What is(are) your declared or planned minor(s)? (PRE)
Q4. Do you plan on being a math teacher? (PRE)

○ Yes ○ No

Q5. Please indicate your confidence with the following: (BOTH)

	Not at all confident	Slightly confident	Somewhat confident	Fairly confident	Completely confident
a. How confident are you with statistics?	○	○	○	○	○
b. How confident are you with the R programming language?	○	○	○	○	○

Teaching

Q6. How were your experiences with statistics as a student? (PRE)
Q7. What makes for a successful statistics class? (BOTH)
Q8. What kind of a teacher do you want to be? (BOTH)
Q9. In your opinion, what is the purpose of education? (BOTH)
Q10. In the context of education, how would you define equity? (BOTH)

Equity

Q11. Below are some sample definitions of equity. Which most closely aligns with your meaning of equity in education? (BOTH)

○ Equity is the same treatment for everyone so that all students have an equal chance to meet the same standards and an equal opportunity to master those standards.

○ Equity is investing in students most at-risk, those whose success or failure in life depends on their school experience.

○ Equity compensates for social injustice to specific groups of students who have not received fair treatment or a fair share of the resources by giving preference, when all else is equal, to underrepresented groups.

○ Equity involves a safety net for individual differences (backup programs, differentiated curricula, and other resources) so that when one program does not work for a particular student, other options are available.

○ Equity involves maximum return on investment: a concentration of scarce resources on those students who are most likely to succeed.

○ Equity requires being responsive to students' backgrounds, experiences, cultural perspectives, traditions, and knowledge when designing and implementing a mathematics program and assessing its effectiveness.

○ Equity in mathematics education means it is no longer possible to predict mathematics achievement and participation based solely on characteristics such as race, class, ethnicity, sex, beliefs and creeds, and proficiency in the dominant language.

○ Equity happens when students develop a critical consciousness through which they challenge the status quo of the current social order.

Q12. Which factors are most responsible for ensuring children's educational success in the U.S.? (select your top 4, in order of importance.) (BOTH)

______ Students themselves
______ Students' parents / caregivers
______ Students' communities or cultural groups
______ Teachers
______ Standardized tests
______ Schools and districts
______ State and federal governments
______ Media and American cultural leaders
______ Educational researchers
______ Curriculum developers
______ Other

Q13. Please indicate how much you agree or disagree with the following statements (BOTH):

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
a. As a student in a non-math college class, I am comfortable discussing educational equity issues.	○	○	○	○	○	○
b. As a student in a math college class, I am comfortable discussing educational equity issues.	○	○	○	○	○	○
c. As a future teacher, I am comfortable discussing educational equity issues.	○	○	○	○	○	○

d. When discussing equity issues, I worry that I will say the wrong thing and offend someone.	○	○	○	○	○	○
e. When discussing equity issues, others are likely to do or say things that hurt or anger me.	○	○	○	○	○	○

Q14. If you agreed with the above statement (others are likely to do or say things that hurt or anger me), please describe any specific words or behaviors that are hurtful when used. (PRE)
Q15. List anything that the instructor or your fellow students can do to support your engagement with these conversations. (PRE)

Math for Social Justice

Q16. In your opinion, what, if any, is the relationship between teaching mathematics and teaching about social or political issues?
Q17. Would you like to incorporate social or political issues into your mathematics classroom? Why or why not?
Q18. How familiar are you with the expression "math for social justice"? (PRE)

○ Not familiar at all

○ Slightly familiar

○ Moderately familiar

○ Very familiar

○ Extremely familiar

Q19. If you are familiar with the expression "math for social justice," how would you define this? (PRE)
Q20. What are some strengths to teaching mathematics for social justice? (POST)
Q21. What are some weaknesses to teaching mathematics for social justice? (POST)
Q22. Is teaching mathematics for social justice something you want to incorporate into your future classes? Why or Why not? (POST)

Conclusion

Q23. How do you racially or ethnically identify? (PRE)
Q24. What is your gender identity? (PRE)
Q25. Is there anything else I should know about you? (PRE)
Q26. What are you going to take away from this class? (POST)
Q27. What do you want to learn more about? (POST)