Title

Toward Critical Data-Scientific Literacy: An Intersectional Analysis of the Development of Student Identities in an Introduction to Data Science Course

Author
Olivares Pasillas, Maria Concepcion

Publication Date
2017

UNIVERSITY OF CALIFORNIA

Los Angeles

Toward Critical Data-Scientific Literacy:

An Intersectional Analysis of the Development of Student Identities in an Introduction to Data Science Course

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Education

by

María Concepción Olivares Pasillas

2017

ABSTRACT OF THE DISSERTATION

Toward Critical Data-Scientific Literacy:

An Intersectional Analysis of the Development of Student Identities in an

Introduction to Data Science Course

by

María Concepción Olivares Pasillas

Doctor of Philosophy in Education

University of California, Los Angeles, 2017

Professor Thomas M. Philip, Co-Chair

Professor Douglas M. Kellner, Co-Chair

The national imperative to increase the presence of women and people of color in science, technology, engineering, and mathematics (STEM) coupled with the growing presence of Latinos in the United States has led to the dramatic rise of programs and initiatives aimed at improving access to and equity in STEM careers and education for Latino youth. Through the use of critical social theory and critical theory of education as guiding frameworks, the dissertation examines an instantiation of STEM reform efforts to analyze the classroom participation structure that emerged in a piloted introduction to data science course at a local high school in one of the largest school districts in the country. The study is particularly

concerned with identifying emergent classroom norms and practices, and understanding whether and how they came to support and/or hinder students' opportunities to learn richly with data through an analysis of the development of student learning identities. This qualitative case study draws on audio-recorded student interviews, video-recorded classroom observations, and field notes collected during the second year of the curriculum's implementation. To identify classroom norms and practices as they relate to the development of student identities as data science doers, the study examines the classroom participation structure (Cobb and Hodge, 2002) and employs Cobb, Gresalfi, and Hodge's (2009) interpretive scheme for analyzing the development of mathematical student identity (also see Cobb & Hodge, 2010). While the multiperspectival approach of this study will provide innovatively insightful contributions to a number of fields including education, cultural studies, data and computer science, the study will also push how educators, learning science researchers, curriculum writers, and policymakers think about the pursuit of equity in STEM education in general and data science-oriented programs and initiatives in particular as they relate to STEM reform efforts.

The dissertation of María Concepción Olivares Pasillas is approved.

Robert L. Gould

William A. Sandoval

Thomas M. Philip, Committee Co-Chair

Douglas M. Kellner, Committee Co-Chair

University of California, Los Angeles

2017

Dedication

Para mis queridos padres, Leonor Cerda y Pedro Olivares.

TABLE OF CONTENTS

List of Figures and Tables

Figures

TABLES

ACKNOWLEDGEMENTS

As I reflect on the achievement of completing my PhD, I am overwhelmed with gratitude to those who, in one way or another, helped me along the way by offering their guidance, friendship, life experiences, and emotional support. I was fortunate enough to cross paths with individuals who left meaningful and lasting impressions in my heart and in my scholarship—a fortune for which I am thankful beyond words.

I would like to express sincere appreciation and gratitude to my honorable doctoral committee. As a whole, I appreciate the honesty, criticality, and direction you offered—it is thanks to you that I was able to push my work and my critical thinking forward; thanks to you, I now find myself standing in the light at the end of the tunnel. Dr. William Sandoval, thank you for taking interest in my work, for seeing the potential in it, and for lending your support at a very crucial moment in my educational trajectory. Dr. Robert Gould, thank you for welcoming me to the Energize project, expressing interest in my perspective and insight, and supporting me in my post-doctoral endeavors. Dr. Douglas Kellner, thank you for believing in my potential as a doctoral applicant and for your guidance and support early on as a Master's student and later as a doctoral student in SSCE. Your work and your teachings proved foundational to my developing sense-of-self as a critical scholar in the field of education and helped grow my thinking around critical theory, cultural studies, and critical media literacy. You helped foster a criticality in me that informs how I view formal and informal education—one that now inspires me to embrace interdisciplinarity and multiperspectivality in my scholarship. Dr. Thomas M. Philip, I am deeply indebted to you for the confidence you have shown in me as a critical scholar, for your unwavering support, and for your honesty. Thank you for trusting in my potential, for seeing promise when belief in my own waned. Thank you for holding me accountable to my academic and social justice goals by asking the tough questions, and for encouraging me to pursue

prestigious post-doctoral opportunities. My doctoral journey was a difficult one, one that tested my emotional, psychological, and intellectual fortitude; one that made me question myself, my ability, and my potential. At times, I didn't even recognize myself, I didn't understand how someone who was once so confident, so sure of herself, found herself asking "Why am I here...What's the point?" I am thankful to you for lending your guidance and for being honest and direct when I most needed it. I now find myself in a very privileged position: on the brink of concluding this chapter in my educational trajectory and about to start a postdoctoral fellowship thanks to your encouragement and support. For this and more, I thank you.

My doctoral research would not have been possible without my work with the "Energize" project, which set out to pursue equity goals in STEM for traditionally underrepresented students. Thank you, again, Dr. Philip for taking me on as a graduate student researcher with this project. I am also thankful to Dr. Gould and Energize project staff for designing the introduction to data science course at the core of my doctoral research, and to the National Science Foundation for their support of the Energize project. Immense gratitude also goes out to Ms. Gellar and administrative staff at "Medical Science High School" who welcomed me into their school, and allowed me to observe the implementation of IDS and interact with students. As for the students whom I had the privilege of meeting and interviewing throughout my time with Energize—thank you so very much for your willingness to share your thoughts and experiences with me. I wish you all nothing but success in your future endeavors.

To the staff and administrators of the GSE&IS Office of Student Services, thank you for the support and resources you have provided over the years. Harmeet Singh, I am especially grateful to you for your guidance and support during my time in SSCE. It is largely because of you that I was able to successfully navigate the PhD program. Thank you for your warm spirit, for your kindness, and your friendship.

I am also profoundly grateful to those who offered their mentorship, helping to shape and enrich my life over the past several years. Thank you, Dr. Alfonso Gonzales for being a positive influence and role model when I first started at UCLA as an undergraduate. Thank you, Dr. David G. García, for seeing potential in me and recommending me to the Ronald E. McNair Research Scholars Program. Your support then helped me be where I am today. I am immensely grateful to the McNair Research Scholars Program wherein my participation marked the beginning of a passion for educational research for social justice. As such, I would like to thank Dr. La'Tonya Rease Miles who was the director of McNair during my time in the program and provided much needed guidance. A big "thank you" also goes out to then assistant director of McNair, Orlando Luna—thank you, Orlando, for your mentorship throughout the years and what would blossom into friendship. Thank you for taking interest in the many changes in my life and lending an ear. I would also like to thank Dr. Pedro Nava who was my McNair mentor—thank you, Pedro, for your mentorship and humor. You knew me during a time in my life of rapid change and transformation and I'm glad to have known you and had the privilege of your counsel and friendship.

My heartfelt appreciation also goes out to Dr. Rhonda Hammer who played a large role in cementing my interests in critical media literacy once I reached graduate school. Thank you, Rhonda for encouraging me to stay centered and to think pragmatically about completing my degree. Thank you for being supportive and for being real with me. I am also thankful for the pleasure of learning from Dr. Daniel G. Solórzano—Danny, you are an inspiration to your students and I am very fortunate have experienced sitting in your classroom. Thank you for the passion with which you pursue social justice through scholarship, thank you for your love of humanity, and thank you for your life-long dedication to addressing issues of educational injustice for Latinos and other marginalized groups.

Thank you, also, to the many friends I've made along the way. To Mañanita Paez, Dr. Argelia Lara, Dr. Tolu Bamishigbin, Eliana Nevarez-Jussen, Dr. Dimpal Jain, Dr. Ryan Santos, Dr. Lluliana Alonso, Dr. José M. Aguilar-Hernandez, Dr. Anthony O'Campo, Dr. Nancy Acevedo-Gil, Dr. Bert Cuevas, Dr. Daniel Lou, Dr. Janet Rocha, Christine Vega, Cristina Medina, and Erica Soto—thank you all for being a memorable part of my doctoral journey and for helping to shape, in one form or another, the friend, scholar, partner, and mother that I am today. Dr. Michaela J. López Mares-Tamayo: thank you so much for your friendship and encouragement. Thank you for including me in your life and welcoming me into your circle, I'm so glad that life brought us together and that I can call you my friend. Many of you have gone on to do incredible work and I consider myself very lucky to have you as role models.

I must also acknowledge three very good friends and cohort members who have served as mentors/mentees, friends, and family. To Johnny Ramirez, Liliana Islas, and Iris Lucero—thank you for your friendship, for our talks and meetups, and for being the realest. Johnny, thank you for being a friend from the very start of this PhD. We've been through ups and downs together during our time in academia, and despite it all, we've encouraged each other to keep going, to keep at it, because our presence and our mission matters. Our conversations have proven therapeutic and have given me strength, and I can only hope that the same goes for you. Lily, thank you for being your fabulous, humorous, critical, and intellectual self and for reminding me to keep going, to keep pushing. Sitting down and working together over the past year, whether online or in person, truly helped me get here. Thank you for checking in on me, for your encouraging words, for reminding me what I'm capable of. Thank you for trusting me, for embracing my friendship, and proving to be a breath of fresh air during a time riddled with anxiety, stress, work, work, and more work. From the bottom of my heart, thank you, Boo. To my *comadre*, Iris—I don't know where to start. We started this journey together and have shared both rewarding and painful experiences in the academy and I am thankful to have had you by my side during these times. The ivory tower can be a lonely, isolating place, but I had the honor of walking its halls with my *comadre* by my side. I have fond memories of us sitting for hours on end at Iso Café, cranking away toward a higher education and toward a better life for ourselves and our families. I remember enduring the stress of quals together, giving each other feedback, encouraging each other, backing one another when our spirits were low. I am thankful to you beyond measure and for countless reasons. You've been a mentor, a sister, a friend, a companion, and a confidante. Suffice it to say that I doubt I would be where I am and who I am were it not for you. Most sincerely, *comadre*, thank you.

I must also acknowledge a group of individuals who have inspired me to be a better person, to educate myself, and to pursue social justice through educational research: my family. I want to thank my nieces and nephews for being a source of motivation for improving educational opportunities afforded to Latinos. Thank you, Eleanor Muñoz, Jessica Muñoz, Jose de Jesus Muñoz, Jr., Gabriela Muñoz, Andres Muñoz, German Alcaraz, Noemi Alcaraz, Angel Alcaraz, Cielo Alcaraz, Gabriel Alcaraz, Jazmin Muños, Jacqueline Muños, Nayeli Muños, Antoine Gutiérrez, Lauren Gutiérrez, Donovan Brown, Alexander Lupercio, Micah Rocha, Jenessa Sanchez, Vincent Sanchez, and Emma Sanchez for being the love at the core of my work. I have made an honest effort to be a positive role model to you all and hope that you will work your best to reach your goals knowing that I'm here to offer any help and guidance along the way. I want to acknowledge and thank my siblings for showing their support in various ways when I most needed it. Here's to you Jose de Jesus "Chuy" Muñoz, Jose Trinidad "Trino" Muños, Guillermina "Guille" Muñoz, Janette Olivares, Lorena Olivares, and Pedro Francisco Olivares. Chuy, you have been like a second father to me, constantly encouraging me to keep going, to believe in myself, and to know my worth. You have given me strength when I needed it. I have cried to you, expressed my frustrations and fears. You've listened and been the big brother I needed. You've kept me grounded and headstrong and proven to be a guiding light. *Muchas gracias, Hermano, por ofrecer palabras duras y sinceras cuando más las necesito. Por eso y más, te aprecio y te doy las gracias*. To Lorena (A.K.A. "Boo Boo"), thank you for answering my texts all hours of the day and night, except when you don't. Thanks for forwarding ridiculous videos that oftentimes bring humor into high-pressure moments in my day. Also, and most importantly, thanks being there for me and for being proud of me (because I know you are ;)).

I also want to thank my sisters-in-law, Monica Pasillas and Ivonne Sanchez—thank you for welcoming me into your family and encouraging me along the way. To my mother-in-law, Raquel Bañuelos, thank you for opening your doors and your heart to me. I could not have wished for a more loving *suegra*. *Gracias por escucharme, por confiar en mí, por platicar sobre cosas de la vida, y por ser mi segunda madre. La quiero y la aprecio, ¡mucho! Gracias por su apoyo, su amor, y su amistad.* I would also like to thank my aunt-in-law, Ofelia Bañuelos—*gracias por su cariño y por darme ánimo para terminar mi doctorado. Gracias por desearme lo mejor y por celebrar los triunfos de mi pequeña familia.*

There is yet another person who has been with me, through thick and thin, from the very start of this journey, my husband, Miguel Alberto Pasillas. Miguel, thank you for cheering me on through this process. It wasn't always easy for us and life got pretty hard as we both struggled to get through grad school while raising a rambunctious child. Your graduation day came and I know you couldn't have done it without me, because I couldn't have reached this point without you. We're like-minded in many ways, you and I, but we're also very different and I've come to learn that we balance each other out. Thank you for being my other half, for teaching me to see the world from a different angle, for bringing music and mariachi into our home, and simply for being you. Thank you for making my coffee just the way I like it and setting it by my desk as I worked on this manuscript. Thank you for picking Sofie up from school so I could keep typing away. Thank you for making dinner, walking the dogs, and all the other little things you did that meant the world. We make a great team and we're raising a strong, intelligent, and kind girl who knows that she is capable of anything. Thank you and I love you.

To my wonder-ful child, Sofía Guillermina Pasillas-Olivares: thank you for making me a better and stronger person. You have taught me to keep going at all costs, because everything I do, I do for you. The moment you were born, my life changed, my passion and motivation changed because from that point on everything I did or did not do with my life would prove consequential for yours. You made me hunger even more for social justice, you made me see my scholarship as something that will influence your world and your educational experiences. You made me a better researcher, simply by asking "Why?" It is thanks to you that I learned to articulate ideas, justifications, and explanations that I often took for granted. You inspire me, Sofía—you are brave, you speak your mind, and you wonder—you always wonder. You love to learn, and you make me want to be part of a larger effort to improve educational opportunities for children like you and your cousins so that you can continue to wonder and learn and make this world a better place. Thank you, Sofia, for teaching me things that no college course has ever taught me. *Mamá* loves you.

Lastly, but most importantly, I would like to acknowledge my parents, Leonor Cerda and Pedro Olivares as my pillars of emotional fortitude through it all. I find myself here thanks to the undying support of my mother and father, who instilled in me a strong sense of hope for the future, resiliency in the face of hardship, and self-respect. These are the working-class lessons that my mom and dad imparted on my siblings and me: "*Su educación es la única herencia que les podemos dejar*" (your education is the only inheritance we can bestow); "*Cosas materiales van y vienen, pero su educación nadie se la podrá quitar*" (material things will come and go, but your education is something no one can ever take from you); "*Trabajamos con la espalda para que ustedes trabajen con el cerebro y tengan mejor vida*" (we work with our backs so you can work with your brain and make a better life for yourself). These words resonated in my mind and fueled my spirit throughout this journey and were constant reminders that I carry the struggles and sacrifices of my parents everywhere I go. *Amá y Apá, humildemente les ofrezco el resultado de nuestros esfuerzos, esta disertación. Ustedes llegaron a este país con muy poco, pero con una esperanza para algo mejor y una gran fortaleza para sobrevivir y sobresalir. Ha sido mi meta demostrarles que sus sacrificios no han sido en vano. Mis padres, son ustedes mi orgullo. Humildemente les doy gracias por su apoyo, su consejo, y su amor*. Love you, Mom & Dad.

VITA

MARÍA CONCEPCIÓN OLIVARES PASILLAS

EDUCATION

University of California, Los Angeles Los Angeles, CA
Master of Arts, Education 2011

University of California, Los Angeles Los Angeles, CA
Bachelor of Arts, Chicano Studies and History 2009

RESEARCH EXPERIENCE

University of California, Los Angeles	Los Angeles, CA
Graduate Student Researcher, Center X	Fall 2013-Spring 2017
Research Assistant, Center for the Study of Urban Poverty	Fall 2009-Fall 2010
Undergraduate Researcher, McNair Research Scholars Program	Summer 2007-Spring 2009
Research Fellow, Research Rookies	Spring 2005

HONORS/AWARDS

Graduate Division University Fellowship	2016
George Kneller Prize	2016
Graduate Summer Research Mentorship Program	2012
Graduate Division University Fellowship	2011
Departmental Honors, Cesar E. Chavez Center for Chicana and Chicano Studies	2009
Edith and Lew Wasserman Scholarship	2009
Spring Dean's Honor List	2008
Oral History Association Annual Meeting Scholarship	2008
UCLA Vice-Provost Recognition Award Westwind/Aleph Conference	2007-2008
Winter Dean's Honor List	2007
UCLA Vice-Provost Recognition Award Westwind/Aleph Conference	2004-2005

PUBLICATIONS

Philip, T.M., Rocha, J., & **Olivares-Pasillas, M. C.** (2017). Supporting Teachers of Color as they Negotiate Classroom Pedagogies of Race: A Case Study of a Teacher's Struggle with "Friendly-Fire" Racism. *Teacher Education Quarterly 44*(1), 59-79.

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 & Instruction 34*(4), 361-388.

Philip, T.M. & **Olivares-Pasillas, M.C.** (2016). Learning Technologies and Educational Equity: Charting Alternatives to the Troubling Pattern of Big Promises with Dismal Results. *Teachers College Record*, Date Published: August 24, 2016 http://www.tcrecord.org ID Number: 21616.

Valenzuela, A.J., & **Olivares Pasillas, M. C.** (2011). *Resilience and Masculinity in Day Labor: Implications for Mental Health*. In Noguera, Hurtado & Fergus (Eds.), Understanding the Disenfranchisement of Latino Males: Contemporary Perspectives on Cultural and Structural

Factors: Routledge.

Olivares, M.C. (2007) “He Wears Wristbands” published in:
García, D. G. (2008). Transformations Through Tearto: Culture Clash in a Chicana/o History Classroom. *Radical History Review*, *102*(2008), 111-130.

CONFERENCE PARTICIPATION/SCHOLARLY PRESENTATIONS

Philip, T.M., & Olivares Pasillas, M.C. (2017). Improvising with technology for expanded meaning making: Possibilities and limitations in the classroom. Paper presented at the annual meeting of the Jean Piaget Society, San Francisco, CA.

Philip, T.M, Olivares Pasillas, M.C., & Rocha, J. (2014). Learning about Big Data for democratic participation. Paper presented at the annual meeting of the American Educational Research Association (AERA), Philadelphia, PA.

Philip, T.M, Rocha, J., & Olivares Pasillas, M.C. (2014). The inadvertent consequences of curricular reform in urban schools that cursorily appropriate social justice frameworks. Paper presented at the annual meeting of the American Educational Research Association (AERA), Philadelphia, PA.

Olivares Pasillas, M. C. & Lucero, I. (2012). Como te ven, te tratan: A critical race analysis of Chicana eyebrows and “masked” resistance. Paper presented at the annual meeting of the American Educational Research Association (AERA), Vancouver. BC, Canada.

Olivares Pasillas, M. C. & Lucero I. (2011). A critical race analysis of Chicana eyebrows and “masked” resistance. Paper presented at the annual meeting of the Mujeres Activas En Letras y Cambio Social (MALCS), Los Angeles, CA.

Olivares Pasillas, M. C., Ramirez, J., Dorman, J., & Jheng, J. Y. (2010). Crit-icizing Hip Hop. Documentary presented at project screening for course entitled Introduction to Education and the Social Sciences, Los Angeles, CA.

Olivares Pasillas, M. C., Ramirez, J., Dorman, J., & Jheng, J. Y. (2010). Is this for reals? Hip hop music analysis presented to and discussed with youth from the Pico Youth and Family Center (PYFC), Santa Monica, CA

Olivares, M. C. (2009). Domestic Rasquachismo as an improvisational aesthetic. UCLA, Chicana/o Studies Departmental Honors theses presentation, Los Angeles, CA.

Olivares, M. C. (2009). Domestic rasquachismo as an improvisational aesthetic. Presentation of research at McNair Community Research Symposium: Los Angeles History and Culture, Los Angeles, CA.

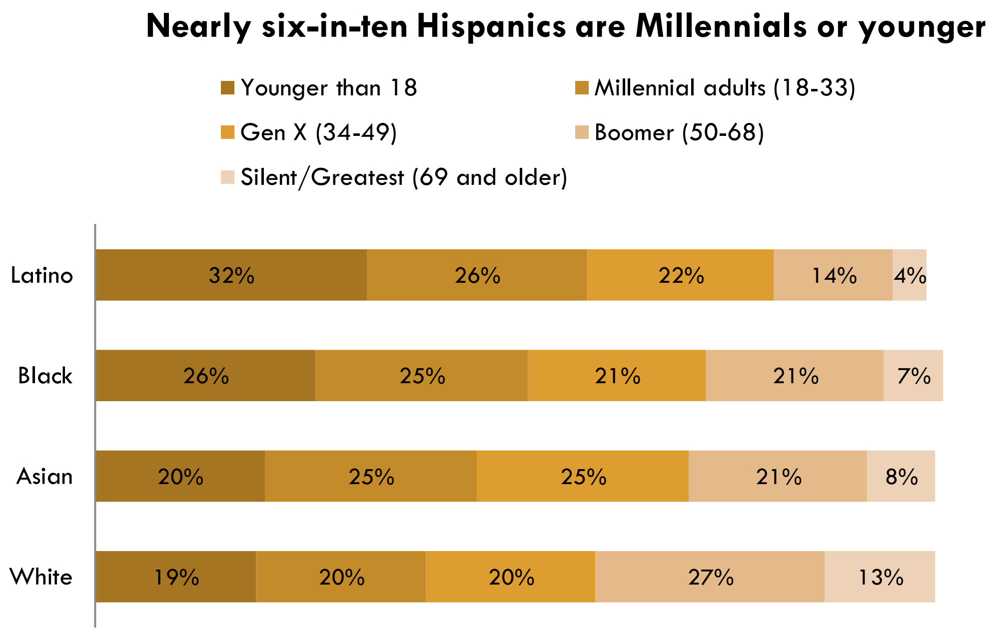
Olivares, M. C. (2008). Domestic rasquachismo. Paper at the Ronald E. McNair Research Scholars Summer Symposium, Berkeley, CA.

Olivares, M. C. (2008). “Domestic rasquachismo: An improvisational aesthetic.” Paper presented at UCLA Research Scholars Summer Symposium. Los Angeles, CA.

CHAPTER ONE
The STEM Imperative

In 2002, then president George W. Bush announced his promotion of the No Child Left Behind Act, a "framework for bipartisan education reform that…called for bipartisan solutions based on accountability, choice, and flexibility in Federal education programs" (2003). A main feature of the law was the call for increased accountability for student performance via mandated annual student testing. In 2004, the Bush Administration also put forth a series of measures known as "A New Generation of American Innovation." The primary goal of this agenda was to promote American technological innovation by focusing on three areas of technological development: hydrogen fuel, health information, and broadband (Promoting Technology and Innovation: President Bush's Technology Agenda).

Over a decade later, the United States continues to push for technological innovation and education reform that will increase the presence of women and underrepresented minorities in the fields of science, technology, engineering, and mathematics (STEM). Today's national STEM efforts are characterized by an urgency to remain globally competitive; diversify STEM careers and education; and capitalize on the US minority population which has reached unprecedented levels, slated to represent upwards of 50% of the total US population by 2050 (U.S. Census Bureau as cited in ASHE Higher Education Report, 2011). The Latino presence alone now accounts for over 17.6% of the U.S. population (The United States Census Bureau, 2015). Latinos are easily the largest and fastest growing ethnic minority in the US, expected to account for "60 percent of our nation's population growth between 2005 and 2050," (White House Initiative on Educational Excellence for Hispanics). Moreover, not only are Latinos now the largest minority groups in the U.S., as the youngest ethnic group in the country, Latinos are characterized by their youth (See Figure 1.1), making improvements to K-12 education more pressing than ever before and vital to the future of STEM fields and innovation.



NOTE: Whites, Blacks and Asians include only those who are single race and not Hispanic. Hispanics are of any race. Figures may not add to 100% due to rounding. Source: Pew Research Center analysis of 2014 American Community Survey (IPUMS). “The Nation’s Latino Population is Defined by its Youth”

Figure 1.1 “Nearly six-in-ten Hispanics are Millennials or younger” is a replication of the table provided by the Pew Research Center: Hispanic Trends retrieved from http://www.pewhispanic.org/2016/04/20/the-nations-latino-population-is-defined-by-its-youth/ph_2016-04-20_latinoyouth-01/

The urgent call to the improve STEM education and career opportunities for Latinos positions the fastest growing and largest racial/ethnic minority group as “vastly underused resources and a lost opportunity for meeting our nation’s technology needs” (National Academy of Sciences, National Academy of Engineering, and the Institute of Medicine, 2011 as cited in Gonzalez & Kuenzi, 2012). Increasing the presence of women and underrepresented groups in STEM is imperative not only to ensuring a thriving tech and engineering industry, but also to upholding a global perception of the United States’ economic prosperity and military prowess (Gonzalez and Kuenzi, 2012). Underscoring a national focus on global perception of American prosperity and power, The Business-Higher Education Forum stated that “increased global competition, lackluster performance in mathematics and science education, and a lack of national focus on renewing its science and technology infrastructure have created a new economic and technological vulnerability as serious as any military or terrorist threat" (Cited in ASHE Higher Education Report, 2011). As a result, there is a clear and general agreement over "problems posed by racial, ethnic, and gender disparities in STEM education and employment [yet this] has not translated into widespread agreement on either the causes of underrepresentation or policy solutions" (Gonzalez and Kuenzi, 2012, p. 24). Capitalizing on "lost talent" and preventing further losses in STEM is essential to maintaining the nation's economic viability, but we must also examine the nuances that underpin this lack of representation and participation in ways that *humanize* so-called "underused resources" by acknowledging, valuing, and incorporating epistemologies and learning processes of non-dominant groups by and large excluded from notions of legitimate STEM learning and doing. Despite decades-long efforts to both improve educational outcomes for students of color and diversify STEM fields, the lack of quality educational opportunities for students of color is emblematic of deep-seated and long-standing educational inequity that characterizes the American education system today. It is important to note that although educational inequity affects a number of non-dominant groups including African-American and Latino students, I am particularly oriented toward a focus on Latino students as the largest and fastest growing non-dominant group in the U.S. In an effort to contextualize Latino educational attainment in the U.S., in the next section I will discuss pressing issues related to Latino education and educational attainment in K-12 and beyond, with a particular focus on STEM education and equity-oriented initiatives that seek to increase access to STEM for historically non-dominant groups.

Latinos and Educational Inequity

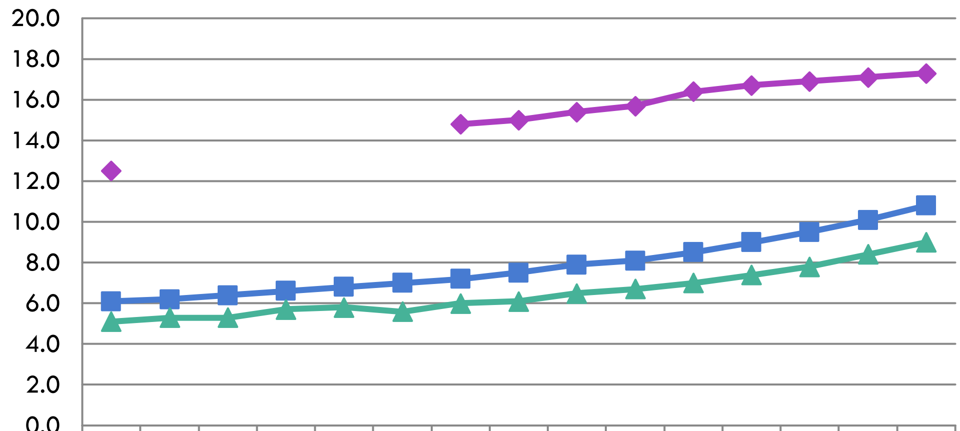
Studies indicate that attrition rates for Latinos have dropped to an all-time-low. For Latinos ages 18-24, the proportion of those who left high school without a high school diploma or equivalent decreased from 33 percent in 1993 to 12 percent by October of 2015 (Krogstad, 2016). The significance of this drop is further magnified by the fact that while the rate for Latinos dropped 21 percentage points, drop-out rates for Black, White, and Asian students dropped by only single-digits—nine, four, and four percentage points, respectively (Krogstad, 2016). While it is important to acknowledge the strides made in reducing attrition rates among Latinos, Latinos continue to withdraw from high school at higher rates than all other ethnic groups (Krogstad, 2016; Covarrubias, 2011). Of those who enroll in college, about 17 out of 100 enroll at a public two-year college, and about 18 enroll at a four-year-institution in pursuit of a bachelor's degree (Krogstad, 2016). Ultimately, only 15 out of 100 Latinos ages 25-29 complete college with a bachelor's degree or higher (Krogstad, 2016).

Furthermore, data collected by the National Center for Education Statistics (NCES) indicate that out of all bachelor's degrees conferred by postsecondary institutions in 2014, an estimated 10.8 percent were conferred on Latinos, up from 6.8 percent a decade earlier. With regards to bachelor degrees in STEM[1] in particular, only 9 percent were awarded to Latino students in 2014, up from 5.8 percent in 2004 (NCES). Although on the rise, these figures must be measured against the proportion of Latinos in the U.S. population and can only be considered "equitable" when gains achieve parity with the presence of Latinos nationwide (Pérez Huber, Vélez, and Solórzano, 2014). U.S. Census Bureau population estimates indicate that as of 2014,

[1] I calculated STEM field totals by adding degrees conferred for the following major fields of study according to NCES tabulations: biological and biomedical sciences; computer and information sciences; engineering; engineering technologies and engineering related fields; mathematics and statistics; and physical sciences and science technologies. For an example of 2012-2013 and 2013-2014 tabulations see the NCES—Digest of Education Statistics website https://nces.ed.gov/programs/digest/d15/tables/dt15_322.30.asp

Latinos represent 17.3 percent of the total population (see Figure 1.2). Given the rapid nation-wide growth of Latinos in the U.S., slow growth in educational gains will not suffice in closing an educational gap that is sure to widen.

Bachelor Degrees Conferred on Latinos by Postsecondary Institutions in Percentages, 2000-2014



	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Latino Share of the Population	12.5						14.8	15.0	15.4	15.7	16.4	16.7	16.9	17.1	17.3
Earned Bachelor's Degree (All fields)	6.1	6.2	6.4	6.6	6.8	7.0	7.2	7.5	7.9	8.1	8.5	9.0	9.5	10.1	10.8
Earned Bachelor's Degree in STEM	5.1	5.3	5.3	5.7	5.8	5.6	6.0	6.1	6.5	6.7	7.0	7.4	7.8	8.4	9.0

Figure 1.2 Source: For bachelor's degrees conferred, see National Center for Education Statistics—Digest of Education Statistics, retrieved from https://nces.ed.gov/pubsearch/getpubcats.asp?sid=091#061. Data on the Latino share of the population reflects figures reported in Stepler and Brown's (2016) "Statistical Portrait of Hispanics in the United States," retrieved from http://www.pewhispanic.org/2016/04/19/statistical-portrait-of-hispanics-in-the-united-states-key-charts/#hispanic-rising-share. Please note that Stepler and Brown (2016) do not provide statistical data for the Latino population from 2001 to 2005.

Figure 1.2 above provides a proportional comparison of the percentage of Latinos in the U.S. population against the percentage of bachelor's degrees, both for all fields of study and STEM fields in particular, awarded to Latinos from 2000-2014. A promising finding here is that the estimated growth rate for bachelor's degrees conferred on Latinos is higher (11.4 percent increase per year) than the estimated growth rate of the Latino population nationwide (2.4 percent increase per year). In other words, given the growth rate, it is possible that over time Latinos will earn bachelor's degrees in proportion to their national presence and eventually achieve degree conferral rates that surpass their national presence. Another, less promising, finding is that the growth rate for Latinos earning bachelor's degrees in STEM (an estimated .9 percent yearly increase) is nearly three times less than the Latino population growth rate and nearly 13 times less than the growth rate for Latinos earning bachelor's degrees overall. Moreover, from 2000-2015 bachelor degrees conferred on Latinos have been predominantly concentrated in the field of business, which alone exceeds the number of degrees conferred to Latinos in all STEM fields. An exploration of why Latinos have earned more bachelor degrees in business and not STEM fields is beyond the scope of this study, but these findings indicate two things: firstly, for more than over a decade, Latinos have consistently and overwhelmingly earned bachelor degrees in a particular field unrelated to STEM; and secondly, postsecondary educational attainment in STEM has not seen substantial changes—despite the widely acknowledged imperative to increase the presence of Latinos in these fields.

To understand the factors at play here, we need to look at educational opportunities and resources available to Latino students prior to entering postsecondary education. At the postsecondary level, STEM courses show a pattern of negatively affecting continued interest and motivation of students of color—this is particularly the case for "gateway" courses like Chemistry 101, known colloquially by students as "weeder courses" (Barr, Gonzalez, and Wanat, 2008; Mervis, 2010). For example, a study that measured changes in student interest in premedical study notes significant decreases in interest among women, African American, and Latino students surveyed. The study asserts,

> the negative influence of chemistry courses on continued interest in premed is experienced more so by women and URM [underrepresented minority] students. In light of the fact that 74% of URM premedical students at Stanford are women, it is not at all surprising that Stanford's URM students are less than half as likely as non-URM students

to persist in their interest in premed and eventually apply to medical school. (Barr et al., 2008, p. 510-511)

Thus, unless students have a strong STEM-related background that precedes postsecondary education, the pursuit of STEM-related study at the postsecondary level is not likely to prove promising or welcoming. Thus, it is necessary to look at issues of access, exposure, and quality learning opportunities in STEM available to Latino students at the primary and secondary levels of education.

Examining the underrepresentation of women and people of color in computer science professions in California, a report from the Level Playing Field Institute (Martin, McAlear, and Scott, 2015) posits that one of the leading causes of underrepresentation is the lack of access to computer science courses in public high schools. Their study finds that the availability of computer science courses, any computer science course including AP computer science, is higher in public schools with the lowest presence of students of color (0-50 percent), while the inverse is true for schools with the highest presence of students of color (Martin et al., 2015). A similar pattern was found in schools with the highest percentage of low-income[2] students in the total student body. Only four percent of all schools with a 76-100 percent low income student population offered AP computer science, compared to 43 percent of all schools with a 1-25 percent low-income student population (Martin et al., 2015). Figure 1.3 provides a breakdown of the availability of computer science courses in California public high schools based on the percentage of low-income students in the student body.

[2] Martin et al. (2015) define “low income” as “Free/Reduced Price Lunch eligibility (through federally-determined poverty guidelines) for the National School Lunch Program” (p. 21).

Availability of Computer Science Courses in California Public Schools by Percentage of Low-Income Students					
Percentage Low-Income students in total student body	Number of CA public high schools	Number and percent of schools offering AP Computer Science		Number and percent of schools offering *any* computer science	
1-25%	198	85	**43%**	120	**61%**
26-50%	305	43	**14%**	101	**33%**
51-75%	403	33	**8%**	130	**32%**
76-100%	378	14	**4%**	92	**24%**

Figure 1.3 Source: This table is a replication of the table presented in Martin et al. (2015, p. 21)

These figures highlight the disparities in educational opportunities for low-income high school students and indicate that schools with fewer low-income students provide their student body with greater opportunities to engage in some sort of computer science study. Among the largest school districts in California, the top three with the highest percentage of underrepresented students of color enrolled a combined three percent of students in the districts in some sort of computer science course. With regards to AP Computer Science course offerings in particular, schools with the lowest percentage of low-income students offered AP Computer Science at a rate (43%) that exceeds that of all schools with 26-100 percent low-income students put together (26%) (Martin et al., 2015). This finding indicates that increasing the presence of Latino students in STEM requires complex and intersectional approaches to address complex and intersectional issues that go beyond a simple lack of tech resources in the classroom.

While some STEM reform efforts focus on providing low-resourced schools with new technologies (Gazzar and Jones, 2014; Szymanski, 2015; Yarbrough, 2014; Philip and Garcia, 2013a, 2013b, 2014; Margolis and Suarez Orozco, 2014), others introduce students to innovative STEM curriculum and experiences in hopes of piquing student interest in STEM at a young age and ultimately lead to equitable outcomes and participation in STEM fields. There is much to be said for introducing students to new technology, resources, and curriculum, but STEM equity scholars argue for the need to carefully consider assumptions and underpinnings that frame STEM equity initiatives. Scholars argue that reframing how we view scientific inquiry; what counts as scientific knowledge; and who can do science is foundational to promoting actual equity in STEM education (Brickhouse, 1994; Stanley and Brickhouse, 1994; Calabrese Barton, 1998; Calabrese Barton & Yang, 2000). The same can be said for new data science and data science-oriented programs that have begun to indicate positivist continuities with parent fields of computer science and mathematics (Ebach, Michael, Shaw, Goff, Murphy, and Matthews, 2016; Ekbia, Mattioli, Kouper, Arave, Ghazinejad, Bowman, Suri, Tsou, Weingart, and Sugimoto, 2015; Couldry, 2014). Genuine efforts aimed at improving Latino education in the 21st century must acknowledge and address the causes of historical educational inequity; and promote the learning of critical literacies necessary for democratic participation in a tech-driven society (Philip & Garcia, 2013a; Philip, Olivares-Pasillas, & Rocha, 2016; Philip, Schuler-Brown, & Way, 2013).

In this dissertation, I draw from scholarship critical of traditional STEM fields and their positivist roots to contextualize my understanding of dominant treatments of STEM knowledge and scientific inquiry in schools. I recognize that data science differs from traditional STEM fields like biology and physics in that it draws heavily from mathematics and computer science, but at the core of data science lies a strong fidelity to positivist treatments of knowledge and the scientific method foundational in traditional STEM fields (Ebach et al., 2016; Ekbia et al., 2015). Given that data science is a relatively new field of study undergoing rapid development due in large part to incessant technological innovation and evolution, I find it highly necessary to contextualize data science as a hybridized extension of traditional STEM fields which center positivist treatments of knowledge, inquiry, and scientific discovery. In Chapter Two, I will review existing literature as it pertains to fields that focus on the study and use of data and attest to the persistence of the myth of objectivity within data science-related fields. In this chapter, however, I pull from existing scholarship that addresses issues of equity in traditional school science for the purpose of introducing and situating my research as a critical project for STEM equity in the 21st century. I will expound on this work in the context of the myth of scientific objectivity in Chapter Two as well.

Statement of the Problem

Traditional notions of science and, necessarily, school science have directly contributed to the exclusion of women and people of color in STEM and in newer fields like data and computer science which build on traditional foundations of science to examine technological phenomena and foster technological innovation. According to Brickhouse (1994), the main reasons that account for the scant representation of non-dominant groups in the sciences are 1) deficit thinking about the intellectual abilities of non-dominant groups, and 2) their unfair treatment in schools. She argues that while well-intentioned, programs aimed at increasing the participation of women and people of color in science remain shortsighted in their treatment of science and scientific knowledge as value-free. Calabrese Barton (1998) finds that even in science teaching that strives for cultural relevance, “it is the teaching methods and applications of science that are challenged, not the underlying scientific concepts and principles,” which are treated as enduring and objective rather than epistemologically informed (p. 529). The very framing of science and scientific knowledge—premised on a white, male, middle-upper class epistemology—upholds harmful views of excluded groups, their intellectual abilities, and their epistemologies (Brickhouse, 1994; Stanley & Brickhouse, 1994; Calabrese Barton, 1998; Calabrese Barton & Yang, 2000).

For example, in their deconstruction of the “culture of power” embedded in school science and the implications that it has for the STEM education of and opportunities for non-dominant groups, Calabrese Barton and Yang (2000) argue that hegemonic notions of what counts as legitimate forms of science education undermine rich scientific learning opportunities for non-dominant students in school, and also delegitimize rich scientific learning that these students engage in outside of school. Citing Delpit (1988), Calabrese Barton and Yang (2000) write that “[t]he ‘culture of power’ represents a set of values, beliefs, ways of acting and being that for sociopolitical reasons, unfairly and unevenly elevate…people” from dominant groups and subordinate those from non-dominant groups such as women, people of color, and those from low socioeconomic backgrounds (p. 873).

Examining how the culture of power influences education in general, particularly science education, for non-dominant students, Calabrese Barton and Yang (2000) examine the educational experiences of and opportunities afforded to Miguel, a young Puerto Rican father living in a homeless shelter with his wife and daughter. Despite gaining personally meaningful scientific experiences outside of school, these experiences and his growing enthusiasm for science and nature were “neither acknowledged formally by his teachers nor cultivated in school” (Calabrese Barton & Yang, 2000, p. 872). For Miguel, not only did the culture of power that pervades school science education invalidate his out-of-school experiences in science and contribute to his leaving high school; it also instilled in him problematic notions regarding who can become a scientist, what counts as legitimate scientific knowledge, and that his own Puerto Rican culture was deficient for not promoting American assimilation via hegemonic standards for academic success, financial stability, and autonomy (Calabrese Barton & Yang, 2000).

The notion of an objectivist science, shaped and maintained by the existence of the culture of power, not only affects the educational opportunities afforded to non-dominant students, but also leaves long-lasting imprints of the nature of legitimate education and knowledge among these very students. The lack of criticality on the framing of scientific knowledge and inquiry supports the continued exclusion of particular groups from access to and equity in STEM education in ways that prove self-perpetuating, self-renewing, and detrimental to the lives and livelihoods of non-dominant groups (Stanley & Brickhouse, 1994). Moreover, the value-free framing of science has served to legitimate deficit views of the intellectual potentialities and capabilities of women and people of color, upholding the validity of standardized testing, and promoting tracking programs that have capped learning opportunities for many (Delgado Bernal, 1999; Oakes, 1985).

The role of scientific objectivity in the perpetuation of educational inequity. By simultaneously obscuring and protecting the culture of power, the myth of science as objective endures as the core of standardized assessments premised on dominant understandings of what counts as valued and legitimate knowledge. Thus, the hegemonic ideology that informs the culture of power leads to the design of social institutions as projects that preserve the dominance of one homogenous group through the subjugation of all others. When we consider that the culture of power has long presided over science, technology, engineering, and mathematics, it becomes clear that the science that is overwhelmingly taught in schools and universities is not a science that is designed to value the ways of knowing and processes of learning of non-dominant groups. For this reason, critical scholars in the learning sciences are confronting the messiness and “desettling” settled expectations in STEM education (Bang, Warren, Rosebery, and Medin, 2012).

Borrowing from Harris' (1995) construct of "settled expectations" in critical race theory, Bang et al. (2012) apply this notion to schooling in reference to the deep-seated and normalized delineations of "acceptable meanings and meaning-making practices" in the classroom that while upholding understandings and ways of knowing of dominant groups, shape "deficit-oriented discourses concerning students from nondominant communities…[which] control the scope of what constitutes an acceptable explanation, argument, or analysis; what 'smart' looks and sounds like; whose narratives and experiences are valued and for what purposes" (p. 303). Settled expectations in STEM education are ideological in nature[3] in that they promote valued (i.e. traditional) forms of scientific knowledge and ways of doing science through education, and "devalue and dismiss boundary expanding forms of knowledge, experience, and meaning-making with which students approach scientific phenomena" (Bang et al., 2012, p. 304). This means that scientific discourses and practices that students like Miguel engage in are, by default, excluded from acceptable forms of school science in the absence of desettling interventions.

Rosebery, Warren, and Tucker-Raymond (2016) provide another pointed example of how traditional school science, as a project of the culture of power delineated by settled expectations, is designed to exclude, and thus devalue, the epistemologies and learning discourses of students from non-dominant groups. They caution that sense-making practices of non-dominant groups can be misinterpreted in school settings in problematic ways, to the detriment of students and student learning (Rosebery et al., 2016). In their own work with Haitian American students, they found that what might be misconstrued as a verbal altercation among students in the classroom is actually "a form of intellectual theatre organized around claims and evidence [where] [s]tudents express, defend, and dispute various points of view on a question, deploying evidence and logic in a process [they] found similar to agreement and disagreement sequences documented in

[3] This is not to say that settled expectations are not ideological in other disciplines.

professional scientific activity" (Rosebery et al., 2016, p. 1574; Warren & Rosebery, 1996). By adopting an "expansive view of human meaning-making as fundamentally heterogeneous and multivoiced, both within and between socially and historically constituted communities" (Bang et al., 2012), the project sought to desettle settled expectations established for school science by capitalizing on the ways of knowing and discourse practices that were central to the cultural lives of these students as members of a non-dominant group while exploring two science curricula on water. The teacher's valuation and respect for students, their communities, and their intellectual contributions to the science classroom allowed students to go beyond the curriculum's focus on water consumption and conservation and bring critical issues relevant to their lives into classroom discourse, emerging unanticipated and dynamic understandings of water. Significantly, by embracing emergent learning negotiated by the teacher and students in the classroom, students were able to begin thinking about the nature of water in ways that challenged what Bang et al. (2012) refer to as the settled "nature-culture divide" that pervades the sciences.

Closely related to settled expectations, the nature-culture divide refers to settled beliefs that nature—that which science ultimately seeks to understand—exists apart from human culture and that the scientific *meanings* ascribed to objects of nature and natural phenomena are true, static, and have no connection to nor are they shaped by culture and cultural life. "Meaning," however, must be conceptualized and constructed by individuals and is thus, in fact, influenced by culture. For Bang et al. (2012), desettling the nature-culture divide "entails imagining multivoiced meanings of core phenomena as open territory for sense-making in the science classroom, similar to the kinds of meaning-making opportunities that are available to scientists in the field" (p. 308). Thus, Bang et al.'s work supports the call for and centering of dialogue in the

Freirean sense to both make science more equitable and to expand our knowledge and interpretation of scientific phenomena. "True dialogue," Freire (1970/2006) argues,

> cannot exist unless the dialoguers engage in critical thinking—thinking which discerns an indivisible solidarity between the world [nature] and the people [culture] and admits of no dichotomy between them—thinking which perceives reality as process, as transformation, rather than as a static entity—thinking which does not separate itself from action, but constantly immerses itself in temporality without fear of the risks involved. (p. 92)

This is not to say that we must do away with science as we know it, but we do need to reconceptualize what counts as science, science-doing, and science knowledge in ways that give credence to the rich learning norms and practices of students of color by encouraging and valuing classroom dialogue that welcomes the "multivoiced" contributions of students whose voices have traditionally been excluded in STEM (Brickhouse, 1994; Calabrese Barton, 1998; Calabrese Barton & Yang, 2000; Stanley & Brickhouse, 1994; Bang et al., 2012; Freire, 1970/2006). By acknowledging and valuing non-traditional learning processes, ways of doing science, and "boundary expanding knowledge," (Bang et al., 2012, p. 304) we expand opportunities for youth of color to participate in STEM in ways that are innovative. Moreover, without equitable access to rich and inclusive STEM learning opportunities, the constitutional right to democratic participation for students from non-dominant groups in a data-driven society is severely hampered (Philip et al., 2013). Given that we do not know enough about how data science, as a new field, includes or excludes diverse epistemologies, we must heed the warnings of critical scholars as they pertain to the foundations of scientific inquiry. Data science is not a field wholly new and never-before-seen. Instead, data science has emerged from established

STEM fields, and as such, is positioned to inherit foundational views of scientific inquiry and knowledge that persist in its parent fields unless there is an equity-driven effort to desettle mechanisms that function to exclude non-dominant groups and their epistemologies.

In reality, adopting equity-driven efforts toward epistemological inclusion of non-dominant groups in STEM fields is directly related to increasing opportunities for democratic participation of all citizens given that society is increasingly shaped by scientific discovery and technological innovation in the era of Big Data. Now, more than ever before, including youth from non-dominant groups in the process of scientific inquiry and knowledge construction are essential to achieving an egalitarian society that benefits and builds on dynamic understandings of ourselves and our ever-changing natural and technological world. Thus, in the next section I will engage in a discussion of the importance of STEM learning in general and of the cultivation of data-scientific literacy in particular for democratic participation in our increasingly tech-driven world in order to convey the highly complex and multi-layered pursuit of educational equity in STEM for non-dominant students in the era of Big Data.

STEM learning for democratic participation. Increasingly vital to STEM learning for democratic participation is examining the role that the scientific and technological innovation that now characterize society play in the lives of youth born into an era saturated with mobile technologies that collect an immense amount of user data. Every single student enrolled in the introduction to data science course that I observed owned a smart phone and enjoyed regular access to the Internet, not unlike American youth in general. A report by the Pew Research Center finds that youth from all ethnic groups ages 13-17 have access to and utilize mobile technologies quite regularly with 92 percent of teens reporting going online on a daily basis and 24 percent reporting being online “almost constantly” (Lenhart, Duggan, Perrin, Stepler, Rainie, and Parker, 2015). According to the same report, teen Internet use is primarily facilitated by smartphone use and ownership, and is more frequently accessed by African-American and Latino teens (Lenhart et al., 2015).

Therefore, informed by the critical views presented earlier, my study considers the ways that settled expectations in data science, in a time of rapid technological innovation and mobile technology saturation, are perceived by students from non-dominant groups through a case study of an introductory data science course. Additionally, I identify whether and how this introduction to data science curriculum, as an instantiation of reform-oriented STEM initiatives, provided avenues for the Latino students who constituted the entire classroom student body to desettle notions of what it means to do data science and be a data scientist. In making a case for the significance of developing multi-literacies for democratic participation, particularly data science literacy, I now turn to a discussion of the role of data in our technological society.

Data Science for a Data-Driven Society

We produce data close to every second of our lives; every time we update our Facebook status, use our mobile phones to locate the nearest Starbucks, save a few dollars at the grocery store by using a loyalty card, and even while we sleep by using wearable technologies to track our sleep patterns. In one way or another we are [perhaps] willing participants in the compilation of not just data, but 'Big Data'. Our constant and widespread use of internet-enabled technology has streamlined systematic data generation and sharing to the extent that researchers in numerous fields can now expeditiously access vast data repositories to reveal patterns previously unseen (Chong and Shi, 2015; Chen et al., 2016; Chen, Mao, and Liu, 2014; boyd and Crawford, 2012). Citing Manyika et al. (2011), Philip et al. (2013) write, "The colossal amount of data generated from transactions and sensors, and as byproducts from [online] activities…promises corporations potential profits and savings in the tune of hundreds of billions of dollars" (p. 104). Digital sociologists, danah boyd and Kate Crawford (2012) define Big Data as a socio-technical phenomenon that exists at the intersection of technology, analysis, and mythology and is, thus, brought into being through

> (1) maximizing computation power and algorithmic accuracy to gather, analyze, link, and compare large data sets… (2) drawing on large data sets to identify patterns in order to make economic, social, technical, and legal claims… [and] (3) the widespread belief that large data sets offer a higher form of intelligence and knowledge that can generate insights that were previously impossible, with the aura of truth, objectivity, and accuracy. (p. 663)

Additionally, Big Data is also understood as defined by the '3 Vs': volume—it consists of extremely large data sets; velocity—data are collected and streamed rapidly; and variety—data sets consist of numerous and diverse variables, data formats, and structures (McCartney, 2015; Selwyn, 2015; Zheng, Zhang, and Wang, 2014). However, as Chen, Mao, and Liu (2014) note, as of yet there is no consensus regarding a proper definition for Big Data due to a difference in opinion among professionals and researchers in different fields.

While statisticians, computer scientists, social scientists, librarians, and other professionals enjoy a privileged awareness of the ways information about all aspect of our lives are collected, the majority of the American population are oftentimes unaware of their accidental and incidental contributions to the ominous data set. This is a disturbing thought when we consider how technology—key to the generation of Big Data—saturates our lives and that of adolescent men and women.

In their analysis of data collected over the span of 15 years, the Pew Research Center found that in 2015 American adult Internet use rose to 84 percent, up from 52 percent in the year 2000 when the center first began a systematic measurement of American Internet use (Perrin and Duggan, 2015). In another study, data show that smartphone use among US adults has quickly increased from 35 percent in 2011 to 68 percent in 2015, with adults ages 18-29 reporting the most smartphone ownership (Anderson, 2015). Smartphone ownership among teenagers is also on the rise at 23 percent for those ages 12-17 in 2011 (Lenhart, 2012). Contemporary education philosopher and cultural theorist, Douglas Kellner (1998) posits that our current use of technology and its rapid development constitute “one of the most dramatic technological revolutions in history” (p. 103). On this radical shift, Kellner writes, “This Great Transformation poses tremendous challenges to education to rethink its basic tenets, to deploy the new technologies in creative and productive ways, and to restructure education in the light of the metamorphosis we are now undergoing” (p. 103). Strasburger (2015) articulates how the meaning of education and “to be educated” has not only changed, but morphed to include literacies pertaining to things like social media when he writes, “To be educated in 2015 means that someone can read, write, download, text, and perhaps even tweet” (63). To be educated in the 21st century now necessitates that students learn to read and write the word and world shaped by a widespread evolution of technology that transcends physical spaces. Education in today’s world takes place not only in the physical space of classrooms, homes, and other public spaces, but also in amorphous spaces located within the technology we carry in our very own pockets. Technology is continually evolving without cease and as such, so should our understanding of schooling, learning, and literacy.

The national STEM imperative, the pervasiveness of technology, and the Internet's omnipresence all warrant that students learn to use technology as tools for learning, but also that they think critically about the politics of these tools, their purpose, affordances, and limitations. In today's day and age, we need to encourage and support our youth in thinking critically about what types of data are collected, how they are collected, the implications that the Big Data operation has on their lives and the role it plays in perpetuating or challenging power dynamics that fuel issues of inequity. As the likely arbiters of a tech-driven society, digital-natives[4] must develop the necessary literacies if they are to achieve democratic participation, personal fulfillment, and professional success in our ever-shifting American democracy undergoing dramatic technological and demographic transformation (Philip et al., 2013; Kellner & Share, 2007; Kellner, 2003; The United States Census Bureau, 2015; White House Initiative on Educational Excellence for Hispanics).

Research Questions

My dissertation examines an instantiation of STEM reform efforts to analyze the development of student learning identities in a piloted introduction to data science course at a high school in one of the largest school districts in the country. My study is particularly concerned with identifying emergent classroom obligations and understanding whether and how they came to support and/or hinder students' opportunities to learn richly with data. Accordingly, through this study I seek to address the following research questions:

1. How did participation in a STEM-reform oriented introduction to data science classroom provide opportunities for students to think richly and critically with and about data?

[4] The term "digital natives" was first introduced by Prensky (2001) to refer to individuals born in or after 1980, after the launch of the internet and other digital technologies. This generation has also come to be known as N-Gen (net generation), D-Gen (digital generation), and Millennials.

2. How did these opportunities contribute to the development of students' critical social identities as data science-doers?

Theoretical Framework

I approach this study from the perspective that a humanistic STEM imperative is one that is not only characterized by an urgency to remain globally competitive; diversify STEM careers and education; and capitalize on the growing Latino presence. A promising and truly equitable STEM imperative must be driven by a simultaneous dedication to cultivate critical *reflection* on the politics of technology, interconnectivity, and data through critical pedagogy and purpose-driven *action* through the development of new and critical literacies to support the democratic participation of our youth in society.

As such, STEM initiatives should seek to cultivate critical data science literacies that encourage women and students of color to engage in highly reflective examinations of their increasingly tech-driven world; considerations of the practical and ulterior motives and functions of popular technology; and the implications that this technological and scientific revolution has for their daily lives and lived experiences as members of non-dominant groups. In other words, genuine efforts to revolutionize STEM education through data science-oriented initiatives must go beyond trying to change students or the curriculum, but must heed the call of critical scholars and embrace a critical pedagogical approach that guides students in critical reflection on the framing of data science, the purpose of data science education, and positions students as potential contributors to data-scientific knowledge and innovation. For Freire (1970/2006), to engage in action (i.e. changing students or the curriculum) without reflection (i.e. examining the epistemological foundations of settled scientific knowledge) is to engage in activism as "action for action's sake," wherein action, devoid of reflection, necessarily suffers and negates true praxis; that is, true transformation of reality, in this case the reality of the dire state of diversity, equity, and access to STEM for non-dominant groups and their sociohistorical educational inequity overall (p. 87-88). By demystifying scientific knowledge and science artifacts, STEM initiatives can promote the fact that students from non-dominant groups have the potential to change and create technology and other science artifacts to address issues relevant to their lives and to the needs of their communities. Instead of learning to use science tools and skills afforded to them, students will learn that artifacts are of human design, and as humans, they too possess the potential to better attune these artifacts to the heterogeneity of their lived realities.

Accordingly, my study is theoretically informed by information literacy frameworks combined with critical social theory (CST) and a critical theory of education (CTE)—frameworks that center the importance of critical pedagogy and the development of new literacies for democratic participation in an ever-shifting society dominated by technological innovation. Existing data-related literacy frameworks argue for the need to cultivate information literacy as a tool for learning to access, evaluate, use, and manage information to solve everyday problems relevant to life in a data and technology saturated age (Partnership for 21st Century Learning, 2015; Association of College & Research Libraries, 2000). CST in education is concerned with "advancing the emancipatory function of knowledge," and promotes the role of critique in cultivating students' ability to think critically about institutional and conceptual dilemmas as they relate to systems of power pervasive throughout society. CTE, as proposed by cultural theorist and education philosopher, Douglas Kellner (2003), holds that the technologically saturated and dominated society of today necessitates a democratic restructuring of education predicated on the needs of society. CTE is a poststructuralist project inclusive of critical theory aimed at acknowledging and valuing epistemologies of marginalized groups and necessitates critical analyses of disciplinary and cultural artifacts traditionally regarded as apolitical and objective. As such, an intersectional CST/CTE framework supports a call for the development of new critical literacies in data science that will allow youth to not only use the technology of today, but to understand the complexities and politics inherent in a society dependent on multimodal data-generating technology. Thus, I employ a combined information literacy and CST/CTE lens to argue for the importance of cultivating, not just information literacy, but critical data-scientific literacy.

Significance of the Study

The development of critical data-scientific literacy, or what Philip et al. (2013) term Big Data literacy, could not be of greater relevance to the lives of youth born into a revolutionary period characterized by technological innovation, constant interconnectivity, and data generation on a grand scale. Critical reform-oriented STEM initiatives, those that engage in a simultaneous and renewing process of reflection and action, can be viewed as efforts to restructure education and develop new and dynamic literacies that can foster critical analyses of technology and mobile devices. Ultimately, the cultivation of critical data science literacy can support the type of knowledge and political awareness necessary for democratic participation in a society undergoing a Great Transformation.

Given the relative newness of reform-oriented STEM efforts that specifically focus on data science literacy, it is imperative that programs and curricula aimed at transforming educational institutions toward equitable outcomes for nondominant groups be explicit about equity and mechanisms by which to achieve them in educational settings. In light of the critical views presented earlier in this chapter, STEM reform must adopt an iterative approach in line with design experimentation that can build on successful changes to STEM education and reassess the purpose of inefficient or ineffective learning interventions (Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003). Moreover, Gutiérrez and Jurow (2016) argue that educational reform efforts designed to improve educational outcomes for students from non-dominant groups must confront the intersection of educational disparities and issues of social justice to which non-dominant groups are subject to through social design experimentation "in order to make central the realities of peoples' lives because the possibilities of learning and development are deeply situated in unevenly developed historical, spatial, and social circumstances" (p. 568).

This requires a clear and honest understanding of exactly what we are attempting to transform and how students are positioned within those efforts. For example, if a STEM reform initiative seeks to attune students to data science without introspection regarding curricular content, then it is students from non-dominant groups and that which differentiates them from those from dominant groups (i.e. their race, culture, language, gender, socioeconomic status) that are positioned as that which must be transformed. Programs that follow this approach will fail in their inability or unwillingness to be critical of settled expectations in data science, thus perpetuating the mechanisms of inequity seen in its parent fields. Furthermore, while programs and initiatives can be designed with good-intentions, program evaluation must adopt a reflexive approach from inception to conclusion to evaluate what works and what needs re-working. Only by identifying that which does not work toward achieving educational equity can we address inefficiencies and build-on these programs.

Another central idea here is that learning science researchers, educators, and policymakers must embrace the role of active learners, as this is not reserved exclusively for students. This is of critical importance because it decenters, or de-settles, the notion of teachers, researcher, and policymakers as beholders of a one true knowledge. To refuse the identity of learner for non-students negates the epistemological contributions of students and the potential to think of science knowledge as multi-voiced because it establishes static learning paths that will not change to accommodate new and unanticipated lines of inquiry, dynamic and new interpretations of traditional scientific phenomena, and the lived realities of the students whom reform initiatives seek to help. Related to this last point, Freire argues that "[m]any political and educational plans have failed because their authors designed them according to their own personal views of reality, never once taking into account (except as mere objects of their actions) the *[people]-in-a-situation* to whom their program was ostensibly directed" (p. 94, emphasis in original).

This study is significant and unique in that it applies an identity lens in its analysis of data science learning outcomes, allowing us to transcend the dichotomous evaluation of whether a data science program is successful or not. By analyzing the types of identities students developed in this data science course and the learning norms and practices that contributed to the development of particular identities, we will be able to gain an intimate understanding of particular classroom processes that prove promising for the development of strong STEM identities, and specifically Critical Social Data-Scientific (CSDS) identities. What's more, this study should be viewed as a unique opportunity to learn about concrete classroom practices that foster student' abilities and opportunities to think richly with data. The study will provide a glimpse into what rich learning with data can potentially look like and serve as a building block for the development of future STEM reform initiatives that seek to cultivate a personally meaningful and socially responsible critical data scientific literacy among youth in general, and non-dominant youth in particular.

In the following chapter, I will review existing scholarship to elucidate concepts I began to touch on in this introductory chapter, beginning with the myth of scientific objectivity, the nature-culture divide, settled expectations, and the culture of power. I will then revisit the myth of scientific objectivity as it relates to data science-related fields that seek to make Big Data actionable in the section entitled "The Myth of [Data] Scientific Objectivity, revisited."

CHAPTER TWO
Literature Review

Approaches to STEM Reform

Reframing science knowledge. Critical scholars in the learning sciences argue that in order to make science more accessible to groups traditionally excluded from STEM, STEM education reform must challenge traditional views on the purity and factuality of science knowledge—that is, the myth of scientific objectivity (Brickhouse, 1994; Stanley & Brickhouse, 1994; Calabrese Barton, 1998; Calabrese Barton & Yang, 2000; Bang et al., 2012). While this argument has more effectively been made regarding traditional teachings in liberal arts, Stanley and Brickhouse (1994) find that it is not as easily applied to the sciences precisely because science appeals to a "universalist epistemology: that the culture, gender, race, and ethnicity, or sexual orientation of the knower is irrelevant to scientific knowledge" (Stanley & Brickhouse, 1994, p. 388). This division between scientific knowledge and facets of human identity is what Bang et al. (2012) refer to as the "nature-culture divide": a binary that negates the fluid and interdependent relationship between nature and culture. While proponents of science as based on a universalist epistemology have argued that the scientific method and peer-review reduce individual bias, multiculturalists counter that because the scientific community lacks diversity and is rather homogenous, peer-review and the scientific method remain blind to epistemological groundings of scientific knowledge (Stanley & Brickhouse, 1994; Bang et al., 2012). Thus, objectivist framings of science necessarily uphold widely held belief systems, values, and epistemologies of a largely homogenous community. To be sure, Western school science obscures and denies the knowledge systems of marginalized groups by delineating "the scope of what constitutes an acceptable explanation, argument, or analysis; what 'smart' looks and sounds like; whose narratives and experiences are valued and for what purposes" (Bang et al., 2012).

Bang et al. (2012) argue that the set of assumptions that delineate normative and widely held beliefs about a colorblind and objectivist science continue unabated through the denial of the ways in which culture shapes the purposes of science doing, interpretations of scientific discovery, and promotion of science knowledge and theory. In reality, all knowledge systems are socially constructed, thus the concept of a value-free science is a myth. With this in mind, the push for a multicultural framing of science serves as impetus to re-conceptualize notions of whose knowledge counts, what counts as science, and who is able to do it. This critical awareness of the role of the epistemological underpinnings of STEM education, however, remains absent even in reform-oriented educational policy. For example, “science for all,” an educational policy that seeks to cultivate scientific literacy among all students, particularly non-dominant students, makes wide-sweeping assumptions about the “all” and takes for granted notions of science and science-doing (Calabrese Barton, 1998). According to Calabrese Barton (1998), research regarding “science for all” can be grouped into three veins: access to resources; knowledge of rules of participation; and the need for culturally relevant teaching. She argues that while these three affordances are important to pursue in making science accessible to all, particularly marginalized groups like students of color and those living in poverty, little attention has been paid to the assumed meanings of science. And so, while “science for all” initiatives that attempt to engage culturally relevant teaching as a way of challenging ineffective teaching methods and applications of science, scientific concepts and principles premised on a dominant group epistemology remain unaddressed and unchallenged (Calabrese Barton, 1998).

The crux of the critique of STEM education and scientific knowledge as objective centers on the need to problematize assumed meanings of science that, when unquestioned, validate hegemonic values that systematically privilege dominant groups while promoting deficit discourses about non-dominant groups (Stanley & Brickhouse, 1994; Bang et al., 2012; Brickhouse, 1994; Calabrese Barton, 1998; Calabrese Barton & Yang, 2000). In this way, settled views and expectations of science as objective and based on a universalist epistemology function to uphold and privilege what Delpit (1988) termed the "culture of power."

Culture of power. "The culture of power," first coined by Delpit (1988) in addressing the debate between skills-based and process-based instructional methodologies in writing, refers to "a set of values, beliefs, ways of acting and being that for sociopolitical reasons, unfairly and unevenly elevate groups of people—mostly white, upper and middle class, male and heterosexual—to positions where they have more control over money, people, societal values than their non-culture-of-power-peers" (Calabrese Barton & Yang, 2000, p. 873). To paraphrase Calabrese Barton and Yang (2000), these arbitrary delineations contribute to social stratification that upholds systems of power and privilege for those that have access to the culture of power and are positioned atop a tiered society, and necessarily erect obstacles for those that do not and are not. While Delpit (1988) argues that the culture of power pervades society and also reigns in the institution of education, Calabrese Barton and Yang (2000) extend this perspective by arguing that the culture of power is at the core of universalist views of STEM education and scientific knowledge. According to Delpit (1988), the culture of power in American schools consists of five premises that ultimately work to silence the voices of nondominant students; they are:

1. Issues of power are enacted in classrooms
2. There are codes or rules for participating in power; that is, there is a "culture of power"
3. The rules of the culture of power are a reflection of the rules of the culture of those who have power

4. For those who are not already a participant in the culture of power, being told explicitly the rules of that culture makes acquiring power easier
5. Those with power are frequently least aware of—or least willing to acknowledge—its existence, and those with less power are often most aware of its existence. (p. 282)

The first premise holds that issues of power are enacted in the classroom through, for example, the power and authority of the teacher over students; curriculum to dictate what is taught and how; textbooks that impart knowledge through the author's/publisher's lens; and state mandated schooling (Delpit, 1988). The second premise holds that there is a particular language, decorum, and self-presentation that hold clout in society and in educational spaces. The third premise refers to the fact that these clout-laden codes and rules are intimately known by those from dominant groups and thus reflect the culture of those in power. The fourth premise posits that those from non-dominant groups can, to some extent, acquire power and participate in the culture of power more easily if explicitly informed of the codes and rules for participation valued within the culture of power. Lastly, the fifth premise addresses three important facets of access to power. It highlights the automatic privilege of those from dominant groups as participants in the culture of power to the extent that some of those who have access to the culture of power, and thus to power, are not explicitly aware of it. Further, those who have access and are aware of it are less inclined to acknowledge it because of the uncomfortable nature of confronting one's own privilege in light of other's oppression or disadvantage. Delpit (1988) argues that this presents a distinctly uncomfortable feeling, for example, for those "who consider themselves members of liberal or radical camps" (p. 283). The third facet of this premise refers to the fact that "those who are less powerful in any situation are most likely to recognize the power variable most acutely" (p. 284).

The culture of power pervades American society and educational institutions to the extent that even within progressive circles, those that are not members of dominant groups, and thus lie outside of the culture of power or lack access to it, are referred to as "disadvantaged" and "at-risk," and are positioned as members of an exclusionary culture (Calabrese Barton & Yang, 2000). It is, then, no surprise that groups described with such labels tend to be low-income youth of color—that is, members of non-dominant groups. When it comes to schooling experiences of youth of color, their lower levels of academic achievement are often viewed through two perspectives. The first perspective blames students of color for their educational outcomes and takes the deficit position that these students must be "fixed." Proponents of this view propose targeting these students and placing them in remedial classes aimed at "fixing" the problem within the student. The second perspective holds that the issue of low academic achievement and outcomes are a result of the educational institutions, and not deficiencies inherent in the student. This second perspective begins to examine the role schools play in disadvantaging students and acknowledges the relationship between educational institutions and power (Calabrese Barton & Yang, 2000; Delpit, 1988). Further, while the culture of power privileges students from dominant groups, it undermines students' abilities and willingness to engage with science and scientific knowledge in personally meaningful ways particularly because traditional science, as we know it, is not epistemologically oriented to the ways of knowing that are familiar and valuable to students from non-dominant groups. To illustrate the consequential dangers of a culture-of-power-laden science for non-dominant students, Calabrese Barton and Yang (2000) write,

> School-based science practices have led to an overwhelming number of students believing that science is a body of knowledge which consists of events, facts, and theories existing "out there" (Cobern, 1996), that science is static rather than dynamic (Yager,

> 1990), that only the very brightest of people can do science (Lemke, 1990), that science does not connect with their personal lives (Brickhouse, 1994; Barton, 1998), and that once they fulfill their scholastic requirements, they will be "done" with science for the rest of their lives (Kahle & Meece, 1994). These kinds of views of science have contributed to low achievement levels in school science, low attitudes toward science and science careers, and low numbers of women and people of color entering the sciences as career choices in the United States (Kahle & Meece, 1994). (p. 876)

Furthermore, the critical views presented above make clear that STEM reform initiatives have gone too long without attuning themselves to the foundational issue of STEM education as premised on epistemological exclusion and subsequent denigration of non-dominant groups. Any effort to meet the STEM imperative on all levels (i.e. national, local, and personal) must allow the valorization of epistemological contributions of students from non-dominant groups.

Toward science as co-constructed knowledge. Two principal reasons cited for the disproportionate representation of women and people of color in the sciences are 1) deficit thinking about the intellectual abilities of women and people of color, and 2) their unfair treatment in schools (Brickhouse, 1994). In order to rectify the lack of diversity in STEM; challenge deficit notions of non-dominant groups, their intellectual contributions, and meaning-making practices; and expand possibilities for scientific discovery, science educators, researchers, and policymakers must work toward valuing and embracing the ways of knowing of non-dominant groups.

Among scholars critical of the positivist framing of science in schools there is widely-held consensus that the movement to reform STEM education must engage in critical examinations of the framing of science and actively work toward re-framing of science and notions about what constitutes legitimate science doing, who can do it, and how (Brickhouse, 1994; Calabrese Barton, 1998; Calabrese Barton & Yang, 2000; Stanley & Brickhouse, 1994; Bang et al., 2012). One way to do so involves viewing science through the framework of poststructural feminism (Calabrese Barton, 1998), which enables a positioning of science knowledge as co-constructed meaning constituted through "a reflexive relationship between 'science' and 'all'…[where reflexivity and deconstruction of normative subjects usher in a valuation of]… their taken-for-granted historical, social, theoretical, and linguistic structures" (Calabrese Barton, 1998, p. 529).

Calabrese Barton (1998) argues that scientific knowledge is created through a co-constructive process wherein "the knower, the known, and the context in which they interact"—what she terms the teacher-student-science triad—are historically, socially, and politically shaped; thus, none are objective in nature. Through the interaction of these socioculturally constructed elements, science teaching, science learning, and science knowledge become imbued with sociocultural meanings. In other words, all three are socially constructed (Calabrese Barton, 1998; Stanley & Brickhouse, 1994). Truly, science-knowledge, in the school context, is ever-shifting and continually co-created through the reflexive relationship between "science" and the "all." Calabrese Barton (1998) asserts "A science for all can truly be for 'all' only if it is removed from the center and allowed to be a positional and dynamic construction of multiple realities" reconceptualized as socially, politically, and culturally constructed (p. 539).

To illustrate what this can look like in a learning setting, I will discuss two examples of efforts to challenge positivist framings of science and science knowledge. The first draws from poststructural feminism to reframe science knowledge as dynamic, multi-voiced, and co-constituted (Calabrese Barton, 1998); the second pursues a project to "desettle" settled expectations in science and reframe science as built on a fluid and interdependent relationship between nature and culture (Bang et al., 2012).

Science knowledge as dynamic, multi-voiced, and co-constituted. Calabrese Barton (1998) writes of her experience engaging in this reflexive type of science with children in an after-school science program at a homeless shelter. It was at the shelter where she met K'neesha, a seventh grade African American girl who was living in the homeless shelter with her mother and sister. K'neesha and her family had moved several times within that year due to financial instability. As a result, K'neesha was attending her third school that year and had fallen behind academically. Despite the availability of adequate learning materials at the school, its commitment to helping her catch up in science, and her teacher's efforts to make science relevant through hands-on activities and by drawing connections between science and students' daily lives, K'neesha expressed that her science class "did not mean anything" to her (Calabrese Barton, 1998, p. 533). She ultimately failed the science unit on digestion she was working on at the time.

In contrast to the traditional school science unit on digestion, Calabrese Barton worked with K'neesha and other children at the homeless shelter to come up with a science project that would reframe traditional notions of an objective science and transcend typical approaches to "science for all" wherein the "science" is assumed as true and neutral and students are expected to change to fit settled science expectations (Calabrese Barton, 1998; Bang, et al., 2012; Rosebery et al., 2016). Together, they developed a science project on pollution in their community inspired by students' frequent remarks about their community as "dirty, run-down, and polluted" (Calabrese Barton, 1998, p. 534). This is one way that science in the after school program began to take form in a way that reflected the multiple voices in the program, taking shape as a "positional and dynamic construction of multiple realities" (Calabrese Barton, 1998, p. 539). K'neesha was notably more involved and personally invested in this project compared to the unit on digestion, particularly with respect to its development and ideas for data collection. Calabrese Barton felt that this had to do with the reframing of what science is, who can do it, where, how, and why.

The project on pollution began with a chart listing the complaints they had about their communities, how those issues made them feel, and additional problems that arose as a consequence of those issues. With Calabrese Barton's guidance, students developed a research project attuned to their concerns about their community, and advanced ideas for data collection, thus beginning to "do science" and reframing *what* constitutes scientific inquiry and *who* can engage in it. K'neesha suggested interviewing community members about how they felt about their community, pollution in their community, and if and how they contributed to it, redefining *where* science can be done. The children also reframed the *how* of science throughout the project beginning with its conceptual development attuned to their lived experiences, and its execution through data collection via video recorded interviews, which was K'neesha's idea. A reflexive science, in K'neesha's case, was a science where she was validated as a scientist, where her concerns about her community were central to the project, and her methods were engineered and realized. K'neesha's after-school science was not about guiding K'neesha in doing science as prescribed, but was instead concerned with the systematic development of scientific inquiry to address issues that were personally meaningful to K'neesha and her community, and thus redefined the *why* of science.

Science as built on an interdependent relationship between nature and culture. In another STEM reform-based effort out of the Chéche Konnen Center, an attempt to "desettle" normative disciplinary constructs in school science, specifically biology, Bang et al. (2012) document how a multi-subject teacher of a 6^{th}-8^{th} grade classroom drew from Haitian immigrant students' skills in *bay odyans*, "a discourse practice widely held among Haitians," (p. 311) to study the relationship between nature and culture. By treating the school biology curriculum as emergent and inquiry-based, the teacher was able to successfully capitalize on continuities between students' epistemologies and science. The approach taken in reconstructing normative expectations in biology in this classroom resulted in meaningful learning experiences for students. Central to this effort was a commitment to confront "settled expectations" (Harris, 1995) that pervade and persist in school science. Settled expectations refers to "the set of 'assumptions, privileges, and benefits that accompany the status of being white…that whites have come to expect and rely on' across the many contexts of daily life [Harris, 1995, p. 277]" (Bang et al., 2012, p. 303). This reform-based approach to teaching biology centered the critique of science as universalist and objective and sought to reframe science through its adoption of students' discursive practices as legitimate contributions to learning about biology.

While critical scholars in the learning sciences offer their perspectives in the context of traditional school science, their core critique is not exclusive to these disciplines but is instead directed at the longstanding STEM tradition of claims to objectivity as a systemic mechanism for the exclusion of non-dominant groups in service of maintaining the preeminence of dominant groups and their epistemology. The enduring overrepresentation of dominant groups and underrepresentation of non-dominant groups in fields that make up the genealogical lineage of a budding data science (i.e., computer science, applied mathematics) indicate that data science, too, is prime to inherit the dominant trait of epistemological exclusivity in the absence of a critical interrogation of the constitution of knowledge in the field. To better understand how existing critiques of scientific objectivity apply to the field of data science, still in its infancy, I now turn to a discussion of existing scholarship on data science, Big Data, and data-generating learning technologies.

The Myth of [Data] Scientific Objectivity, Revisited

The growing body of scholarship on Big Data and data science concerns itself predominantly with examining the revolutionary nature of digital data generation and its application in organizational science. Thus, much of the attention garnered by Big Data regards understanding new and ever-expanding opportunities for organizations to boost efficiency and productivity by making Big Data actionable through the execution of Big Data analytics (McAbee, Landis, and Burke, 2017; Chen, Chen, Gorkhali, Lu, Ma, and Li, 2016; Chong and Shi, 2015). Big Data analytics refers to "a set of [advanced analytic] techniques that allow researchers and practitioners to identify relations between observed variables and/or cases" in large data sets for the purposes of data-driven decision-making (McAbee et al., 2017, p. 280; IBM, n. d.). In efforts to improve productivity, organizations are investing in new technologies and highly skilled personnel to manage large data streams and leverage insights (McAbee et al., 2017; McAfee & Brynjolfsson, 2012).

While there is no shortage of interest in capitalizing on the insights of Big Data, there is less attention, and much less socio-criticality, where epistemology and politics of Big Data are concerned. Much like traditional science and mathematics fields that position dominant knowledge systems as universal and objective (Bang et al., 2012; Calabrese Barton and Yang, 2000; Calabrese Barton, 1998; Brickhouse, 1994; Stanley and Brickhouse, 1994), data and data-generating technologies, too, are often misrepresented and subsequently misread as apolitical and objective (Selwyn, 2016; Selwyn, 2015; Couldry, 2014; boyd and Crawford, 2012).

Arguing pointedly against the myth of Big Data objectivity and the need for critical sociological perspectives in the era of Big Data, boyd and Crawford (2012) assert that while computational scientists might claim their work "as the business of facts and interpretation," (p. 667) where there is human intervention there is subjectivity (Ebach et al., 2016; Ekbia et al., 2015; Couldry, 2014). In short, claims to objectivity are false in their refusal to acknowledge that all human decision making is epistemologically grounded in the lived experiences, sensibilities, and philosophical perspectives of individuals (Couldry, 2014). Coupled with traditional science's longstanding claim to objectivity, perceived data and tech-neutrality mean the likely reproduction of positivism in the field of data science. As a matter of fact, scholars in the humanities and social sciences have noted that many dilemmas observed in the traditional ("historical") sciences, including epistemological exclusion as a result of positivism, have emerged and to a large extent have come to define Big Data and Big Data analytics within data science-related fields (Ekbia et al., 2015; Ebach et al., 2016).

For this reason, it is important to understand that subjectivity plays a role in all phases of the data life cycle where human interpretations, observations, and choices are necessary for making decisions about the structure and design of data collection methods, management, processing, interpretation, and representation (boyd and Crawford, 2012; Selwyn, 2015; Halford, Pope, and Weal, 2013). To put it another way, "Big Data is not self-explanatory," but is instead dependent on human intervention to yield actionable insights (Bollier, 2010, p. 13; boyd and Crawford, 2012).

Critical digital sociological perspectives of Big Data, technology, and education. In response to the increasing momentum of positivist discourse regarding Big Data and data-generating technologies in education, the burgeoning field of digital sociology offers critical perspectives concerning the role Big Data play in reproducing disciplinary and social inequalities, and intensifying managerialism in schools; (Selwyn, 2015; boyd and Crawford, 2012). Thus, in the following sub-sections, I expound on these areas in light of perspectives offered by critical scholars in the learning sciences discussed earlier.

Reproduction of social inequalities. The hype that surrounds the Big Data phenomenon rests on the promise of never-before-seen opportunities to quickly access massive data stores of diverse types of information. This has led to heightened demands for Big Data analysts with advanced statistical analysis and programming skills applicable in a multitude of industries including, but not limited to, government, business, healthcare, education, and social media (McAbee et al., 2017; Chen et al., 2016; McCartney, 2015). Therefore, participation in the field of data science is contingent on matters of *access* and *skills*. Critiques offered by boyd and Crawford (2012) help explain why this is problematic:

> Wrangling APIs, scraping, and analyzing big swatches of data is a skill set generally restricted to those with a computational background. When computational skills are positioned as the most valuable, questions emerge over who is advantaged and who is disadvantaged in such a context. This, in its own way, sets up new hierarchies around ‘who can read the numbers’, rather than recognizing that computer scientists and social scientists both have valuable perspectives to offer. (p. 674)

Grounded in the enduring myth of scientific objectivity, the privileging of ‘who can read the numbers’ as a more valuable skill in data science than qualitative analyses reproduces existing tensions in academia regarding quantitative research as a matter of facts production and qualitative research as a matter of storytelling (Couldry, 2014; boyd and Crawford, 2012). Among Big Data enthusiasts lies the belief that purely quantitative analyses of correlation suffice for analyzing data, making claims about causality, and, subsequently, about what is true and matters in the social world (Couldry, 2014; boyd and Crawford, 2012). These beliefs reflect yet another phenomenon that has arisen in the era of Big Data: Big Data hubris, defined as "the notion that big data replaces, rather than supplements traditional data acquisition and analysis, namely 'small data'" (Ebach et al., 2016, p. 2). This point of view poses a direct threat to the analytical and interpretive offerings of the humanities and social sciences by suggesting the impending obsolescence of qualitative methods for understanding the social world (Ebach et al., 2016). Further, Couldry (2014) argues that by appealing to objectivity

> [Big Data's] effect is to reinforce our belief that such data offer a new route to social knowledge…Each such myth, by rationalizing a certain perspective on how we can come to know the social, obscures our possibilities for imagining, describing and enacting the social *otherwise*…the power of the myth of big data emerges [in challenging] the very idea that the social is something we can *interpret* at all. (p. 882, emphasis in original)

Although Big Data is a socio-technical phenomenon profoundly informed by and informing the social, Big Data hubris inherent in the myth of Big Data objectivity creates the illusion that studies of correlation, user profiling, and predictability supersede social-scientific and humanistic methods for understanding the social (boyd and Crawford, 2012; Couldry, 2014). Like traditional STEM fields, the constitution of scientific legitimacy in data science reflects settled and narrow expectations regarding how to do science (data analytics), who can do science (data analysts), and ultimately who is able to contribute to the creation of social knowledge (top-level stakeholders and data analysts) (Bang et al., 2012; Calabrese Barton and Yang, 2000; Calabrese Barton, 1998; Brickhouse, 1994; Stanley and Brickhouse, 1994). Therefore, by challenging the social (i.e. Big [social] Data) as the work of interpretation and instead positioning it as the work of analytics, the myth of Big Data objectivity exists in service of the enduring culture of power and settled expectations in STEM (Ebach et al., 2016; Ekbia, et al., 2015; Couldry, 2014; Bang et al., 2012; boyd and Crawford, 2012; Calabrese Barton and Yang, 2000; Delpit, 1988; Harris, 1995).

In addition to reproducing *disciplinary* inequalities, Big Data has already begun its role in the reproduction of *social* inequalities due to issues of differential access to new technologies, internet use, Big Data for analysis, and skills necessary to actuate analysis (boyd and Crawford, 2012). Through the creation of a social hierarchy of ‘data classes’ “ordered along lines of technical and statistical expertise,” divisions between the Big Data rich, those who have the skills necessary to do Big Data analytics, and the Big Data poor, those who do not, widen (Selwyn, 2015, p. 71; Ekbia et al., 2015; boyd and Crawford, 2012). Ekbia et al. (2015) provide that the popularity and profitability of Big Data is founded on socioeconomic, cultural, and political shifts that Big Data, itself, enables through its power “as a polarizing force not only in the market, but also in arenas such as science” (p. 1537). Yielding massive business profits, Big Data has been referred to by some as a new “asset class” and “new oil”—a fact that helps contextualize Facebook’s increase in ad revenue informed by Big Data analytics from $300 million in 2008 to $4.27 billion four years later (Ekbia et al., p. 1537). Of course, we all contribute to the generation of data on a grand scale in one way or another, but proximal benefits are limited to an exclusive group of individuals who possess the digital capital—“the reach, scale, and sophistication of his or her online behavior”—necessary to capitalize on Big Data (Ignatow & Robinson, 2017, p. 3; Couldry, 2014). As a matter of course, emerging inequalities in who has access to information technologies and highly valued skills of data analytics mimic and amplify existing socioeconomic inequalities (Ignatow and Robinson, 2017).

What is more, while Big Data enthusiasts revel in the purported accessibility of limitless data, access is not equally granted nor widely distributed (boyd and Crawford, 2012). Powerful companies like Google and Facebook possess the means to finance the collection of Big Data, what Van Dijk (2005 cited in Ignatow & Robinson, 2017) refers to as information capital. However, their willingness to finance Big Data analytics does not mean that the public will be granted open access to the data that is collected, particularly when it comes to proprietary data (boyd and Crawford, 2012). Writing in the context of social media, boyd and Crawford (2012) posit that "[s]ome companies restrict access to their data entirely; others sell the privilege of access for a fee; and others offer small sets to university-based researchers" (p. 673).

Moreover, social divisions between the Big Data rich and the Big Data poor are reinforced through the university system where access to Big Data is more readily acquired by well-resourced top universities that can afford to pay for access (boyd & Crawford, 2012). Citing Capek, Frank, Gerdt, and Shields (2005), Ekbia et al. (2015) observe that "[t]he complex ecosystem in which Big Data technologies are developed is characterized by a symbiotic relationship between technology companies, the open source community, and universities" (p. 1527). Adding to this point, in 2013, information technologies firms supported Big Data science-related course offerings at nearly 26 institutions of higher learning—figures that have presumably increased since (Cain-Miller, 2013). These relationships are further complicated by the fact that students from top universities have a greater likelihood of being invited to develop their skills at large companies that head data-collection ventures, creating a direct pipeline for students majoring in data science-related disciplines to enter the field of Big Data analytics (boyd and Crawford, 2012).

Unfortunately, research shows that the educational pipeline available to non-dominant students, as yet, is more bleak than promising. When we consider the well-documented lack of access to quality educational experiences and opportunities for students from non-dominant groups, and the fact that they do not enjoy equitable representation in data science-related fields and STEM in general, it is not difficult to discern how emergent disciplinary and social divisions perpetuated via the myth of Big Data objectivity are poised to perpetuate deep-seated educational inequity for non-dominant students in the new field of data science (Martin et al., 2015; Pérez Huber et al., 2014; Covarrubias, 2011; Barr et al., 2008).

Toward data science as co-constructed knowledge. Unless there is a critical sociological understanding of the ways in which certain types of knowledge are legitimized as true and universal in the fields that constitute the foundation of data science, including mathematics, statistics, and computer science, then data science too runs the risk of reproducing and intensifying the power structures that create inequity in STEM through the simultaneous valuation of dominant epistemologies and negation of non-dominant ones (Selwyn, 2015). In other words, it does not suffice to define equity in data science as exposure to processes involved in data analytics for students who have not traditionally benefitted from access to STEM education and careers if they are learning to work with data in ways that reinscribe their epistemological exclusion and leave unquestioned the new manifestations of positivism in data science and treatment of data-generating technologies (Selwyn, 2015).

Assessing Learning Within STEM Reform

Reframing "success" in data science learning outcomes. While critical and mindful design of equity-oriented data science initiatives represents a difficult yet necessary effort, assessing learning within programs piloting new data science curricula represents another that is equally difficult and necessary. Achieving equity in STEM education not only requires challenging the myth of scientific objectivity, but also a reframing of assessment that moves beyond narrow definitions of success that focus on competency measures and the development of technical skills (Carlone et al., 2011; Kearns, 2011; Au, 2011; Jacob, 2005; Hursch, 2005; Kohn, 2000). The technical rationality inherent in traditional learning assessments works to the advantage of students from dominant groups and to the disadvantage of students from non-dominant groups by shifting the focus of education from fostering students' personal and intellectual development to developing the technical skills and information necessary to perform well on standardized measures. This, invariably, means the perpetuation of educational inequity for those students who have been traditionally underrepresented in higher education, particularly and more so in STEM.

Accordingly, Carlone et al. (2011) assert that science educators and researchers must think critically about what counts as equity and what constitutes equitable outcomes in reform-based science. They caution that "Doing well on achievement measures does not necessarily, by itself, imply a successful outcome" because this does not indicate that a student "affiliates" with science—that is, sees themselves as a "smart science person" (Carlone et al., 2011, p. 462) or perceives science as personally meaningful. Indeed, these measures convey very little about student learning (Khalifa, Jennings, Briscoe, Oleszweski, and Abdi, 2014; Au, 2011; Hursch, 2005; Kohn, 2000). In their efforts to understand how students in two fourth grade classes came to think of and affiliate with being a "smart science student," Carlone et al. (2011) found that "those who expressed outright disaffiliation [with being a smart science student] were not necessarily those who did not perform well on assessments. In fact, N'Lisha, an African American student, was one of the top performers on both written and performance assessments" but she identified three White students as “science” people,” adding “We aren’t like them” (p. 461-462).

There is a significant distinction to be made between satisfactory performance on standardized measures aligned with hegemonic science standards and the execution of learning processes that indicate meaningful engagement in scientific learning. If the latter is consistent with the values and ways of knowing of those that lie outside of the culture of power, then standardized measures will not account for nor value the unique learning processes because, as projects of positivism, they are not designed to do so (Selwyn, Henderson, and Chao, 2015; Kaufman, Graham, Picciano, Popham, and Wiley, 2014; Khalifa et al., 2014; Hursch, 2005). Furthermore, meaningful engagement in scientific learning and self-identification or perception of oneself as “a science person” holds more promise for encouraging a student to pursue an education and career in STEM than does satisfactory science task completion with little personal investment. Hence, assessment of STEM reform initiatives must also challenge the notion of assessment itself by taking the “affective dimensions (and dispositional outcomes) of learning” into consideration and “examin[ing] the ways the promoted ways of ‘being scientific’ in a classroom are meaningful, believable, and achievable for a diverse groups of students” (Carlone et al., 2011, p. 463). This means that a reframing of “success” in STEM equity outcomes requires analysis of how meanings of science, legitimate science knowledge, and legitimate science-doing are culturally shaped, and thus co-constituted, in the classroom (Carlone et al., 2011; Cobb & Hodge, 2002; Cobb et al., 2009; Cobb & Hodge 2010). To do so is to challenge positivism and the settled expectations of technical rationality that uphold and protect the Culture of Power in education through appeals to the myth of scientific objectivity and to the disadvantage of non-dominant groups through epistemological exclusion (Ebach et al., 2016; Ekbia et al., 2015; Bang et al., 2012; Calabrese Barton and Yang, 2000). Aware of the problematic implications and ineffectiveness of standardized measures for understanding student identification with STEM, namely mathematics, the work of Cobb and colleagues purposefully moves toward alternative measures for gaining a better and more nuanced understanding of student performance in and identification with math-doing. Therefore, in the following section, I will discuss this scholarship and its usefulness and appropriateness for understanding student performance in and identification with data science-doing.

Supporting the development of strong student STEM identities. In this section, I will draw from the work of Cobb and his colleagues to make sense of data science education principally because it draws heavily from its parent field of mathematics. While I recognize that data science is different than mathematics due to its focus on data collection, analysis, and management; as well as its use of coding and computer programming, the longer history of scholarship on mathematics education will be immensely useful for providing a dynamic portrait of data science education.

Echoing the critical views of scholars presented earlier, Cobb and colleagues argue for the restructuring of mathematics education for equity via the cultivation of strong student STEM identities. Their work stems from a design-based research tradition that seeks to support rich mathematical student learning and bring about strong student identification with mathematics as math-doers via a process of iterative instructional design (Cobb & Hodge, 2002; Cobb et al., 2009; Cobb & Hodge 2010). Cobb et al.'s proposal to restructure mathematics specifically through the personal development and enrichment of students' identities acknowledges the objectifying and silencing nature of top-down efforts to improve educational outcomes. In essence, supporting the development of strong student STEM identities is about epistemological inclusion, where students are positioned as co-creators of knowledge.

What it means to be a successful student in a mathematics class is a ‘normative’ construct co-constituted by and through the social interactions that take place in the classroom between the teacher and the students (Cobb, Wood, Yackel, & McNeal, 1992). Normative notions of what it means to be a successful math student as developed in a classroom pose implications for the quality of learning that can take place, and hence for opportunities afforded to students to develop strong affiliations with math-doing (Cobb et al., 1992). Like scientific knowledge, the culture, values, and imperatives that emerge in a given mathematics classroom are socially constructed, with normative expectations, behaviors, and practices established by a community of individuals. Furthermore, traditions established by a community of individuals in turn “influence individuals’ construction of scientific or mathematical knowledge” by discerning what constitutes acceptable problems, solutions, explanations, and justifications within that particular tradition (Cobb et al., 1992, p. 575).

In their interactional analysis, Cobb et al. (1992) observed two elementary school mathematics classrooms, one at the second-grade level and another at the third-grade level. Illustrating the drawbacks of engaging with STEM disciplines as objective and fixed, Cobb et al. (1992) found that in the third-grade mathematics classroom—where mathematical knowledge was treated as pre-existing, pre-determined, and external to the students in the class—the mathematical learning that occurred could be characterized as procedural mathematical learning without conceptual understanding. In this particular classroom students learned that they were expected to provide the answer anticipated by the teacher and not necessarily to learn the conceptual underpinnings of their mathematical activity on place value numeration. This example illustrates the argument conveyed earlier—that competency in science knowledge and skills do not account for the conceptual ways in which students are able to interact with science knowledge in the classroom (Carlone et al., 2011). Additionally, "[t]he manner in which the children routinely cooperated even when the mathematical rule was not immediately evident to them indicated that, at a minimum, they had learned to act as though they believed that mathematics consists of fixed, objective rules" (Cobb et al., 1992, p. 589). This ultimately contributed to students' enculturation into positivist beliefs about math (Lave, 1988 cited in Cobb et al., 1992) as universal, objective, factual, and fixed (Brickhouse, 1994; Calabrese Barton, 1998; Calabrese Barton & Yang, 2000; Stanley & Brickhouse, 1994; Bang et al., 2012).

Unlike the procedural instructions that the teacher conveyed and students took up in the third-grade classroom, the second-grade classroom consisted of challenges and back and forth exchanges among students and the teacher, "There was a taken-as-shared [normative] understanding of the task that reflects a prior history of negotiations of meanings and interpretations in the classroom" (Cobb et al., 1992, p. 590). This means that the activity of negotiation and meaning-making was central to math learning in this classroom, which ultimately lead to students' ability to learn mathematics with understanding, as opposed to without, and in ways that were personally meaningful (Cobb et al., 2009; Cobb & Hodge, 2002; Yackel & Cobb, 1996).

Moreover, Cobb et al., (2009) propose an interpretive scheme designed with the explicit purpose of analyzing the types of identities that students develop in the mathematics classroom to inform instructional design and teaching. Cobb et al. (2009) argue that the personal identities that students develop as math-doers are important to consider in instructional design because they indicate the ways in which students come to identify with the normative identity of a successful mathematics students as co-constructed in the classroom. Pulling from Cobb & Hodge's (2002) earlier work on classroom social structure, they posit that the normative identity of a successful mathematics student is a social construction shaped by the classroom microculture, which, when treated as a community of practice (Wenger, 1998), consists of three aspects:

1. General classroom norms,
2. Specifically mathematical norms, and
3. Classroom mathematical practices (Cobb et al., 2009; Cobb & Hodge, 2000; Cobb & Yackel 1996).

Existing literature on "how students come to understand what it means to do mathematics as it is realized in their classroom and with whether and to what extent they come to identify with that activity" (Cobb et al., 2009, p. 41) indicates three cases of student math affiliation: students come to identify with, merely cooperate, or resist engagement in and affiliation with mathematical activity as it plays out in the classroom (Boaler & Greeno, 2000; Martin, 2000). As such, a strong affiliation or student identification with mathematical classroom activity, and "taken-as-shared" mathematical understandings as co-constituted in the classroom have the potential to inspire in students personally meaningful rationales[5] for math learning.

Furthermore, student development of personally meaningful rationales for identifying with mathematical activity in the classroom is dependent on the ways in which students are able to exercise legitimate forms of agency within the classroom. Opportunities for students to

[5] As cited in Cobb and Hodge (2010), D'Amato (1992) "distinguishes between two ways in which learning in school can have value for students" (p. 185). The first way involves students seeing extrinsic value, or structural significance, in learning. Students that see the structural significance in learning in a mathematics classroom will identify with mathematical activity in efforts to perform well and be successful as a means to an end such as gaining admission to college. The second way involves students seeing intrinsic, or situational significance, in learning "in which students view their engagement in classroom activities as a means of maintaining valued relationships with peers and of gaining access to experiences of mastery and accomplishment" (p. 185).

exercise legitimate forms of agency are determined by how authority is distributed in the classroom. The distribution of authority in the classroom refers to “the degree to which students are given opportunities to be involved in decision making about the interpretation of tasks, the reasonableness of solution methods, and the legitimacy of solutions” in the classroom (Cobb et al., 2009, p. 44). Consequently, the extent to which students are able to make legitimate mathematical contributions in the classroom has implications for the types of agency they are able to exercise in the classroom. A significant point to consider is that while agency can be described as fluid, exercised on a continuum, and the extent to which one is able to engage in agency, Cobb et al. (2009) “use the term *agency* in a more technical sense that moves beyond the view of agency as an amount and focuses on the ways in which students can legitimately exercise agency in particular classrooms” (p. 44-45). As such, a classroom wherein authority is distributed to the teacher and students, as would be the case in classrooms that incorporate discursive practices of students from non-dominant groups and/or follow an inquiry-based instructional design (see Bang et al., 2012; Carlone, 2004), for example, would provide opportunities for students to exercise conceptual agency. A classroom wherein authority is distributed solely to the teacher, such as classrooms that follow the traditional elicitation-response-evaluation pattern (see Cobb et al., 2002), would provide opportunities for students to exercise disciplinary agency.

By and large, the normative identity of what constitutes an effective and successful math-doer in a classroom is socially constructed through the norms and practices of the classroom social structure. Opportunities for students to develop strong affiliations with mathematical activity by viewing it as structurally or situationally significant to them as individuals can be supported or obstructed by the nature of authority distribution in the classroom and, subsequently, the types of agency students are able to exercise. Thus, a systematic analysis of the classroom social structure can prove of immense use, unlike high-stakes assessments, in improving students' educational outcomes through efforts to understand the nature and quality of student identification with mathematical activity and identifying ways to support student views of math, and STEM in general, as personally meaningful and enriching.

Emergence of New Tools and Multiliteracies for Equity-Oriented STEM

As a result of the strong push to increased access to STEM for historically non-dominant groups, a number of scholars have noted relatively new approaches to engaging youth in STEM education, particularly in data science. Below, I discuss data science (Philip et al., 2013; Hogenboom, Holler Phillips, and Hensley, 2011; Calzado Prado & Marzal 2013; Stephenson & Schifter Caravello, 2007) for the development of critical literacies for the 21st century.

Data literacy in STEM reform. Within the last decade, data literacy has emerged as a highly relevant and necessary field for life in the 21st century where large data sets are made available to the public and data is incessantly collected by ubiquitous mobile technology. While the term "data literacy" has increasingly come into use to loosely refer to "the ability to read and interpret data, to think critically about statistics, and to use statistics as evidence" (Hogenboom et al., 2011, p. 410; Calzado Prado & Marzal, 2013; Carlson, Fosmire, Miller, & Sapp Nelson, 2011), there is no definite consensus about the skills that constitute data literacy, the purpose for cultivating data literacy, and approaches to teaching it. Although there is no formal definition for data literacy, it is understood as consisting of a number of abilities and practices. Articulations of the importance of data literacy range from describing it in terms of Big Data analytics (McAbee et al., 2017; Carlson et al., 2011; Gunter, 2007) to emphasizing the need for data literacy as a critical reading of the information-driven world, technology, and issues of privacy and surveillance (Philip et al., 2013; van Dijck, 2014; Tygel & Kirsch, 2015; Bhargava, Kadouaki, Bhargava, Castro, & D'Ignazio, 2016; D'Ignazio & Bhargava, 2016). A number of scholars have also called for the need to cultivate other data-related literacies for the information age. These include information literacy (Elmborg, 2006; Zurkowski, 2013), quantitative literacy (Steen, 1999), statistical literacy (Rumsey, 2002), and numeracy (Stephenson & Caravello, 2007) to name a few. Still others describe the need to think critically about data and the myth of objectivity in the era of Big Data, but do not refer to this as literacy (Selwyn, 2015; Couldry, 2013; boyd and Crawford, 2012). I believe that data literacy encompasses a number of skills that are in ways extensions of other articulations of data-related literacies. Accordingly, not all the literature that I discuss in this section pertains to "data literacy" per se, but instead pertains to data-related literacies articulated by scholars as essential for life in the information age.

Shapiro and Hughes (1996) argue for the need to develop an information literacy curriculum that positions information as a liberal and technical art. This process, they write, entails conceptualizing information literacy as encompassing "the old concept of 'computer literacy'…the librarians' notion of information literacy and a broader, critical conception of a more humanistic sort [of literacy]" (Shapiro & Hughes, 1996, p. 2) consisting of seven distinct yet intersecting competencies.

1. *Tool literacy* refers to understanding and knowing how to use existing information technology tools (i.e. software, hardware) relevant to the educational and occupational spaces inhabited by an individual.
2. *Resource literacy* involves the ability to effectively identify, locate, and access information resources.

3. *Social-structural literacy* involves an awareness of the socially constructed and value-laden nature of information and the ways information becomes imbued with meaning.
4. *Research literacy* refers to the ability to be able to use current research technology tools to conduct research.
5. *Publishing literacy* involves knowing how to compose, prepare, and share research online with the research community.
6. *Emerging technology literacy* refers to the ability to adapt to and use affordances of new technologies in a time when technology is developing quite rapidly so as to not be constricted by the limitations of older technologies.
7. Lastly, *critical literacy* refers to the ability to be critical of information technologies. This means having the ability to evaluate the intellectual, personal, and social affordances and limitations of technology (Shapiro & Hughes, 1996).

These dimensions of data-related literacy are useful for understanding three salient themes in the literature. These themes orient data-related literacy as 1) a functional [technical] tool; 2) an interdisciplinary resource; and 3) a resource possessing humanistic potential. Further, I have grouped Shapiro and Hughes' (1996) seven dimensions into the theme that best describes their relative function (Figure 2.1). Each theme can be understood as characterized by one or more of Shapiro and Hughes' (1996) dimensions of literacy. The dimensions of literacy should not be interpreted as restrictive measures for thinking about data-related literacy within a given theme. Instead, I borrow from Shapiro and Hughes (1996) because after surveying literature on data-oriented literacy, their dimensions represent appropriate guiding competencies that can help inform a mapping and developing understanding of data literacy and related literacies.

Themes in Data-Related Literacy Literature
Data-related literacy as a skill set
Tool literacy
Data-related literacy as an interdisciplinary resource (builds on previous literacy)
Resource literacy
Research literacy
Publishing literacy
Data-related literacy as a resource possessing humanistic potential (builds on previous literacies)
Social-structural literacy
Emerging technology literacy
Critical literacy

Figure 2.1

Additionally, the themes should not be understood as mutually exclusive but instead as an organizational tool for examining my survey of scholarship on data-related literacies. Below I expound on each of the three sections and layout the diverse meanings of data-related literacy. Gaining an explicit understanding what constitutes literacy within equity-oriented STEM reform efforts is necessary for designing and implementing critical and reform-oriented data science initiatives as it necessitates the clear articulation of intended learning goals in the classroom. Additionally, it allows a mapping and reconceptualization of what it means to be critically data science literate.

Data literacy as a skill set. Literature in this vein views data-related literacy as a practical and functional scientific-mathematical skillset. Data literacy can be loosely understood as "the ability to understand, use, and manage science data" (Carlson et al., 2011, p. 633); and the ability to "access, assess, manipulate, summarize, and present data" (Gunter, 2007, p. 2). Data literacy here refers to one's ability to master use of a functional tool, "a suite of data acquisition-, evaluation-, handling-, analysis- and interpretation-related competencies," that enables one to carry out a function and complete tasks related to statistical data (Calzado Prado & Marzal, 2013, p. 2; Carlson et al., 2011; Gunter 2007). Definitions seem to equate data literacy with Big Data analytics, defined as advanced skills of statistical analysis and computing (McAbee et al., 2017; Chen et al., 2016; Chong and Shi, 2015). This definition is consistent with how companies like IBM and those within data science-related fields view data literacy. Further, the myth of Big Data objectivity informs this conceptualization of data literacy as it places data literacy squarely in the realm of the math-sciences and does not acknowledge the need for critical qualitative analysis (Couldry, 2014; boyd and Crawford, 2012; Hogenboom et al., 2011; Qin & D'Ignazio, 2010; Gunter, 2007; Love, 2004). Additionally, some scholars espouse data literacy here with an eye toward increasing one's viability within "an innovative knowledge economy and increasingly data-driven society" (McAuley, Rahemtulla, Goulding & Souch, 2014, p. 53; Carlson et al., 2011; Steen, 1999). Articulations of data-related literacy in this vein are both broad and exclusive. They are broad in that some scholars equate data literacy with other math-science literacies like statistical literacy, numeracy, and quantitative literacy. They are exclusive in alluding to the use and development of data literacy in fields like statistics, mathematics, and information technology, without lending attention to the use and development of data literacy in the social sciences and for citizenship and democracy.

Data literacy as an interdisciplinary resource. Literature in this vein expands the purpose and function of data literacy beyond acquiring a skill set highly relevant for the information age. Here, data literacy is described as involving the development of an interdisciplinary and comprehensive science knowledge that necessitates thinking and synthesizing across disciplines to "access, interpret, critically assess, manage, handle and ethically use data. From that perspective, information literacy and data literacy form part of a continuum, a gradual process of scientific-investigative education" (Calzado Prado & Marzal, 2013, p. 126). As such, Calzado Prado and Marzal (2013) argue, "data literacy can be viewed both as a whole and as an integrated assemblage of other competencies" (p. 126). Relatedly, Steen (1999) argues that numeracy is "both more than and different from [traditional school] mathematics" in that mathematical skills are essential, but unlike traditional school mathematics, numeracy is concerned with learning to use and read data for daily life, to inform decision-making, and enable informed social participation in a democracy (para. 15). Informal inferential reasoning (IIR) (Rubin & Hammerman, 2006; Makar & Rubin, 2009) follows a similar logic and rationale for working with data and challenging traditional and formal approaches to teaching statistical literacy (National Governors Association Center for Best Practices, Council of Chief State School Officers, 2010) that deemphasize formal statistical inference as core to statistical reasoning. Existing scholarship notes that students have found it challenging to generalize and infer from data for everyday life and decision-making when engaging in formal statistical inference due to the decontextualization of abstracted data. As such, IIR calls for "more holistic, process-oriented approaches to learning statistics" that emphasize statistical inquiry as a process to support meaning-making with data (Makar & Rubin, 2009, p. 83; Konold & Pollatsek, 2002). Literature in this vein emphasizes the cross-disciplinary nature of data-related literacies and the importance of utilizing data literacy skills in different contexts toward personal and democratic means (Steen, 1999; Gunter, 2007; Miller, 2010).

Data literacy as a resource possessing humanistic potential. The third category in the review includes scholarship that builds on the previous two themes and argues for critical data-related literacies. Data-related literacy here refers to learning to use/and question mathematical tools and concepts to support personal development, informed citizenship, and democratic participation. These articulations also explore the implications of a data and technology-saturated society for non-dominant groups. Literature in this section follows Elmborg's (2006) articulation of critical information literacy as involving the development of "a critical consciousness about information, learning to ask questions about the library's (and the academy's) role in structuring and presenting a single knowable reality" (p. 198). Elmborg's (2006) critical information literacy emphasizes, like critical scholars in the learning sciences (Calabrese Barton, 1998; Brickhouse, 1994; Stanley & Brickhouse, 1994; Calabrese Barton & Yang, 2000, Bang et al., 2012) and digital sociology (Selwyn, 2015; boyd and Crawford, 2012), the need to examine the problematic nature of knowledge based on hegemonic constructs, and the role that legitimating institutions (in his case the library and academy) play in upholding varied manifestations of positivism.

Scholars have noted parallels between Freire's (1968) critical pedagogy for Popular Education and a critical data-related literacy. Examples of critically-oriented data literacies include Popular Data framework (Bhargava et al., 2016), critical data literacy (Tygel & Kirsch, 2015), and Big Data literacy (Philip et al., 2013). These types of literacies are all similar in that they all employ the collection, use, management, and critical understandings of data; however, their approaches, purposes, contexts, and tools differ.

Popular Data framework for developing data literacy draws inspiration from Freire's

Popular Education (hence, Popular Data) to explore ways in which quantitative and qualitative data can be used to build "participatory and relevant invitations for learners to build a stronger and impactful data literacy" (Bhargava et al., 2016, p. 200). The Popular Data framework for cultivating data literacy engaged youth in a generative process of analyzing data to turn it into a story relevant to the lives of youth and a subsequent telling of that story through the painting of a mural (Bhargava et al., 2016). This interpretation of data literacy expands how we learn about data, how data can be used outside of STEM, and the different ways data can prove to be empowering and lead to praxis.

Critical data literacy, as defined by Tygel and Kirsch (2015), applies Freire's (1968), Popular Education Literacy Method to learning about and with data. Critical data literacy is explicitly concerned with the development of statistical-mathematical methods to "critically analyze data, understand the context where they are generated, and the reality pictured with them" (Tygel & Kirsch, 2015, para. 17; D'Ignazio & Bhargava, 2016; Bhargava et al., 2016). Tygel and Kirsch's (2015) Freirean inspired critical data literacy seeks to contribute to the democratization of data as a means of combating widespread data illiteracy that will only grow given rapid and new developments in technology.

Moreover, Big Data literacy (Philip et al., 2013) emphasizes the need to learn about Big Data, data collection, analysis, interpretation, and visualization in data science. Big Data literacy also calls into question value-free treatments of technology, also known as 'rational technology', and seeks to examine the role ubiquitous technology plays in a number of issues including privacy, corporatization of personal data, and instructional technology in the classroom (Philip et al., 2013).

The last two approaches to data-related literacy necessitate the use and problematization of new technologies as these have growing implications for issues of digital illiteracy, critical social awareness, and democratic participation (Tygel & Kirsch, 2015; Philip et al., 2013). While Tygel and Kirsch (2015) call for future work to derive tangible ways of applying the theoretical constructs presented in their conceptualization of critical data literacy, Philip et al. (2013) pose that some existing approaches to teaching data literacy through data science are driven by uncritical views of new technologies as apolitical and objective, placing too much faith in the transformative abilities of technology and cutting edge content without addressing issues critical to democratic participation and issues of power. Additionally, this treatment of new technologies resonates with the second pillar of technical rationality—a concept which holds that rational technologies are rational in their ability to strip information of subjective meaning through abstraction into quantifiable and objective data (Standaert, 1993). This assertion also follows critiques presented earlier about traditional approaches to teaching statistical literacy that center the need to use statistical tools while deemphasizing statistical inquiry processes that lead to powerful meaning-making with data (Makar & Rubin, 2009; Rubin & Hammerman, 2006). In the following section, I will discuss language as a contributing factor to the uncritical use of new technologies in schools; technology and cutting-edge content; and discuss Philip et al.'s (2013) proposed framework for using new technologies as instructional tools toward powerful meaning-making with data in data science.

The role of language in perceptions of tech-neutrality. The myth of data-scientific objectivity is perpetuated through hyperbolic rhetoric used to describe new data-generating learning technologies and their uncritical, albeit well-intentioned, adoption into the classroom devoid of careful consideration regarding pedagogical affordances; ethics relating to their collection of students' personal data; and automated digital surveillance (Philip, 2017; Selwyn, 2016; Chong and Shi, 2015; Philip and Garcia, 2015; Philip et al., 2013). Reflecting on dominant and enduring views regarding new and mobile learning technologies expressed during a recent international forum Philip (2017) recounts, “the message was quite clear: technology in schools equals innovation; let’s not waste time being negative about technology; let’s just get on with it” (p. 34). “Such a cavalier approach to learning technologies in schools and the flippant reaction to any cautions and critiques,” he contends, “only serve to further jeopardize the learning opportunities of students who have been historically marginalized in schools” (Philip, 2017, p. 34). Indeed, how we talk about new technologies, what we deem worthy of thoughtful discussion and what we dismiss as insignificant within those discussions matters in leveraging the expectations and purposes of using new technology in the classroom (Selwyn, 2016). Selwyn (2016) argues that hyperbolic language used to describe technology in educational settings lends credence to idealized potentials of technology while ignoring realities of use and ulterior, often business related, motives for the introduction of the educational technology industry—valued at over $5 trillion—into the classroom. Uncritical adoptions of new technologies in schools implicitly denies the possibility that some aspects of ‘learning-technologies’ might actually hinder student learning and pedagogical practices meant to foster powerful meaning-making and conceptual reasoning among students (Philip, 2017; Selwyn, 2016: Philip et al., 2013).

Unbridled use of hyperbolic language to describe ‘technology and education’ in a time of heightened interest in the nascent field of data science and other data-related fields contribute to fostering public perception of tech-neutrality, and by extension data-neutrality generated via the use of technology. Undeniably, new [data-generating] technologies provide opportunities for new ways of learning, processing information, and collaborating with peers—these are things we know, but efforts to introduce new technologies into the classroom must be preceded by understandings of the nuances and limitations inherent in different forms of technology and data—as the two are now inextricably linked. Furthermore, STEM reform efforts that incorporate use of new technologies in classrooms and other learning spaces must actively work toward demystifying the notion of value-free technology and data to avoid the shortsightedness of previous efforts that have failed to do so. This is doubly the case for equity-oriented initiatives that incorporate the use of new technologies to learn specifically with and about data.

Technology and cutting-edge content. Philip et al. (2013) caution against the adoption of new technologies and cutting-edge content for data analysis in the classroom without thoroughly thought-out rationales behind their actual purpose, role, affordances, and implications for learning. If the purpose of incorporating new technologies and cutting-edge content in the classroom is to increase democratic participation in ways that confront educational inequities of non-dominant groups, then curricular design of data science programs and technology use must avoid falling into the paradigm of what Philip et al. (2013) term "ideological paradigms of technology and cutting-edge content as an end, means, and equalizer" (p. 104).

STEM reform initiatives that treat technology and cutting-edge content as *the educational end* are reductive in their view of the goals of education and fail to attend to the critical need to improve education for the sake of the people (Philip et al., 2013). The use of technology and cutting-edge content *as a means to educational success* refers to the treatment of new and mobile technologies as inherently revolutionary, a view that overemphasizes, and perhaps misunderstands, student interest in mobile devices without considering how the meaning of technology changes in different contexts (Philip et al., 2013). Moreover, treating technology and cutting-edge content as *the equalizer* assumes that the mere presence of new technologies in the classroom will suffice in ameliorating widespread and deep-seated educational inequality, operating on a highly misinformed understanding of the nature of inequity in schools. Thus, any reform-oriented effort that promotes the use of technology and cutting-edge content as the solution to educational inequity will not only fail to address the true causes of educational inequity, but will also expend precious funding that could go toward more effective approaches to education reform for students of color, all the while intensifying the issues that prompted the very need for reform (Philip & Olivares-Pasillas, 2016). Thus, decisions regarding the use of mobile technologies in the classroom should focus on understanding the affordances that new technology and cutting-edge content provide for learning in the classroom.

Affordances of mobile technology in the classroom: the 3Ts[6]. Philip and Garcia (2013) argue that a useful way to approach the use of new technologies, particularly those popular among youth such as smart phones, is to consider their affordances in light of what they term the 3Ts: text, tools, and talk (see also Philip & Olivares-Pasillas, 2016). To consider the *texts* afforded by mobile technologies in the classroom is to consider how and what texts are made available through the use of mobile technologies, as well as to consider and develop the new literacies necessary to read and analyze these texts (Philip & Garcia, 2013). Viewed as *tools*, mobile technologies offer unique and unprecedented opportunities for students to "collect, analyze, represent, and communicate data in elegant ways to audiences across the globe" (Philip et al., 2013, p. 110), thus, thoughtful consideration of the instructional affordances of mobile technologies in the classroom involves thinking about how particular tools "allow students to

[6] Please note that Philip and Garcia (2013) do not couch their framework regarding tools, text, and talk—the 3Ts—within Philip et al.'s (2013) three paradigms of technology and cutting-edge content as an end, a means, and an equalizer nor do Philip et al. (2013) position the 3Ts as components of this paradigm. For purposes of fluidity and organization, I have inserted a brief review of text, tools, and talk within the paradigm of technology and cutting-edge content as a means because this paradigm mounts a critique of the blind adoption of technology and cutting-edge content as intrinsically transformative. The 3Ts help disentangle this misinterpretation and misrepresentation of technology in the classroom and provide a guideline for the types of considerations that should be made before adopting new technology as instructional tools. The reader should also note that the 3Ts touch on themes and issues discussed in all three paradigms.

meaningfully collect, represent, visualize, analyze, or communicate texts for a particular set of learning goals" (Philip & Garcia, 2013, p. 313). The texts and tools afforded by technology in the classroom, however, are of little relevance to learning without meaningful communication and interaction within the classroom community that both fosters a connection between the teacher and students, and incorporates the type of discourse that carries clout in a particular discipline. Thus, incorporation of new technologies in the classroom should consider how the texts and tools afforded by new technologies support and build upon rich classroom discourse—that is, *talk*. All things considered, Philip and Garcia (2013) offer the 3Ts as generative lenses that can be useful in thinking about the affordances and pitfalls of introducing new forms of technology into the classroom. Taken together, the 3Ts can help ensure that technology and cutting-edge content actually function as learning tools for data science.

Converging Critical Perspectives in and for STEM Reform

In the preceding literature review I discuss existing scholarship of great significance for understanding the equity-oriented impetus for reforming traditional STEM education. In discussing approaches to STEM reform, I included a critical call for the reframing of scientific knowledge. The myth of scientific objectivity has informed traditional science for far too long, creating the illusion that science is derived from a universalist epistemology, is factual, and is value-free (Brickhouse, 1994; Stanley & Brickhouse, 1994; Calabrese Barton, 1998; Calabrese Barton & Yang, 2000; Bang et al., 2012). Critical scholars in the learning sciences argue that unless settled expectations in science are desettled (Bang et al., 2012), the ways of knowing and intellectual contributions of students from non-dominants groups will continue to be excluded from STEM discourse, education, and careers. This is due to the fact that the culture of power (Delpit, 1988; Calabrese Barton & Yang, 2000), a set of unspoken and unwritten rules and value systems, that regulate power in American society function to privilege white middle-to-upper-class heterosexual males through the valuation of hegemonic epistemologies and the simultaneous denigration and subordination of non-dominant groups, their cultures, and epistemologies.

Scholars outside of the math-sciences attest to the fact that positivist views of scientific objectivity have come to characterize Big Data and Big Data analytics within data science-related fields (Ekbia et al., 2015; Ebach et al., 2016). As a result, digital sociologists have begun to offer critical perspectives concerning the role Big Data play in reproducing disciplinary and social inequalities, and intensifying managerialism in schools; (Selwyn, 2015; boyd and Crawford, 2012). Boyd and Crawford (2012) and Couldry (2014) argue that the privileging of advanced skills of statistical analysis in data science-related fields, and dismissal of the affordances of qualitative inquiry reproduce existing tensions regarding the legitimacy of quantitative research versus qualitative research. Therefore, by challenging the legitimacy of qualitative perspectives and making claims to objectivity the myth of Big Data objectivity in the infant field of data science exists in service of the enduring culture of power and settled expectations in STEM (Ebach et al., 2016; Ekbia, et al., 2015; Couldry, 2014; Bang et al., 2012; boyd and Crawford, 2012; Calabrese Barton and Yang, 2000; Delpit, 1988; Harris, 1995). Additionally, Ignatow and Robinson (2017) add that widening inequalities in who has access to technology and training in data analytics parallel and exacerbate existing socioeconomic inequalities. When we consider the well-documented lack of access to quality educational experiences and opportunities for students from non-dominant groups, and the fact that they do not enjoy equitable representation in data science-related fields and STEM in general, it is not difficult to discern how emergent disciplinary and social divisions perpetuated via the myth of

Big Data objectivity are poised to perpetuate deep-seated educational inequity for non-dominant students in the new field of data science (Martin et al., 2015; Pérez Huber et al., 2014; Covarrubias, 2011; Barr et al., 2008). Thus, unless administrators, educators, policy makers, and researchers are willing to open sustained and critical conversations about the politics of data and technology in the era of Big Data, data science risks perpetuating the enduring legacy of epistemological exclusion and educational inequity in STEM for non-dominant students.

Thus, in an effort to reform STEM education and career outcomes for non-dominant students, educators, researchers, and policymakers must actively work toward redefining science as co-constructed knowledge (Calabrese Barton, 1998). A number of researchers in the learning sciences have embarked on this mission by engaging in scientific exploration with youth from non-dominant groups both in and out of school (Calabrese Barton, 1998; Bang et al., 2012). Efforts highlighted in this review demonstrate that scientific knowledge can be co-constituted in ways that incorporate discourse practices and learning processes in ways that not only grow scientific knowledge and curiosity for students and teachers, but also work toward addressing issues relevant to the lives of students (Calabrese Barton, 1998; Bang et al., 2012). While these efforts provide models for developing ways to desettle settled scientific expectations, not all STEM reform efforts reflect these approaches and these approaches are not a one-size-fits-all. However, an element shared by these models consists of engaging in scientific inquiry in ways that explicitly sought to build on student interests, cultures, and epistemologies, and thereby sought to provide opportunities for students to identify with science.

Just as STEM reform efforts must actively pursue an agenda to support the development of strong STEM identities among non-dominant students, so should assessment of learning outcomes in STEM reform initiatives (Carlone et al., 2011). Scholarship included in the review also argues that a systematic analysis of STEM-equity outcomes must avoid the dilemmas inherent in high-stakes assessments and accountability measures and take into account the opportunities available for students to identify with STEM in personally meaningful and enriching ways (Selwyn et al., 2015; Khalifa et al., 2014; Kaufman et al., 2014; Carlone et al., 2011; Cobb et al., 2009; Cobb & Hodge 2010; Cobb & Hodge, 2002). Popular efforts to engage youth in personally relevant scientific study have sought to capitalize on the ubiquity of and widespread access to mobile technologies to learn about data. As a result of STEM reform efforts, increasing public availability of large data sets, and new technologies, data literacy has emerged as particularly popular (Tygel & Kirsch, 2015; Philip & Garcia, 2013; Philip et al., 2013; Calzado Prado & Marzal 2013; Hogenboom et al., 2011; Stephenson & Schifter Caravello, 2007). This review discusses nuances, limitations, and affordances of pursuing data literacy in STEM reform. Ultimately, the scholarship discussed above contributes to my study in its critical reading of scientific knowledge, STEM learning outcomes, new technologies for learning about data, and emergent popular multiliteracies.

CHAPTER THREE
Methodology

Introduction to the Energize Project

My dissertation examines an educational STEM reform effort, hereafter referred to as Energize (all names are pseudonyms), implemented in a large urban school district that serves a predominantly Latino student population. Funded by the National Science Foundation, the course was part of a national imperative to increase the presence of women and people of color in the fields of science, technology, engineering, and mathematics. Energize is a collaborative project between a large urban school district and university-based researchers in the fields of education and in science, technology, engineering, and mathematics (STEM) that seeks to increase and inspire student interest in computer and data science through the design, development, and implementation of computer science and data science units that can be inserted into a number of math and science courses as well as a standalone course entitled Introduction to Data Science (hereafter referred to as IDS). Energize consists of a number of teams responsible for curriculum writing for math and science units, professional development for in-service teachers, tech support, research, and evaluation. I began working with Energize in 2013 as a graduate student researcher with the research team. The research team was responsible for the collection of field notes; audio-recorded teacher and student interviews; and video-recorded classroom observations of data science units as well as the standalone IDS course.

IDS. The standalone IDS course was introduced in the 2014-2015 academic year as a result of an iterative process that arose from the design and implementation of computer and data science units in pre-existing math and science courses. One of the goals of Energize in its design of IDS was to implement an inquiry-based pedagogy as defined by the local school district. IDS drew heavily from fields of computer science, data science, and statistics and was formally recognized as a statistics core math course, approved as a "c"[7] course in the University of California A-G subject eligibility requirements. The course consisted of a combination of classroom lessons and labs. Lessons introduced students to data science skills and concepts, while labs provided students the opportunity to learn to code in R[8] using, open-source data analysis software, RStudio to implement concepts imparted through the lessons. IDS was piloted in 10 classrooms in 10 different high schools during the first year of implementation during the 2014-2015 academic year. In the second year of implementation, IDS was taught in 31 classrooms at 26 different high schools.

The IDS curriculum. In its totality, the IDS curriculum consisted of four units to be taught over the course of one academic year. The classroom that I observed was only able to cover Units 1 and 2. As designated in the curriculum, Unit 1 consisted of three sequential themes: "Data Are All Around"; "Visualizing Data"; and "Would You Look at the Time?" The themes were explored through a total of 17 lessons and eight labs with roughly one instructional day allotted for each. The unit also included two practicums, one halfway through the unit and another and the end. Although they were not formal assessments, practicums called on students to draw from cumulative knowledge on topics gained in preceding lessons and labs. The curriculum indicated five days for end-of-unit projects and presentations. Unit 2 consisted of four themes: "What is Your True Color?"; "How Likely is it?"; "Are you Stressing or Chilling"; and "What's Normal?" Figure 3.1 indicates general learning goals relative to each of the four themes in the curriculum (for an overview of the curriculum, see Appendix A).

---

[7] The University of California A-G subject eligibility requirements designate a letter for each academic subject. The "c" requirement refers to the subject of mathematics and is met after a student has taken "three years of college-preparatory math, including or integrating the topics covered in elementary and advanced algebra and two- and three-dimensional geometry" (UC A-G Subject Requirements).

[8] R is a programming language.

Unit 2 Themes	Learning Goals
What Is Your True Color?	• Students learn that statistics are useful for making sense of large amounts of data. • Students learn to create numerical summaries of data through the use of measures of center (mean, median) and measures of spread (mean of absolute deviation [MAD], interquartile range [IQR]). • Students learn to determine which measures are appropriate for a given distribution and how to calculate them.
How Likely is it?	• Students learn that probability measures the long run frequency of occurrence for chance outcomes. • Students learn that probabilities can be approximated through simulations in RStudio and also via mathematical calculation and learn to do so.
Are You Stressing or Chilling?	• Students learn that permutations of data provide a model that shows us how the world behaves if chance is the only reason for differences between groups or variables. • Students learn to determine if outcomes occur by chance or design by analyzing simulated probabilities against real ones.
What's Normal?	• Students learn that the Normal curve describes the distribution of many real phenomena. • Students learn that drawing the Normal curve over histograms is useful for determining if a distribution is Normal. • Students learn that typical values are located toward the center of the curve and less typical values and extreme values are located farther from the center.

Figure 3.1 Learning goals were derived from the IDS curriculum for Unit 2.

The themes were explored through 18 lessons, nine labs, and three practicums and were to be taught over the course of 30 instructional days; and a final five days dedicated to end-of-unit projects and oral presentations. Furthermore, every theme involved the cultivation of an enduring understanding; a data-related story to promote student engagement; and specific learning objectives for statistics, mathematics, data science, applied computational thinking using RStudio; and real-world connections. Additionally, the curriculum indicated specific learning objectives and data file/data collection methods per each theme.

The site of observation: Medical Science High School. The dissertation draws from observational classroom data captured through video recording and audio-recorded student and teacher interviews collected during the second-year implementation of IDS Unit 2 at Medical Science High School (hereafter referred to as MSHS).

MSHS is a small pilot high school located on one unifying campus that is home to two other small pilot high schools in the working-class urban neighborhood of Easton Park, located in the southeast region of a large metropolitan county. US census data estimates for 2015 indicate that at over 90 percent, the population in Easton Park is overwhelmingly Latino. Latino residents are mainly of Mexican descent; however, about three percent are of Salvadoran descent. Additionally, over 50 percent of its overall population is foreign-born. The percentage of Easton Park residents with less than a high school diploma is about 2.5 times higher than that of the entire county and close to five times higher than that of the country as a whole (Table 3.1). What's more, the percentage of Easton Park residents with a high school diploma or higher is about two times lower than that of the county and the country as a whole. The estimated median household income in Easton Park in 2015 was $35,917 while the median household income in the county and the country was $56,196 and $56,516 respectively. In addition, the per capita income for Easton Park residents was $12,496.

MSHS is part of one of the country's largest school districts, with a student body comprised of over 97% Latino students. The unifying school campus, and its three pilot schools, opened in 2012 in an effort to address and alleviate the over-populated and over-enrolled comprehensive local high school that has, for many years, matriculated high school-age (grades 9-12) students living within Easton Park and neighboring municipalities.

Easton Park Educational Attainment and Income Comparison for Adults 25 and Older, 2015			
	Easton Park	County	U.S.
Education			
Less than a high school diploma	57.3%	22.7%	11.6%
High school diploma or higher	42.7%	77.3%	88.4%
Bachelor's degree or higher	5.7%	30.3%	32.5%
Income			
Median household income	$35,917	$56,196	$56,516
Income per capita	$12,496	$28,337	
Persons in poverty	27.3%	18.2%	13.5%

Table 3.1 Note: City and county figures were derived from census American Community Survey (ACS) 5-year estimates located at www.census.gov. Figures for the U.S. were drawn from Ryan & Bauman (2015) and Proctor, Semega & Kollar (2016).

With a student enrollment of 686[9] students, MSHS is small school in comparison to Easton Park High School (hereafter referred to as EPHS), the larger comprehensive high school, which enrolls upwards of 1,500 students. This means that the student body and faculty size are much smaller, there are fewer course offerings, and there is a sense of community. During my initial visit to MSHS, the principal expressed that the strong sense of community meant that students were familiar with the faculty and administration and vice versa. Additionally, MSHS is a Title 1 school with over 80% of the student body eligible for free or reduced lunch.

School schedule. MSHS offered year-long courses and operated on a block schedule which means that out of eight subject classes, students attended all odd numbered classes one day and even numbered classes another. The weekly schedule consisted of alternating even/odd days throughout the week so that students would ultimately attend all classes an equal amount of time. The duration of each class was 90 minutes except for the first block of the day which was 110 minutes, allowing 10 minutes for students to have school-provided breakfast in class. Classes were also shorter in duration on Tuesdays, designated "PD Day," and lasted 60 minutes.

[9] Data on MSHS regarding the student body, student enrollment, and free/reduced lunch eligibility was gathered by the California Department of Education (CDE) through the California Longitudinal Pupil Achievement Data System (CALPADS).

My role in Energize. As a member of the research team, I was assigned to observe one of 31 implementation sites. I observed the classroom of one of two teachers identified by the school district liaison as the strongest in the program. I was placed at this particular site because of my proximity to it relative to other sites during the first year of implementation. I subsequently remained at the same site for the second year of implementation establishing rapport with the IDS teacher, students, and school administrators. Proximity was important to facilitate my ability to conduct daily classroom observations and scheduled interviews, some of which required multiple site visits per day on my behalf. Additionally, the school site was located in the city wherein I attended high school (EPHS) and several teachers who once taught at my high school were now teaching at MSHS, including the principal who was one of my teachers in high school. My familiarity with the city, the population, and some members of the school faculty and administration facilitated my ability to establish rapport and open lines of communication with faculty and staff. This proved indispensable in my ability to understand the social, class, and educational dynamics that characterized the school.

Data Collection

This qualitative study was initially exploratory. Prior to entering the classroom, we (the research team) knew that the curriculum sought to engage students in personally meaningful and powerful ways of collecting, analyzing, interpreting, presenting, and managing data through the use of internet-enabled mobile and computer devices, cutting edge content (Big Data), coding, and a custom app designed by the project explicitly for this program. Additionally, we knew that the two teachers we observed were especially strong mathematics teachers dedicated to the quality learning of their students.

Further, this study relied on participant observation, one-on-one interviews with students, and field notes as methods of data collection. Specifically, I engaged in moderate participant observation during my site visits. This means that although I actively engaged with students and oftentimes answered questions related to the course (when possible), I sought to retain my position in the classroom as the researcher. I will discuss my reason for doing so in the following section. During the first semester, I introduced myself to students as an alumnus of EPHS and shared that I lived the earlier part of my childhood in Easton Park before my family and I moved to South Central Los Angeles. I also shared that I graduated from UCLA with my bachelor's, master's degree, and was currently working on my doctoral degree and welcomed any questions they might have for me regarding my upbringing in Easton Park or my time at UCLA. Thereafter, I would circulate throughout the classroom and greet students. I made an effort to commit their names to memory as soon as possible so that I would be able to address them by name and work toward establishing rapport.

Going into the classroom, I very much viewed myself as a researcher but it soon became clear that students viewed me as a teacher or school administrator and referred to me as "Miss"—a name/title students used when speaking with any adult female faculty or staff member—although I initially asked that students call me by my first name. In addition, students began asking me questions regarding the curriculum and their classwork. Although I did not answer all questions—as I was not trained to teach or problem-solve through the curriculum as teachers were—I answered questions and provided help whenever I could. When I could not help a student with a question, I encouraged them to ask the teacher, Ms. Gellar. I also got the sense that if students were not viewing me as a teacher, they were definitely viewing me as an authority figure. This I gathered through student reactions whenever I walked by which included hushing each other and looking at me with alarm, in fear that I might reprimand them for talking about extraneous topics having nothing to do with their classwork. Similar situations would initially take place in the computer lab as well, as students hurried to close non-course related windows on their screen as I walked by. When this happened, I would either smile at students and continue walking around or stop and engage them in conversation regarding their non-course related activity. I approached these situations in this manner so that students would see that not only was I not there to monitor them, I might actually share interest in things that interested them. These exchanges were brief so as to not encourage behavior that might be viewed negatively by Ms. Gellar or hinder students' abilities to attend to their work, but simply meant to indicate that my role in the classroom was not to monitor or reprimand them. For example, on one occasion while students were supposed to be working on a lab in RStudio, three students were gathered at the end of a row of computers looking at pictures of the UCLA marching band on Google Images. I happened to be a few seats away and as I made my way over to exit the row of computers one of the students, Carlo, shot a nervous glance in my direction. I looked at the screen, smiled, and asked if any of them were in marching band. A female student in the group, Glenda, said she was and Carlo added that she was accepted to UCLA and was going to join the band. I expressed enthusiasm for her acceptance and desire to join the band and congratulated her. Over the course of the year students' guarded nature during my presence subsided as many of them continued with their non-course related conversations and activity whenever I happened to be nearby, presumably having realized that my role in the class was not one of an authority figure.

Video-recording of observational data. During the first semester, I did not video record my observations. Due to the fact that I was an outsider entering the students' classroom, I did not want to seem intrusive by bringing in a video camera from the onset. When I began video recording, some students were visibly shy or uncomfortable at first but this did not happen until the second semester when Ms. Gellar began teaching Unit 2. If I had begun video recording during the first semester, I strongly believe it would have had a negative impact on my efforts to establish rapport with students. Before I began video recording, Ms. Gellar helped me share with students that I would be bringing a video camera to record the class. I assured them that the video recordings were confidential and that their likeness would remain anonymous. I reminded them of the consent forms they and their parents signed and told them that their rights and safety were protected. I also shared that video recordings would not be made public and were collected for analysis by the research team of what aspects of the curriculum work and/or do not work, and how to improve the course for future students. During the first day of video recording, a male student, Luis, asked, within earshot of other students, if I was going to show the recording to Ms. Gellar. I assured him that I would not and that the video would only be viewed for research purposes. In total, I was able to capture 34 hours and 12 minutes of video-recorded classroom observations.

Audio-recorded one-on-one interviews. Toward the end of the 2015-2016 academic year, I conducted one-on-one guided exit interviews with 12 IDS students, six females and six males, that sought to get a sense of the following:

- Their developing STEM identity
- Self-identification with science-doing
- Perception of the real-world importance of data science
- The significance of data science for themselves and their communities
- Their future academic endeavors

The interviews also included a question for two students, Kim and Diego, who were taking the IDS class for the second time after failing it the year prior, which was also the first year of the implementation of the IDS curriculum. This question asked them to speak on how their experience taking IDS the second time differed from their experience the first time and what, in their opinion, accounted for differential experiences. Additionally, I asked follow-up questions for further elaboration in accordance with students' responses; hence, many of these are not included in the guided interview protocol (see Appendix B for the interview protocol).

I consulted with Ms. Gellar to select students that would represent a range of performance levels. Ms. Gellar judged student performance based on a number of aspects of their classroom participation including performance on assignments, participation in class, and understanding of IDS skills and concepts. I scheduled interviews with students either during their free block (for those that had one) or advisory block so as to not interfere with their participation in other classes. By interviewing students during free and advisory blocks I was able to continue conducting video-recorded observations in the mornings when Ms. Gellar taught IDS.

When interviewing students, Cobb et al. (2009) found it more useful to associate the interviewer "with the school rather than with the team conducting the design experiment," by making sure that the interviewer was not involved in conducting or video recording the instructional sessions in the design experiment classroom" (p. 57). This decision was made given Cobb et al.'s (2009) understanding that "an interview is a social event in which the interviewer and interviewees present themselves to and recognize each other in particular ways" (p. 57). In their case, they ultimately found that their "strategy appears to have been effective in that the students did make a number of negative observations about the design experiment class" (p. 58).

Given my own experience growing up as a Latina student in predominantly Latino schools that were under-funded and over-enrolled; research on the disproportionate disciplining and criminalization of Latino and Black youth; and teachers' and administrators' "inherited professional roles in the ongoing surveillance, management, and disciplining of youth" (Raible & Irizarry, 2010, p. 1197; Rocque & Patternoster, 2011; Morris Perry, 2016), I thought it more appropriate to disassociate myself with the school, namely *school authority*, and instead associate myself with the reform initiative, specifically as a researcher with the research team. I felt it necessary to impart to students that my observations would in no way lead to negative implications for them. This is something I actively sought to convey throughout the entire period of observation, and students' eventual willingness to speak freely and be off-task in my presence indicated to me that I had presented myself and they had recognized me in a particular way that ultimately made me privy to significant insights regarding genuine student perceptions of learning in their IDS class and their sentiments regarding data science as a field of study. While the interview protocol was designed with the expectation that interviews would last around 30 minutes, the duration of student interviews ranged from 19 minutes to an hour. In total, audio-recorded interviews lasted six hours and 56 minutes.

Classroom student body. Ms. Gellar's IDS classroom consisted of 41 students of which 24 were female students and 17 were male. Two female students were juniors and the rest of the students in the class were seniors. There were also two seniors, Kim and Diego, taking IDS for the second time after failing the course the year prior. Kim and Diego and a number of other students needed to pass IDS in order to graduate. The school counselor placed a small group of students in IDS because they expressed interest in the course after hearing about it from peers and from Ms. Gellar during the first year of implementation. A larger group of students needed the course to graduate, as it would allow them to meet the math requirement for graduation. Within this group, a number of students were placed in IDS after having failed Algebra II the year prior.

During the first year of implementation a substantial proportion of students enrolled in IDS had failed their previous math class. Ms. Gellar expressed frustration at the counselor's treatment of IDS as a last resort for students who had failed out of Algebra II because it undermined the potential of IDS to cultivate critical science learning by positioning it as a remedial class. As expressed by staff in Energize meetings, this approach to treating IDS as a "dumping ground" was taking place at a number of implementation sites and presented obstacles for the implementation of IDS in a number of ways that I personally observed in Ms. Gellar's class. For example, lessons assumed that students taking IDS had at least a basic algebraic understanding of math concepts from which it sought to build on, but many students required additional scaffolding to bridge the gap between concepts they were familiar with and those they were assumed to be familiar with. This, in part, affected pacing of lessons and labs as Ms. Gellar had to go through lessons at a slower pace, provide students extra time to complete labs, supplement the curriculum with instructional material aimed at helping students make better connections with statistical concepts, and at times provide brief reviews of math skills that students should have learned prior to IDS in order to facilitate their understanding of the skills and concepts imparted in the course. While getting through the IDS curriculum was challenging during the first year of implementation in Ms. Gellar's class, the second year showed improvement as a lower number of students were placed in IDS as a result of failing Algebra II and a number of students were [at the very least] initially motivated to participate in the course having chosen to be in it. Ms. Gellar attributed this change to the retirement of the previous school counselor who was responsible for placing students in courses.

Additionally, while IDS provided some students the opportunity to meet graduation requirements, enrollment in IDS also held implications for students' math pathways at MSHS, which had limited course offerings, compared to larger comprehensive high schools. In an interview that took place during my first year observing her classroom, Ms. Gellar, a seasoned math teacher well-acquainted with the high school math sequence, expressed that if a student did not intend to eventually take an AP mathematics class in high school, then taking IDS would be a good idea and if they later decided to move on in the math sequence they could subsequently take Algebra II. However, the opportunity to move forward in the high school math sequence would require that students take IDS in their junior year, allowing them time during their senior year to take Algebra II, given that IDS does not have a corresponding math sequence and there is no particular course explicitly designed to follow IDS at MSHS[10]. For students inclined to take math courses beyond Algebra II, participation in IDS would have to occur earlier than their junior year. Necessarily, this would require that students plan accordingly well ahead of their final high school years. Ms. Gellar expressed a different sentiment when it came to students who intended to take an AP math class. She felt it was important for students who had a desire to take an AP math class to have a path available for them to do so rather than to be programmed into a place where moving forward in the high school math sequence would eventually prove very difficult. For these reasons, the placement of students into IDS at MSHS had implications

[10] While Ms. Gellar and Energize staff members have expressed that students can take Statistics or AP Statistics after IDS as a likely follow-up to IDS, three issues are worth considering. The first is that MSHS does not offer such a course; secondly Statistics and AP Statistics were not designed to cultivate data science literacies, as is the case with IDS; lastly, they are not instantiations of STEM-reform efforts but are instead traditional courses already established in many high schools.

for both how the curriculum was ultimately implemented, and the affordances and limitations to student pathways in math.

Classroom lessons. Unit 2 of the IDS curriculum consisted of 18 lessons taught in the classroom (as opposed to the computer lab). In the classroom, students were seated at individual desks that were organized into 14 groups of three. Figure 3.2 illustrates the general layout of the classroom. Ms. Gellar started each lesson with a projected slide presentation that introduced the objective of the lesson, vocabulary words, and essential concepts as written in the curriculum. The lessons also included classroom activities, opportunities for group reflection, and whole-class share-alouds.

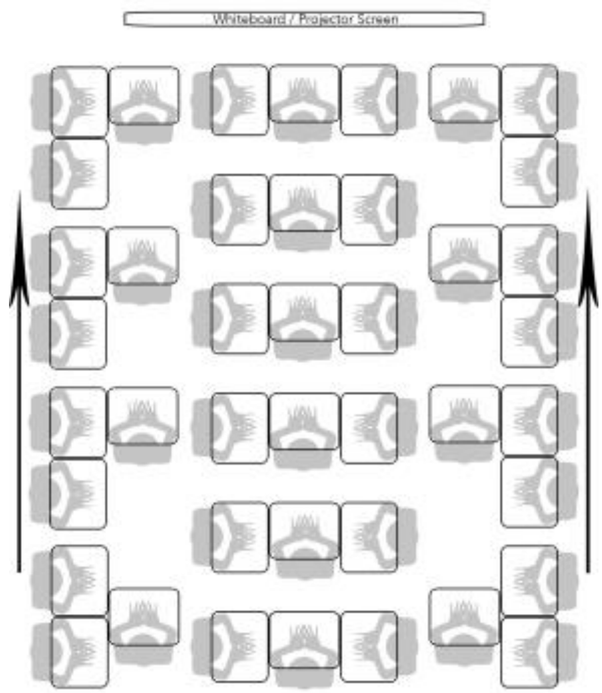


Figure 3.2 General student seating layout of Ms. Gellar's IDS classroom.

Ms. Gellar created lesson worksheets that functioned similarly to the journal that the curriculum called for wherein students were to write reflections on skills and concepts learned during the class. The worksheets consisted of fill-in-the-blanks, complete-the-sentence, brief writing prompts, and included space for students to record data gathered during classroom activities. Also, some visuals that were originally intended as handouts were incorporated in smaller form into the worksheet. Ms. Gellar would pause at moments during the lesson to allow students to copy enduring understandings, lesson objectives, essential concepts, and definitions into the worksheets (for an example of a lesson worksheet, see Appendix C). At the end of each lesson, students turned worksheets in for grading. Once Ms. Gellar handed back graded worksheets students placed them in their IDS binder for their own reference.

RStudio labs. Unit 2 of the IDS curriculum consisted of eight RStudio labs that were interspersed throughout the unit. The class would walk to the computer lab to complete the labs [11]. The labs were designed to introduce students to coding through the software R, a programming language; and RStudio, an open-source data analysis software that student accessed via the internet. The RStudio interface consists of four panes: source, workspace/history, R console, and the file and plot viewer (Figure 3.3).

Students learned to load the labs designed for the curriculum by entering "load lab ( )" in the console pane and indicating the desired lab number inside the parenthesis. Once loaded, a lab would appear in the file and plot viewer, providing students with a slide presentation that guided them through the coding process. These lab slides were also authored by Energize. Labs progressively got more difficult as they required students to incorporate coding knowledge learned in previous labs to analyze large outside datasets, such as the Center for Disease Control

[11] The curriculum was designed with the assumption that lessons and labs would be completed in the same setting—the classroom. However, during the first year of implementation, which took place during the 2014-2015 academic year, Ms. Gellar experienced technical difficulties with the laptops that MSHS provided. She expressed that all of the computers needed to be updated and some would not power up. That year, she made the transition from trying to implement the labs in the classroom to having students walk over to the computer lab. While she had access to the computer lab, this was only because IDS was scheduled during the computer teacher's conference period.

data, and datasets that contained their own data, which they input/collected using the Energize app (participatory sensing component).

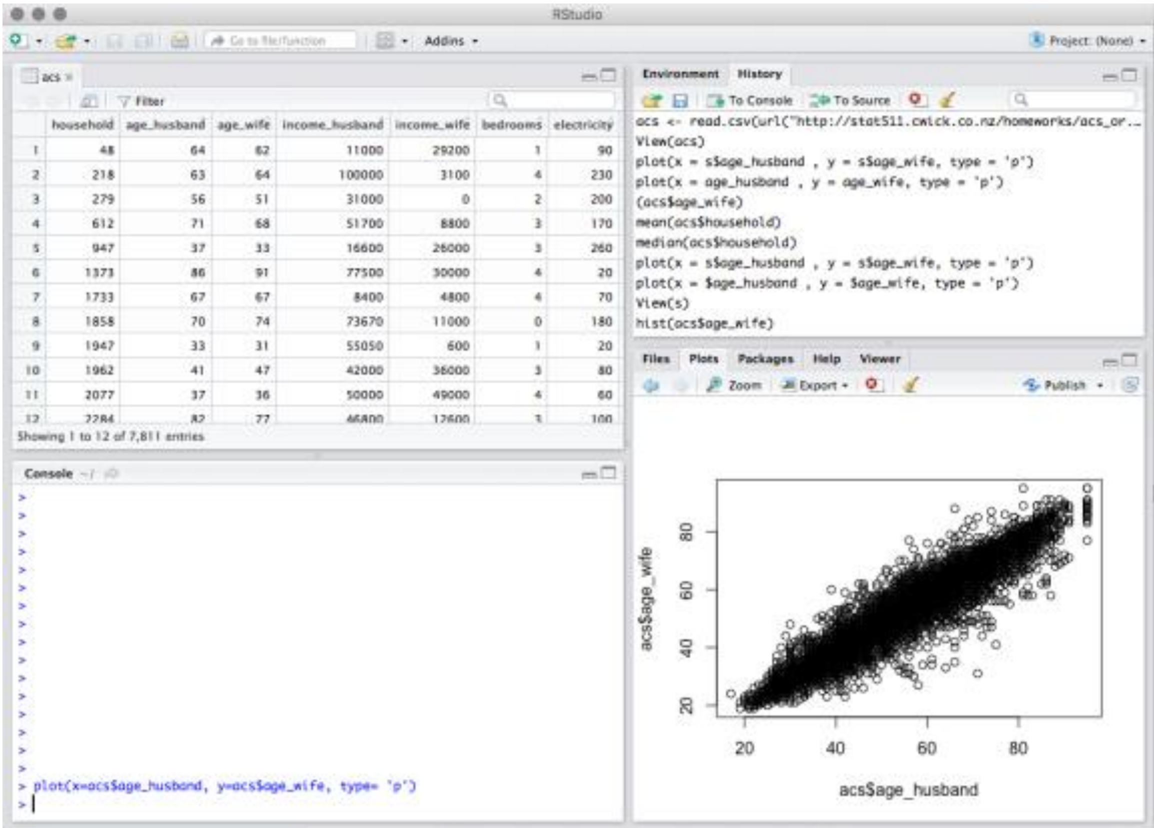


Figure 3.3 This is an example of what RStudio looks like and what students saw upon launching RStudio. The panes are located as follows: source pane (top left); workspace/history pane (top right); console pane (bottom left); file and plot viewer (bottom right).

Students were seated with their classroom groupmates in rows of six with two groups per row (see Figure 3.4 for the lab student seating layout). Classroom dynamics in the computer lab were different due primarily to the fact that labs were completed individually with the lab presentations guiding students through to completion. In the labs, Ms. Gellar took on more of a facilitator role where she introduced the lab, reminded students of deadlines, and allotted time for completion. She also walked around the classroom providing help to students who asked for it. Although students mostly worked independently to complete the labs, they visited each other for help, checked each other's progress, and compared lab responses.

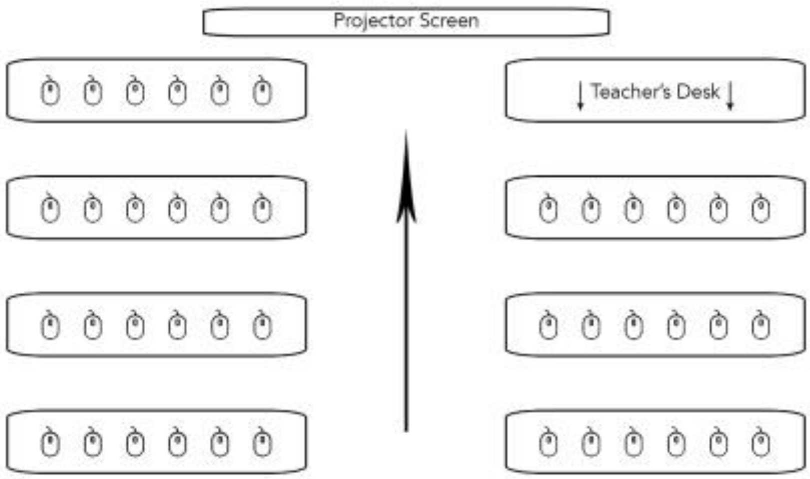


Figure 3.4 MSHS computer lab student seating layout.

Like the class worksheets, Ms. Gellar also created lab Word files (for an example see Appendix D) with a series of questions for students to answer as they typed in different codes as indicated in the lab slides. It is worth noting that unlike a traditional math class, students did not have a textbook to refer to. Additionally, homework was highly minimal and when assigned typically consisted of completing lesson worksheets. Ms. Gellar instituted both the lesson worksheets and the lab Word files to incorporate a writing component, have students engage more deeply with IDS concepts, and to have students complete physical work, which she also needed in order to include assignments in her grade book. The lesson worksheets were particularly useful for students who used them as reference material in the absence of a textbook.

Methodology

To analyze the data corpus, I adopt Cobb, Gresalfi, & Hodge's (2009) "interpretive scheme for analyzing the identities that students develop in mathematics classrooms" as an analytical framework. Their interpretive scheme is rooted in grounded theory, which seeks to build theory from data (Glaser & Strauss, 1967/1999; Strauss & Corbin, 1990), and "focuses directly on the relations between the microculture established in particular classrooms and the identities that students are developing in those classrooms" (Cobb et al., 2009, p. 41). As such, an initial step in the data analysis involves identifying recurring patterns in interactional classroom activity (Cobb et al., 2009; Cobb & Hodge, 2002; Cobb, 1999; Cobb & Hodge, 1992). Furthermore, to systematically identify patterns in classroom activity I conduct a thematic analysis of video-recorded classroom observations. This method, useful for "identifying, analyzing and reporting patterns (themes) within data" (Braun and Clarke, 2006, pp. 79) allows me to examine salient themes relating to students' relatively new engagement with data artifacts and data science. Before discussing the specific methods I employ in my data analysis, I will first discuss the theoretical contributions that Cobb et al.'s interpretive framework offer to my analysis of classroom norms and practices as they relate to the development of student identities as data science doers (Cobb & Hodge, 2002).

Toward an Analysis of the Classroom Microculture and Student Identities

The classroom microculture. A significant construct in Cobb et al.'s (2009) interpretive scheme is that of the mathematics classroom microculture (Cobb & Hodge, 2002). The mathematics classroom microculture can be understood as consisting of "a set of locally instantiated practices…[,] dynamic and improvisational in nature," (Gutierrez, Baquedano-Lopez, & Tejeda, 1999 cited in Cobb & Hodge, 2002, p. 261) that play out in the mathematics classroom. The mathematics classroom microculture is comprised of three different aspects: social norms, sociomathematical norms, and classroom mathematical practices (Cobb & Hodge, 2002). In adopting this framework to the IDS classroom, I use analogous terms socio-data-scientific norms in place of sociomathematical norms, and classroom data science practices in place of classroom mathematical practices. Accordingly, social norms are general normative classroom behaviors that are not specific to the IDS classroom—these are transferable to other subject courses. An example of an emergent social norm in IDS is that students consulted with peers when having difficulty finding a solution to a problem. Socio-data-scientific norms are normative classroom behaviors that are specific to the data science classroom—these are not transferable to other subject courses. An example of a socio-data-scientific norm that emerged in IDS is that when in the lab setting, students routinely reviewed their coding history as a reference tool to recall previously used codes. Lastly, data science classroom practices are interactional activity, including speech acts, which are unique to the data science classroom. In IDS, this included practices such as consistently scrolling through coding history in RStudio as a means of recalling necessary codes necessary for the completion of labs. An analysis of the norms and practices that emerged in the IDS classroom is essential to understanding the nature of student identity constructs as constituted in data science classroom; these include normative identity and personal identity.

Normative identity. The normative identity that students develop in the mathematics classroom as math-doers is a collective identity developed through classroom norms and practices as they play out in that classroom. Normative identity, as defined by Cobb et al. (2009), "comprises both the general and the specifically mathematical obligations that delineate the role of an effective student in a particular classroom" (p. 43). The normative identity, regardless of its nature, is a co-constituted venture jointly negotiated by both the teacher and students. General classroom obligations have to do with legitimate *ways* students are able to exercise agency in the classroom and *to whom* students are accountable. Specifically mathematical obligations have to do with *what* students are accountable for mathematically (Figure 3.5 below). Hence, obligations refer to actions and behaviors that students must engage in to meet the expectations established in the classroom to be considered successful and effective math-doers. Cobb et al. (2009) follow

Engle (2006) and Hull and Greeno (2002) in outlining two aspects of general classroom obligations "that prove to be important when documenting the role of an effective mathematics student and thus the normative identity established in a particular classroom [:]…the *distribution of authority* and the ways that students are able to *exercise agency*" (Cobb et al., 2009, p. 44, emphasis in original).

The distribution of authority determines to whom students are held accountable in the classroom. For example, in classrooms that incorporate discursive practices of students from non-dominant groups and/or follow an inquiry-based instructional design, authority is distributed to both the teacher *and* students (see Bang et al., 2012; Carlone, 2004). Conversely, in a traditional mathematics classroom where the teacher is the primarily determinant of what qualifies as legitimate student contributions, authority is distributed solely to the teacher. In effect, *how* authority is distributed in a mathematics classroom determines the ways in which students are able to exercise agency. Cobb et al., (2009) describe two forms of agency: conceptual agency and disciplinary agency.

Conceptual agency "involves choosing methods and developing meanings and relations between concepts and principles" (Cobb et al., 2009, p. 45). When students engage in conceptual agency, they have the opportunity to make decisions regarding the appropriate use of methods and to participate in the process of meaning-making with regards to mathematical understandings and solutions. To exercise conceptual agency is to make connections between the mathematical artifacts, skills, and concepts learned in the mathematics classroom, and to provide solutions backed by strong mathematical reasoning. Thus, mathematical understandings and classroom activity have the potential to be personally meaningful to students. Conceptual agency and disciplinary agency differ in the *ways* in which agency is exercised. Disciplinary agency “involves using established solution methods,” (Cobb et al., 2009) to solve mathematical problems. When students engage in disciplinary agency, they take up decision-making opportunities to decide which solution methods to use in solving mathematical problems, albeit without developing strong conceptual reasoning, and thus justification, for selecting particular methods.

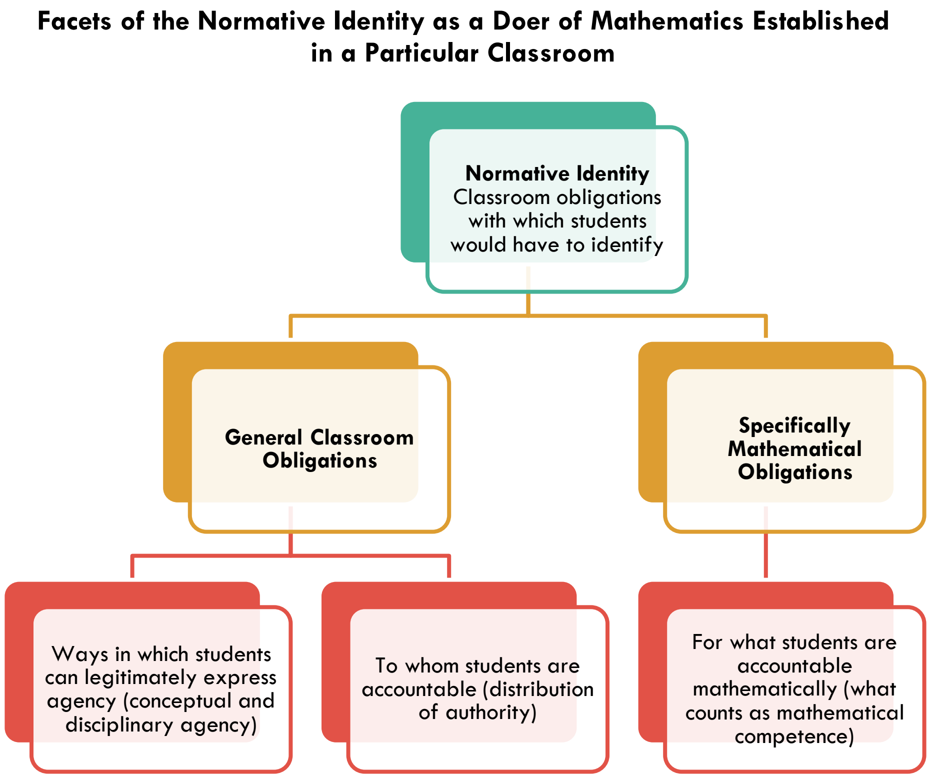


Figure 3.5 The schematic model provided above is a replication of Cobb et al.’s (2009) model located on p. 46.

Generally speaking, opportunities for students to develop strong affiliations with mathematical activity, by viewing it as structurally or situationally significant to them as individuals, can be supported or obstructed by the nature of authority distribution in the classroom and, subsequently, by the types of agency students are able to exercise in the mathematics classroom.

Further, in order for students to identify with classroom mathematical activity as effective math-doers as constituted in the classroom, they must first identify with the aforementioned obligations. This means that instead of viewing obligations as obligations to others, such as to the teacher, students begin to view obligations as obligations to themselves. An analysis of the general and specifically mathematical obligations that emerge in the classroom involves viewing the classroom as a community of practice (Wenger, 1998) wherein the teacher and students, as members of the classroom community, co-construct obligations, norms, practices (classroom activity including speech acts), and normative identity.

As sociological constructs, obligations are closely related to the notion of norms, which are also sociologically constructed within the classroom community. Following Searing (1991), Cobb et al. (2009) define a norm as “a recurrent pattern in joint activity that is regulated by the expectations that the teacher and students have for each other’s actions in particular situations” (p. 44). Analysis of emergent classroom norms is thus useful in gathering obligations that develop in the classroom. Furthermore, an understanding of general and specifically mathematical obligations that students are expected to fulfill in order to be successful and effective math-doers in a mathematics classroom makes empirical analysis of the construct of normative identity possible (Cobb & Hodge, 2002). Constructed by *joint* activity and expectations that the teacher *and* students have for each other, classroom norms and obligations are collective constructions. This means that even in a traditional mathematics classroom where the teacher might be viewed as the beholder of knowledge, students must still choose to cooperate with the teacher and identify with the general and specifically mathematical obligations they must fulfill in order to achieve the normative identity of a successful and effective math-doer as constituted in that particular classroom.

Personal Identity. Whereas normative identity is a collective construct, personal identity is individualistic. However, like normative identity, personal identity is a negotiated construct. What’s more, in the process of participating in the “initial constitution and ongoing regeneration of the normative identity…. [students] develop personal identities," wherein they affiliate with the mathematical activity as constituted in the classroom; become disenchanted with mathematical activity during their cooperation and participation in mathematical activity as constituted in the classroom; or oppose participating in mathematical activity as constituted in the classroom (Cobb et al., 2009, p. 47). In this last case, a student will develop a personal identity that reflects their resistance and opposition to the normative identity (Cobb et al., 2009). Through involvement, whether oppositional or cooperative, in behaviors and practices that shape the classroom as a community of practice, personal identities can and do undergo transformation; in that sense, they are collective. In the sense that personal identities belong to individual students and differ from one student to the next, they are individualistic.

Another key point to consider is that specifically mathematical obligations, the construction of competence, and students' self-perception and perception of others play a crucial role in the types of personal identities that students develop in the mathematics classroom. Specifically mathematical obligations have a bearing on what students are held accountable for mathematically. The successful fulfillment of specifically mathematical obligations as constituted in the classroom, determines the extent to which a student views him or herself as mathematically competent as well as the extent to which they may see their peers as mathematically competent. Correspondingly, Cobb et al., (2009) posit that "[w]hat gets constructed as mathematical competence in the classroom has implications for students' perceptions of their own and their peers' relative capabilities and thus for issues of status and power in the classroom" (p. 48). Furthermore, the construct of personal identity refers to the individual identities that students develop within the mathematics classrooms and the extent to which they identify with, merely cooperate, or resist obligations (Boaler & Greeno, 2000;

Martin, 2000). With this in mind, the personal identity that a student develops in a mathematics classroom and the nature of their affiliation with classroom obligations can indicate whether and how students affiliate with what it means to know and do mathematics. Moreover, while normative identity tells us *what* some students are identifying with and/or resisting, personal identity tells us *why* (Cobb et al, 2009).

Methods

I analyzed my data in two parts and, following Cobb et al. (2009), I analyzed each part in two phases. The first part, Normative Data-Scientific Identity Analysis, involved the analysis of the largest data set, video-recorded classroom observations, with the purpose of understanding the nature of the normative data-scientific identity as constituted in the classroom and the ways in which the classroom community contributed to its constitution. The second part, Personal Data-Scientific Identity Analysis, focuses on a second data set, audio-recorded exit interviews conducted with 12 IDS students for the purposes of understanding the nature of their identification with taken-as-shared ways of doing data science.

Normative data-scientific identity analysis.

Phase I. Phase I of the normative data-scientific identity analysis involved (1) organizing my data and (2) identifying emergent classroom obligations as constituted in Ms. Gellar's IDS classroom. My organizational approach is guided by Enyedy & Mukhopadhyay's (2007) qualitative approach to analyzing video case studies, and my analytical approach is guided by Cobb et al.'s (2009) method for deriving and testing general and specifically mathematical obligation conjectures.

The organizational particularities of Phase I involved the creation of time indexed content logs of the video data. Content logs consisted of summaries of video content, instances of partial transcription, and short analytical notes (Enyedy & Mukhopadhyay, 2007). I then conducted thematic open-coding of these content logs using qualitative data analysis software, ATLAS.ti. My unit of analysis for emergent coding was episodes of data science-related activity and social interaction. This process was inductive and recursive, and meant to capture episodes representative of *emergent* themes. Thus, documenting reoccurring themes and then refining into a coding scheme allowed me to account for the thematic nature of data-scientific activity in Ms. Gellar's IDS classroom. Also, emergent codes were not mutually exclusive; if an episode met multiple code definitions and, thus, represent more than one theme, I coded it accordingly. Relatedly, I coded multiple episodes only once if they dealt with the same theme and involved the same students. The number of students involved in an interaction did not affect the amount of times I coded for a theme.

Multiple rounds of coding revealed salient themes indicative of obligations students felt compelled to fulfill in Ms. Gellar's IDS classroom. To construct obligation conjectures, I took a theme represented by a given code and constructed a statement that encapsulated the expectation students feel obliged to meet with regards to that theme. For example, for codes that predominantly spoke to the theme of argumentation, I revisited relevant episodes to consider precisely *what* they conveyed with regards to standards for argumentation. This involved considering if and how episodes testified to the purposes of argumentation, ways of arguing, when to argue, or whom to argue with. In this way, I developed an initial set of general and specifically data-scientific obligation conjectures, which I preliminarily tested by seeing if they held true for all episodes treated as constituting the conjectured obligation.

I subsequently employed Cobb et al.'s (2009) approach to testing and revising conjectures by analyzing content logs of video recorded data and the video recordings themselves to look for instances wherein students' contributions to the data science classroom (1) violated the conjectured obligation and (2) were subsequently treated as illegitimate contributions by members of the classroom community. If the activity that constituted the violation was treated as illegitimate, for instance through a peer's mention of the 'right' way of carrying out a solution method in response to an alternative approach, this indicated that a norm for participation and corresponding obligation(s) were established, observed, and implemented in the classroom. Thus, the confirmation of classroom obligations was dependent on identifying instances where conjectures were refuted and analyzing the reaction of the classroom community in response to the refuting episode. If violation of obligation conjectures were not met with de-legitimating reactions, I revised my conjectures to better represent the obligation conveyed in coded episodes and tested them again (Cobb et al., 2009). What is more, the process of testing obligations was also useful for allowing me to check my biases as the final list of obligations depended on confirmation derived from the data itself. If an obligation conjecture does not withstand the test of confirmation via refutation, then this might signal, for example, that I am misinterpreting the nature of the obligation, the expectation that that obligation is meant to fulfill, or perhaps even observing a non-existent obligation. Thus, revising and re-testing played an important role in ensuring trustworthiness of conjectures.

Additionally, during my preliminary analysis of emergent themes, I found that salient themes spoke to some aspects of the normative data-scientific identity in Ms. Gellar's IDS class but not all; and so, in order to facilitate my understanding of emergent themes and factors that contributed to their constitution, I modified Cobb et al.'s (2009) approach by conducting supplemental coding for episodes that spoke to three areas: authority distribution; ways in which students were able to exercise agency; and how students reasoned with tools and written symbols in the classroom. To clarify how I leveraged my analysis of emergent themes and supplemental coding, I utilized the former as the basis for obligation conjectures, and the latter to supplement my in-depth analysis of classroom obligations and their constitution in Phase II.

My process of analysis was iterative and premised on the constant comparative method of grounded theory wherein I revisited thematic-codes for refinement and continued to develop my analysis of the data (Glaser & Strauss, 1967/1990; Strauss & Corbin, 1990). Thus, I engaged in thematic analysis at the latent level (Braun and Clark, 2006), allowing me to account for not only what was explicitly said, but also how it was said (tone), facial expressions, and body language along with other nonverbal cues. According to Braun and Clarke (2006), engaging in the latent level of thematic analysis transcends analysis of semantic content and enables an examination of "*underlying* ideas, assumptions, and conceptualizations – and ideologies – that are theorized as shaping or informing the semantic content of the data" (Braun and Clarke, 2006, pp. 84, emphasis in original).

Phase II. Once I established an empirically grounded collection of obligation conjectures, I engaged in the second phase of analysis, which involved an in-depth analysis of general and specifically data-scientific obligations. General classroom obligations are imbued with signification regarding the distribution of authority, and the forms of agency students can legitimately exercise in the classroom (i.e. conceptual agency and disciplinary agency) (Cobb et al., 2009). Hence, I distinguish *general classroom obligations* from specifically data-scientific obligations as those that speak to 1) whom students are held accountable to and "the degree to which [they] are given opportunities to be involved in decision making about the interpretation of tasks, the reasonableness of solution methods, and the legitimacy of solutions"; and 2) the ways that students exercise agency in the IDS classroom (Cobb et al., 2009, p. 44). To provide an example of how I might draw meaning from a confirmed obligation conjecture I borrow the following general classroom obligation from one of Cobb et al.'s (2009) case studies:

- "Listening and taking notes in order to understand the solution methods demonstrated by the teacher." (p. 52)

When analyzed, this obligation reveals that students exercised disciplinary agency by engaging in the activity of "listening and taking notes" to understand *specific methods used by the teacher* to solve a given mathematical problem. In this case, Cobb et al.'s (2009) data corpus indicated that students learned to appropriately match up solution methods to given problems but lacked the conceptual understanding of precisely *why* particular solution methods were appropriate for solving particular problems. While students were not discouraged from asking question to grow their conceptual understandings of the relationship between mathematical problems and solution methods, when students asked clarifying questions the teacher responded by repeating the same steps of the solution methods, thus emphasizing the expectation that students were to follow the procedure demonstrated in order to arrive at a solution. By doing so, the sole purpose of identifying an appropriate solution method was communicated as the need to arrive at a solution—to find the answer. By reiterating the procedural steps she had already demonstrated without further elaboration or discussion regarding the reasoning that supported the appropriateness of a specific solution method, the teacher communicated that finding the answer to a problem constituted an effective and, thus, legitimate mathematical practice in the classroom. In this respect, this general classroom obligation indicated that students did not exercise conceptual agency, at least not in regard to understanding the rationale for utilizing established solution methods. They were, however, afforded opportunities to engage in disciplinary agency. Additionally, there was a clear expectation from the teacher that students were to listen and take notes as those who failed to do so were admonished—providing Cobb et al. (2009) insight into *how* this expectation was imparted and by *whom*.

It is important to note that while I draw from Cobb et al.'s (2009) case study to illustrate what a general classroom norm might look like and how it might inform my understanding of authority distribution and ways students were able to legitimately exercise agency in the classroom, my analysis of general and specifically mathematical classroom obligations will, like Cobb et al. (2009), take into account *all* obligations and not just one. In this respect, it is not unlikely that students might exercise different forms of agency in accordance with different aspects of classroom activity. Also, my analysis of general and specifically mathematical obligations takes into account how expectations are conveyed over a period of time. In other words, if I have observed ten classroom lessons and I identify a pattern in joint activity that is recurrent during one lesson, I cannot say that that pattern points to an obligation if it is not recurrent throughout the duration of the course. Figure 3.6 provides a visualization of the interpretive scheme, specifically tailored to an analysis of general classroom obligations.

Personal data-scientific identity analysis.

Phase I. Phase I of the personal data-scientific identity analysis involved reviewing, transcribing, and thematic open-coding of exit interviews conducted with 12 IDS students, similar to the approach I employed in Phase I of the normative data-scientific identity analysis discussed above (Enyedy & Mukhopadhyay, 2007). Unlike the analysis discussed above, my analysis of the personal data-scientific identities that students developed in Ms. Gellar's IDS classroom reserved the development of conjectures for Phase II.

Interpretive Scheme for Analyzing Specifically Data-Scientific Obligations

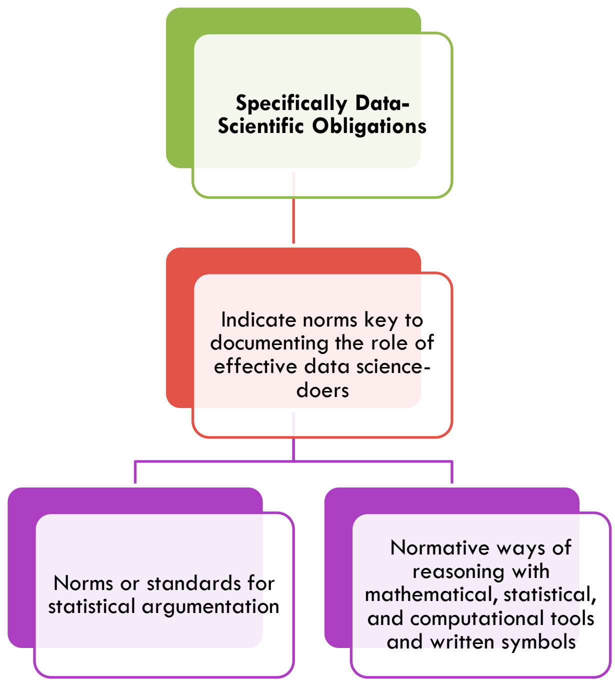


Figure 3.6 The Interpretive Scheme for Specifically Data-Scientific Obligations Model is derived from the model described by Cobb et al. (2009).

Phase II. In the second phase of the personal data-scientific identity analysis I developed, tested, and refined conjectures about the four aspects of students' "views about and appraisals of how the classroom 'works'" (see Figure 3.7) (Cobb et al., 2009, p. 64). According to Cobb et al. (2009) conjectures about understandings and valuations are developed in the process of analyzing thematically related interview segments across students. The nature of refutations in this phase of identity analysis differs from that in the phase of normative data-scientific identity analysis in that here I am not concerned with establishing whether a conjecture speaks to obligations co-constituted by the classroom community. Instead, in this phase of analysis I am concerned with understanding patterns and differences in students' perceptions and affiliations with taken-as-shared ways of doing data science.

Interpretive Scheme for Analyzing Personal Data-Scientific Identities

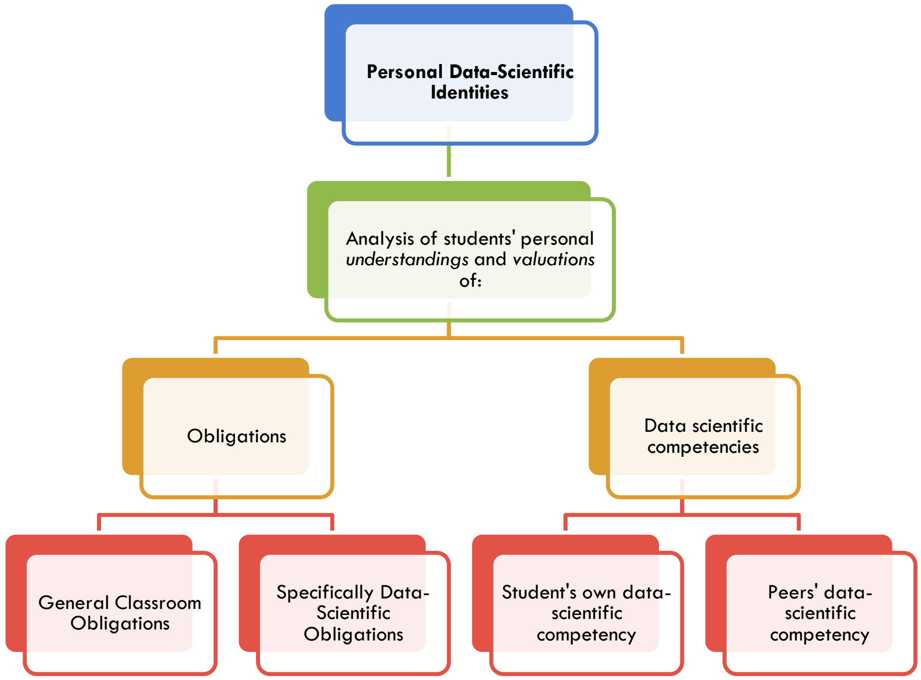


Figure 3.7 I derive this interpretive scheme from that described by Cobb et al. (2009, p. 57).

Thus, the trustworthiness of the analysis of personal identities that students developed in Ms. Gellar's class is established through careful attention to the interplay of conjectures, refutations, interview data, and general and specifically data-scientific obligations. In this phase of analysis, conjectures functioned as analytical tools to help me understand where students' understandings and valuations were consistent, where they differed, and why. Thus, identifying a refutation did not serve to discredit a conjecture; instead, what was of interest was the reason a conjecture was true for some but not for all. This level of analysis allowed me to develop a nuanced understanding of the personal identities that developed among these 12 students and how they related to taken-as-shared understandings of what it meant to legitimately do data science in Ms. Gellar's classroom. In so doing, I was able to account for student perspectives and experiences that may have been obscured in my analysis of the normative data-scientific identity that emerged in this classroom.

Moreover, Cobb et al. (2009) offer that the moral dimension of the classroom helps make the personal identities that students develop in a mathematics classroom tractable. This means that "students' understanding of their general and specifically mathematical obligations involves a sense of 'oughtness' about what they do" (Cobb et al., 2009, p. 47). Accordingly, locating phrases like "you have to," "I decided to," "it [the lab, handout, assignment] tells you to" and other verbal cues that signal a type of moral obligation to meet an expectation were useful in my efforts to understand the personal data-scientific identities that students developed in Ms. Gellar's IDS classroom, and allowed me to gain insight into whether students perceived obligations as obligations-to-others, and thus merely cooperated with the teacher or resisted engaging in data-scientific activity; or perceived obligations as obligations-to-themselves and thus identified with the normative data-scientific identity as constituted in the classroom. While my analysis was not restricted to looking for such phrases to understand the nature of personal data-scientific identities that students developed in Ms. Gellar's IDS classroom, getting a sense of the extent to which students identified with being effective data science-doers was of major importance as identification alone did not mean that data science, as taught and practiced in the classroom, became intrinsically meaningful to a student. It is the *rationale* that supports identification and learning that has implications for whether a non-dominant student is inclined to persist in their STEM trajectory (Cobb et al., 2009).

Citing D'Amato, Cobb & Hodge (2010) define structural rationales for learning as viewing academic achievement as instrumentally valuable "as a means of attaining other ends such as entry to college and high-status careers, or acceptance and approval in the household and other social networks" (p. 185). Further, they define situational rationales for learning as viewing "engagement in classroom activities as a means of maintaining valued relationships with peers and of gaining access to experiences of mastery and accomplishment" (p. 185). While both structural and situational rationales for learning can prove instrumental for a student's desire to perform well academically and pursue continued study in STEM, "students' participation in Discourses that give them access to a structural rationale varies as a consequence of family history, race or ethnic history, class structure, and caste structure within society" (Cobb & Hodge, 2010, p. 185). As such, Cobb & Hodge (2010) posit,

> Discourses that inscribe the achievement ideology wherein society is seen to reward hard work and individual effort with future educational and economic opportunities constitute a resource on which some students but not others can draw as they attempt to make positive sense of their lives in school…the resulting inequities in motivation (Nicholls, 1989) emphasize the importance of ensuring that all students have access to a situational rationale for learning mathematics. (p. 186)

In light of the significance of situational rationales for learning for STEM equity, my analysis of the personal data-scientific identity that students developed in Ms. Gellar's IDS classroom not only helped me grasp the nature of students' identification with taken-as-shared understandings of what it meant to be an effective data science-doer, but also the learning rationales that informed identification.

To provide a brief analytical example that demonstrates how I inferred signification regarding the personal data-scientific identity that students developed in Ms. Gellar's IDS classroom, I draw from an exit interview I conducted with Dolores, a high-performing student in Ms. Gellar's IDS class. When I asked, "Do you think you could be a data scientist if you wanted to?" the following exchange ensued:

***Dolores*: If I really wanted it, probably—just effort. But I don't know—the math…**

Researcher: Why do you say "probably"?

***Dolores*: Anyone can do anything, right? That's my perspective. If you really want it, you can go get it. But I don't want it. It doesn't catch my interest.**

Dolores' response indicates that despite her demonstrated ability to perform well in the class in terms of completing and earning high marks on assignments, being the most vocal female participant in the class, and her status as someone other students turned to for help during lessons and labs, she had no interest in pursuing a career in data science. She also expressed disinterest in mathematics, which the course drew from as a core-statistics course, but did not express negative valuations of her mathematical competence. This I gather from the fact that when she said, "the math" (line 1) her tone expressed apathy for the subject rather than insecurity or a lack of self-confidence. Further, when she speaks a second time in response to my question regarding the uncertain term "probably" (line 2), Dolores expresses confidence in her ability to "do anything" as long as she puts in the necessary effort— "If you really want it," she says, "you can go get it" (lines 3-4). The final two sentences of her response resolutely connect her feelings about effort to issues of interest and aspiration when she says, "But I don't want it. It doesn't catch my interest."

Further, during one of our exchanges in class, she expressed that she was personally invested in doing well in the class because she cared about her grade but not because she found data science intrinsically valuable. Her response, coupled with our exchanges in class, indicate that despite her distaste for math, she did not have negative valuations of her ability to do math effectively. She asserts her viewpoint that exploring data science as a career choice is not so much a matter of mathematical competence but is instead concerned with her interest in math—a perspective presumably informed by her own experiences and sensibilities with regards to math-doing. For Dolores, general and specifically data-scientific obligations represented obligations-to-herself, albeit not due to personal interest or aspirations, but because she wanted to earn a good grade. Therefore, I surmise that Dolores' identification with the normative data-scientific identity as constituted in Ms. Gellar's IDS classroom was supported by a structural rationale for learning data science as opposed to a situational one.

As demonstrated above, while I closely followed Cobb et al.'s (2009) interpretive scheme for analyzing the personal identities that students to developed in Ms. Gellar's IDS classroom by analyzing interviews conducted with students, I also pulled from impromptu exchanges and mini-interviews conducted during video-recorded observations to further contextualize my personal data-scientific identity analysis.

CHAPTER FOUR
Normative Data-Scientific Identity Analysis

In this chapter, I will present my findings as they pertain to obligations students felt compelled to fulfill in their efforts to engage in data-scientific activity in legitimate ways as constituted in the IDS classroom. My forthcoming findings and analysis will show that the normative identity of what it meant to do data science in legitimate ways as constituted in Ms. Gellar's classroom delineated ways of reasoning about disciplinary understandings, solution methods for completing tasks, conditions for peer collaboration, and standards for argumentation. To begin to examine the ways in which this came to constitute what it meant to do data science in legitimate ways, I will begin by discussing Ms. Gellar's role in the classroom to contextualize her approach to teaching IDS. I will then discuss my emergent coding scheme, followed by a discussion of the three general classroom obligations and two specifically data-scientific obligations that I discerned after several rounds of coding, categorizing codes, refining codes, and code definitions consistent with the methods outlined in the previous chapter. I will then delve into an in-depth analysis of the types of classroom activity that constituted the general classroom obligations that emerged in Ms. Gellar's classroom, followed by an analysis of the specifically data-scientific obligations that emerged by discussing the types of classroom and computer lab activity that contributed to their constitution. I will then discuss the role that specifically data-scientific obligations played in defining authority distribution and students' opportunities to exercise agency in the classroom. Ultimately, I argue that by delineating legitimate standards for social participation and disciplinary engagement in the classroom, these types of obligations collectively constituted what it meant to be an effective data science-doer.

Situating Ms. Gellar

As stated earlier in Chapter 3, Ms. Gellar was one of two teachers whose IDS classrooms we, the Energize research team, chose to observe due to the fact that the teachers had a reputation for being especially strong mathematics teachers at their respective high schools. Ms. Gellar was known as an exceptional math teacher with over a decade of experience teaching several traditional mathematics courses. She was very well-respected at MSHS and highly regarded by her students as caring and dedicated to her pupils. In preparation for the first year of implementation of IDS, Ms. Gellar attended several professional development sessions organized by Energize along with nine other teachers that would also be teaching the pilot IDS course. The Energize PDs introduced teachers to the course, and while they allowed them to draw from their pre-existing understandings of mathematics as math teachers, they also introduced them to the concept and burgeoning field of data science. IDS curriculum writers facilitated the PDs and coached teachers on the particularities of teaching IDS at their designated school sites by going through the curriculum, with teachers playing the role of students. This means that teachers went through the entire curriculum including RStudio lab assignments and, necessarily, learned to code in R to enable them to provide assistance to students when they worked through RStudio lab assignments. When students participated in the coding component of IDS, however, the role of teaching and authority as it pertained to determining the reasonableness of solution methods and acceptable student responses was primarily distributed to RStudio lab assignments as opposed to the teacher. This is because RStudio lab assignments were preloaded into RStudio and consisted of procedures and instructions for lab completion that had to be followed closely by students. Thus, unlike the traditional mathematics course and the reform-based course analyzed by Cobb et al. (2009) wherein authority was distributed to the teacher and/or students, in IDS authority was heavily distributed to the RStudio lab assignments, and educational technologies by implication, when students were in the computer lab setting. Situating Ms. Gellar as an Energize-trained teacher is significant for contextualizing her approach to teaching IDS as this was a newly designed course with specific guidelines conveyed during PDs for teacher implementation. After the first year of implementation, Ms. Gellar became an IDS teaching-coach with Energize, co-coaching a new and larger cohort of IDS teachers.

During my time with Ms. Gellar I was able to see firsthand that she was highly dedicated to her students and often sought to reconcile the limitations of IDS as a new course in order to improve her students' learning opportunities and supplement what she perceived as curricular shortcomings. For example, IDS did not have an accompanying textbook which students were often accustomed to having in other courses, particularly mathematics and science courses, and so, Ms. Gellar provided and instituted use of class worksheets for students in the second year of implementation in efforts to help them retain data-scientific disciplinary understandings and reference materials. During our first year together, she also expressed frustration with a lack of formative assessments because she found it difficult to gauge students' grasp of data-scientific concepts and coding skills that were both new to them and new to her. Without a clear understanding of students' disciplinary competency, she felt ill-equipped to fully support their scientific learning. For this reason, Ms. Gellar also designed RStudio lab assignment worksheets to help structure student learning, understand what and how students were learning, and identify where students needed additional scaffolding.

Ms. Gellar's role and responsibilities as a teacher for a STEM reform effort were challenging and highly complex as she was tasked with not only teaching IDS in accordance with PDs, but also preparing students for state-wide standardized assessments, college placement/entrance exams, all the while working toward meeting the goals of her school which emphasized quality instruction over the amount of curricular coverage. In other words, while the curriculum consisted of four units, she was less concerned with ensuring that she taught all four units and more concerned with helping students gain strong understandings of the disciplinary material in ways that could help them meet other competency and proficiency benchmarks. Thus, while she did her best to teach IDS as coached, she also dedicated much time to doing things like designing supplemental work sheets and lessons to meet the needs of her students. I implore that the reader be mindful of the complex nature of Ms. Gellar's role and responsibilities as I discuss the emergence of classroom obligations through classroom practices. It is important to remember, however, that like other classrooms in the district, Ms. Gellar's class was overenrolled with 42 students, and so, her already heavy teacher workload was compounded by her efforts to improve learning opportunities for her students, and, still, she admirably persisted. My role as a researcher and observer in Ms. Gellar's classroom carried its own difficulties and complexities but the ultimate goal of my presence there and my purpose here is to contribute to a collective understanding of STEM reform, to elucidate affordances and limitations therein, and to push forth existing approaches toward equitable outcomes for non-dominant students like the Latino students I had the privilege of meeting and observing in Ms. Gellar's class.

Phase I: Obligations

Phase I of the normative data-scientific identity analysis involved systematically reviewing and cataloguing video-recorded classroom observations to identify what expectations students felt compelled to meet in order to do data science in legitimate ways in the classroom. Below, Table 4.1 relays my emergent coding scheme including the category I assigned to the different types of recurring activity and speech-acts, a description of types of activity and speech-acts ("Definition"), the shorthand code I assigned to each, along with the number of instances and the proportional representation of each activity type. Figure 4.1 indicates the proportional distribution of each code in the classroom setting versus the computer lab setting. It should be noted that codes "TINK" and "LEE" only correspond to labs because they refer to activity that only the labs were conducive to because of their reliance on RStudio and the fact that students only engaged in computer use when in the computer lab. Similarly, I observed two other types of activity coded "IRE" and "CH-T" only during lessons where the teacher's role in the classroom was more focal. I will discuss these forms of activity in the coming sections and in the context of the constitution of general and specifically data-scientific obligations. For the moment, I present Table 4.1 and Figure 4.1 as an organizational tool for conveying to the reader how I began to draw meaning from particular classroom practices. For example, the first four types of classroom activity in Table 4.1 are practices that principally shaped the direction of classroom discourse. I provide the reader with the shorthand code I assigned to each type of practice because I will refer back to them when visualizing and describing the constitution of obligations in coming sections.

Table 4.1 Emergent Coding Scheme

	Definition	Code	N	%
Classroom Discourse	Participating in initiation-response-evaluation method of whole-group classroom discussion by answering questions when prompted to by the teacher	IRE	38	12.7
	The teacher explicitly states a disciplinary expectation that students must meet	EE	40	13.3
	The teacher implicitly conveys an expectation by modeling ways of engaging in classroom activity or producing data science artifacts	IE	12	4
	Students make known that labs state explicit disciplinary expectations for students to meet	LEE	7	2.3
Peer Collaboration	Consulting with peers when directed to by the teacher to complete tasks	CWP	20	6.1
	Engaging in "helping" acts with peers by explaining how to carry out solution methods and engaging in problem-solving together	HELP	23	7.7
	Engaging in "helping" acts to solve problems by copying or doing someone else's work	C	22	7.3
Data Scientific Reasoning	Demonstrating perfunctory and decontextualized treatment of data and data science artifacts	PERF	26	8.7
	Expressing curiosity about aspects of data collection for data they are provided with	CRIT	3	1.0
Argumentation	Indicating disagreement by challenging correctness of peers' data science-related assertions or statistical calculations when solving problems or completing tasks	CH	25	8.3
	Expressing disagreement with Teacher's disciplinary assertions by asking clarifying questions	CH-T	5	1.7
Agency	Making data science-related assertion or response, followed by a phrase that expresses self-dismissal or de-legitimates student's own assertion/response	SDS	14	4.7
Problem-Solving	Following procedures established by the teacher or RStudio labs to complete tasks	PROC	22	7.3
	Solving problems by carrying out a series of calculational steps	CALC	19	6.3
	Appropriately matching solution methods established by the teacher to solve given problems	MATCH	16	5.3
	Experimenting with features of RStudio to complete the labs when directed procedures for lab completion prove ineffective	TINK	8	2.7

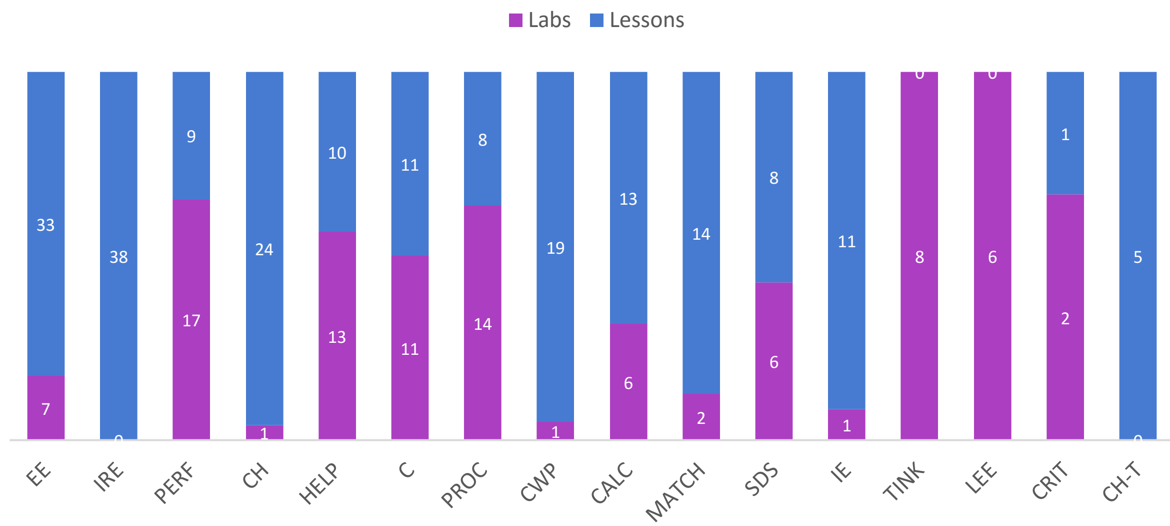


Figure 4.1 Proportional distribution of emergent codes for lessons and labs. NOTE: each column represents 100% for each code indicated below it.

Phase II: In-Depth Analysis of Obligations

General classroom obligations. In this section, I will discuss the general classroom obligations that I derived from the codes presented above including a discussion of the particular classroom practices that contributed to their constitution. Constituted by student and teacher engagement in a number of classroom practices aimed at meeting expectations of what it meant to legitimately do data science and engage in data-scientific activity, the general classroom obligations that emerged in the IDS classroom include:

1. Developing disciplinary understandings and generating disciplinary reference material by listening and taking notes
2. Completing tasks by carrying out established procedures
3. Collaborating with peers for help completing tasks

General classroom obligation #1. The first general classroom obligation, listening and taking notes to gain familiarity with disciplinary skills and concepts and generate disciplinary reference material was directly supported by three types of student and teacher activity (Figure 4.2).

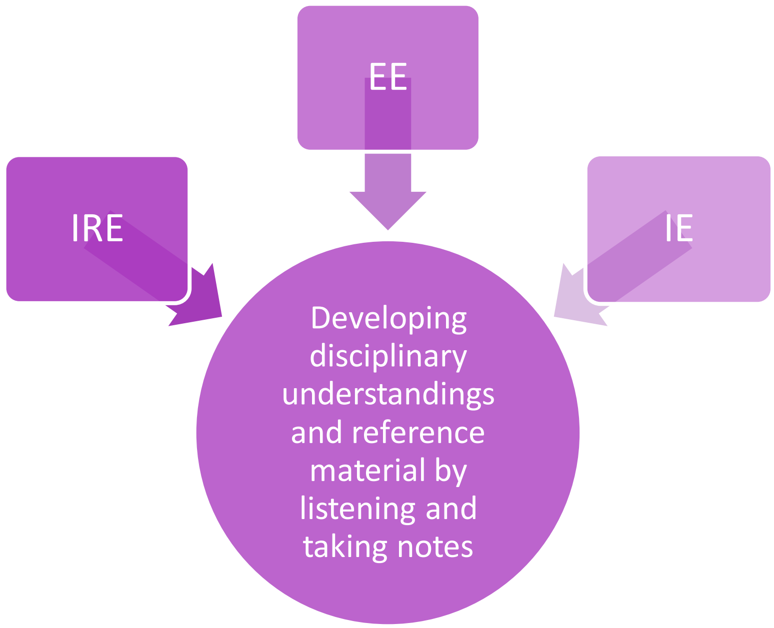


Figure 4.2 Constitution of general classroom obligation #1.

The first type of classroom activity that directly contributed to the constitution of this obligation was participation in the initiation-response-evaluation ("IRE") method of whole-class discourse. When Ms. Gellar introduced students to data-scientific skills and concepts, she initiated discussions by posing questions to which students were expected to respond. Ms. Gellar subsequently evaluated and either legitimated student responses by affirming their correctness or de-legitimated them by re-phrasing students' responses; asking probing questions to guide them toward the appropriate response; or leaving questions open for others to answer appropriately in accordance with the curriculum. Students also contributed to classroom discussion by responding to Ms. Gellar's prompts throughout the duration of the course and showing attentiveness to Ms. Gellar's and students' assertions.

The second type of classroom activity that contributed to the constitution of the first general classroom obligation was the teacher's delivery of statements that conveyed explicit disciplinary expectations. This activity was facilitated by Ms. Gellar's use of lesson slide presentations that stated guiding lesson items such as enduring understandings, lesson objectives, vocabulary words, and essential concepts drawn from the written IDS curriculum. Slide presentations informed students about *what* they were expected to learn and *how* they were expected to perform data-scientific activities. Examples of explicit expectations conveyed through slide presentations include:

- "Students will understand that the mean of the absolute deviations (MAD) is a way to assess the degree of variation in the data from the mean and adjusts for differences in the number of points in the data set."
- "Students will understand the basic rules of probability. They will learn that previous outcomes do not give information about future outcomes if the events are independent."
- "Students will learn how to merge two data sets and ask statistical questions about the merged data."

Ms. Gellar also verbally stated explicit expectations regarding acceptable responses and ways of engaging in data-scientific activities. Examples of these types of explicit expectations include:

- "If you are done, your explanation shouldn't be, 'this needs to be this type of plot.' Explain to me why. I don't want a one-liner."
- "If you give me things that are not appropriate, then that tells me you don't really know what you're supposed to be analyzing. You need to look at the distribution and determine which [measures of center or spread] are more appropriate."

Furthermore, Ms. Gellar also expressed the explicit expectation that students should take notes by copying projected lesson items such as enduring understandings, lesson objectives, vocabulary words, and essential concepts. She regularly stopped at slides that contained these lesson items and asked students to copy them down in the lesson worksheets she provided. Copying lesson items was meant to provide students with referential material that they could revisit throughout the course in the absence of an accompanying course textbook. These types of expectations were conveyed by direct requests for students to copy projected lesson items onto their class worksheet.

Additionally, while Ms. Gellar explicitly stated disciplinary expectation during lessons, the lab slides loaded into RStudio also expressed disciplinary expectations that helped define legitimate ways of doing data science. Due to the fact that students were responsible for completing lab assignments by following steps specified within pre-loaded RStudio lab slides, the teacher did not typically express expectations during labs, other than the expectation that students follow steps outlined in RStudio lab assignments. Hence, most of the episodes I observed of explicit expectations by the teacher ("EE") occurred in the classroom setting during lessons. This also explains why all initiation-response-evaluation ("IRE") episodes occurred in the classroom setting during lessons and not in the computer lab setting where RStudio lab slides functioned as the guiding source of classroom instruction.

The third type of classroom activity that contributed to the constitution of the first general classroom obligation was implicit expectations conveyed via modeling of acceptable written responses and appropriate methods for solving problems. I did not observe as many instances of implied expectations as I did explicit expectations because expectations were mainly explicitly stated in the classroom setting (verbally or via projected lesson slides) or stated within RStudio lab slides in the computer lab setting. The following is an example of an implied disciplinary expectation via modeling:

- “We’ve been setting [comparison statements] up like ‘This percentage of this group is taller or shorter than this percentage of that group’…so far we’ve looked at our shortest male and our tallest female. What else?”

Students demonstrated that they were receptive of Ms. Gellar’s explicit and implicit expectations by engaging in activity and delivering responses that reflected efforts to meet those expectations. For example, Figure 4.3 below provides examples of student responses that sought to meet Ms. Gellar’s modeling of an acceptable response. The quotes below are drawn from a lesson on chance outcomes and shuffling simulations where Ms. Gellar had just declared the purpose of the day’s activity to determine if differences between groups were due to chance or by design. She then asked students to compose five statements that compared two groups, and as an example offered a sample statement as depicted in Figure 4.3.

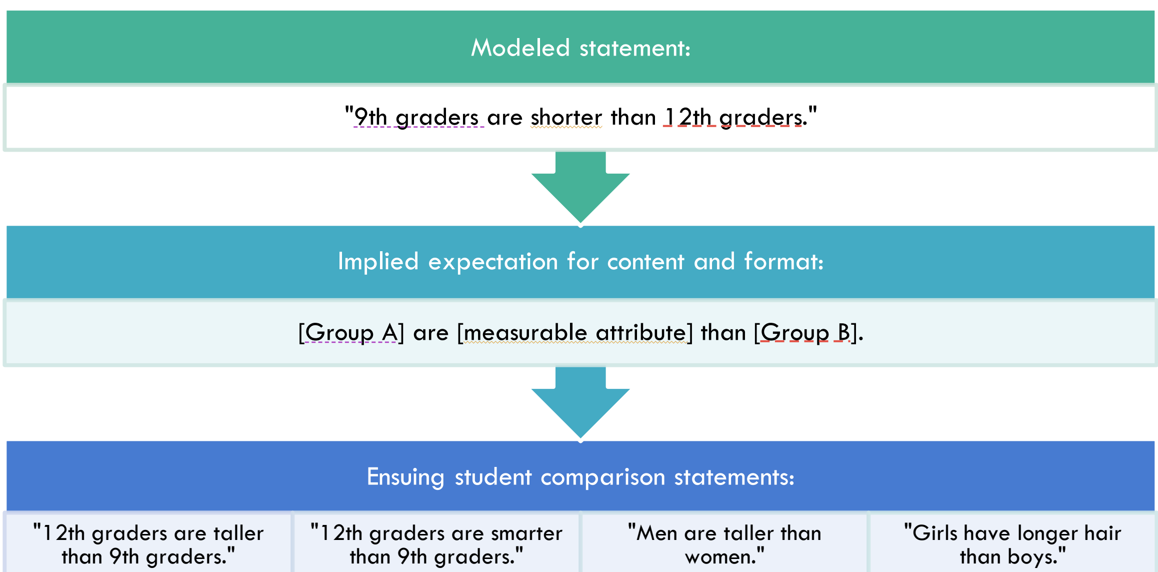


Figure 4.3 Examples of student responses that sought to meet modeled response standards.

Taken together, Ms. Gellar’s approach to guiding classroom discourse and conveying disciplinary expectations, along with students’ participation in classroom discourse and adherence to disciplinary expectations contributed to the constitution of the first general classroom obligation of listening and taking notes to gain familiarity with disciplinary skills and concepts and generate disciplinary reference material. Furthermore, while this obligation was primarily constituted by student and teacher activity during lessons, lab activity also contributed to its constitution by predominantly emphasizing lab assignment completion contingent on students' willingness and ability to follow procedures stated in RStudio lab slides. Labs were designed to introduce students to coding and, thus, could only be completed through student engagement in coding. Due to the fact that the overwhelming majority of students in Ms. Gellar's class had no prior experience with coding, their development of coding skills was completely dependent on following procedures outlined in lab slides. Ms. Gellar encouraged students to use their coding history as a reference tool to recall codes; and so, students were actively creating data-scientific reference material by virtue of working through lab assignments. Also, because following lab procedures was the only way to complete labs assignments, students typically did so, though not without objection.

General classroom obligation #2. The second general classroom obligation, completing tasks by carrying out established procedures, was directly constituted by the first general classroom obligation and four additional types of student activity concerned with problem-solving (Figure 4.4). Below, I will discuss how the first general classroom obligation, and episodes of classroom activity coded "PROC", "CALC", and "MATCH" were elemental in the makeup of the second general classroom obligation.

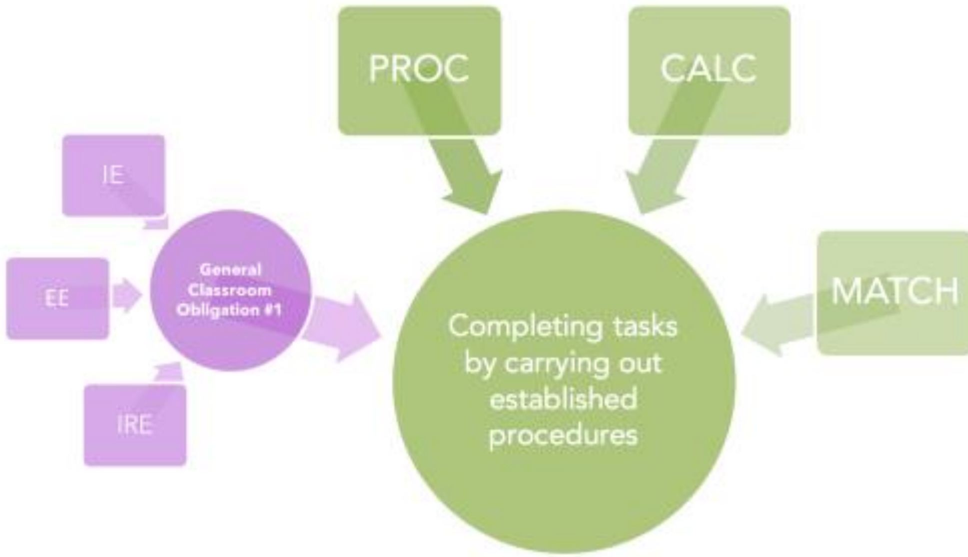


Figure 4.4 Constitution of general classroom obligation #2.

The first general classroom obligation was instrumental in demarcating authoritatively differential responsibilities for students and for the teacher. It positioned data-scientific knowledge as established, static, and true, but never as an evolving element of a new disciplinary field. In this way, the second general classroom obligation supported students' exercising of disciplinary agency wherein they were responsible for understanding imparted disciplinary understandings and applying them appropriately to complete tasks. Thus, for students to be academically successful in the classroom they had to gain familiarity with the skills and concepts necessary to execute established methods for solving problems and completing tasks. By engaging in activities that sought to meet the first classroom obligation of developing disciplinary understandings and reference material by listening and taking notes, students focused on how to do data science, as endorsed by the teacher and the written curriculum.

As a result of students' cooperation with disciplinary expectations inherent in the curriculum and their subsequent participation in the co-construction of classroom obligations, students completed tasks by deferring to procedures prescribed by either Ms. Gellar or the

RStudio lab slides. All three types of classroom activity that constituted this obligation, not including those that constitute the first general classroom obligation, represent dominant problem-solving approaches to task completion and include (1) following procedural steps ("PROC"), (2) performing established calculational steps ("CALC"), and (3) matching solution methods to given problems ("MATCH"). I categorized these types of activities as problem-solving episodes (Table 4.1). Moreover, problem-solving episodes accounted for 16 percent of all coded classroom activity, while the three specific approaches just mentioned accounted for 87 percent of all coded activity that specifically related to problem-solving approaches (see Figure 4.5).

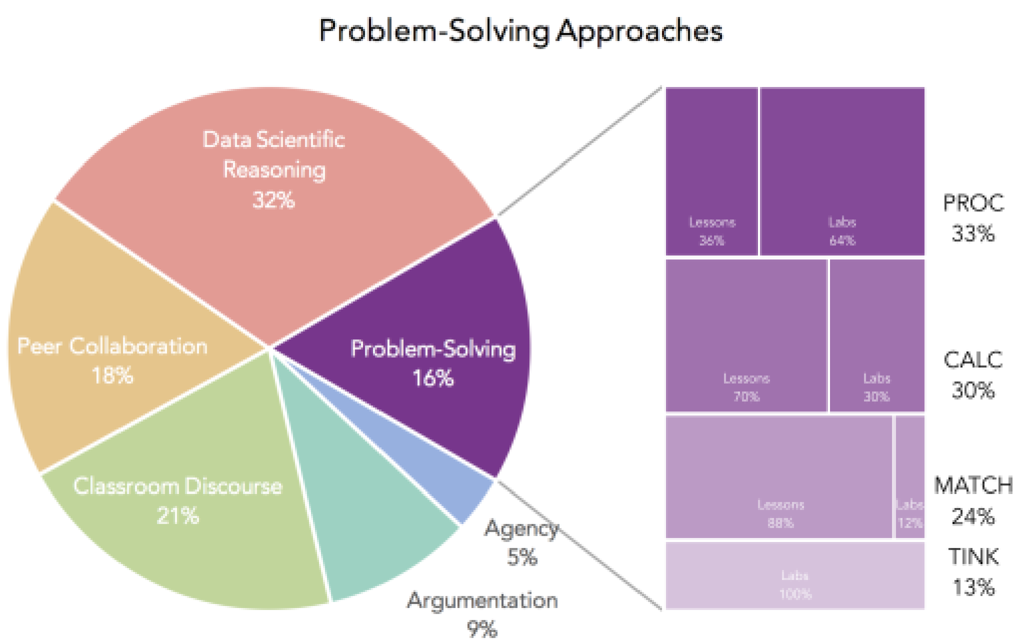


Figure 4.5

While these approaches are similar, there are important distinctions to discuss. "PROC" refers specifically to forms of procedural task completion that do not include carrying out mathematical calculations. "CALC," on the other hand, refers specifically to solving mathematical problems by following calculational steps described by the teacher. While IDS drew heavily from the applied mathematics field of statistics, not all tasks required students to solve mathematical problems by carrying out calculations. For instance, students completed tasks such as lab assignments by following procedures stated within RStudio lab slides and not by carrying out calculations. I attribute this to the fact that lessons were designed to introduce students to statistical skills and concepts used to analyze and interpret data; therefore, there was a greater focus on calculating, for instance, measures of center and spread to describe a data set. Labs, on the other hand, were designed to help students develop technical coding skills by following lab slide procedures to write and run codes. By running codes, students directed RStudio to carry out operations including statistical calculations. The differential emphases of lessons and labs helps contextualize why I observed more cases of procedural task completion in the computer lab setting (14) than I did in the classroom setting (8). Conversely, I observed more cases of students solving statistical problems by following calculational steps when they were in the classroom setting (14) as opposed to the computer lab setting (6). Figure 4.6 provides an overview of the categorical distribution of all coded activity to the left. To the right is a breakdown of the problem-solving category including the proportional distribution of codes within it and the inter-code distribution of classroom activity for lessons (classroom setting) and labs (computer lab setting).

The final type of activity that contributed to the constitution of this obligation was appropriately matching solution methods established by the teacher to solve given problems. Unlike “PROC” and “CALC” which are concerned with the *how* of problem-solving, “MATCH” is concerned with the *what*. Student cooperation in the fulfillment of the first general classroom obligation contributed to the demarcation of decision-making guidelines for determining *what* statistical measures to calculate (i.e., measures of center: mean vs. median; measures of spread:

mean of absolute deviation [MAD] vs. interquartile range [IQR]) and *what* graphical representations to generate (i.e., histogram vs. boxplot). For example, as per Ms. Gellar's explicit expectations, students were encouraged to match the mean, as a measure of center, to data sets with distributions of symmetrical/unimodal shape. Furthermore, they were encouraged to match the MAD, as a measure of spread, with the mean. In essence, if a student decided to calculate the mean for a distribution, then they also calculated the MAD; conversely, if students decided to calculate the median for a distribution, then they also calculated the IQR. Students became adept at matching calculational solution methods with statistical problems, but appropriately matching graphical representations with distributions proved more complex because in addition to having more options to choose from for graphical representations, students also had to evaluate two kinds of characteristics: the *shape* of the distribution and the *size* of the dataset.

General classroom obligation #3. The third general classroom obligation, collaborating with peers for help completing tasks, was directly constituted by four types of classroom activity that defined the function of peer collaboration. Figure 4.6 below provides a visualization of the constitution of the third general classroom obligation.

The first type of activity that contributed to the constitution of the third general classroom obligation involves students consulting amongst themselves when directed to by the teacher to complete tasks. While this activity emerged as a direct result of Ms. Gellar's explicit expectation for students to work collaboratively, explicit expectations discussed earlier refer to those that delineated *how* students were expected solve problems and complete tasks. The activity of consulting with peers, coded as "CWP," refers to activity that resulted from explicit expectations that specifically demarcated the *function* of peer collaboration and *conditions* that called for it.

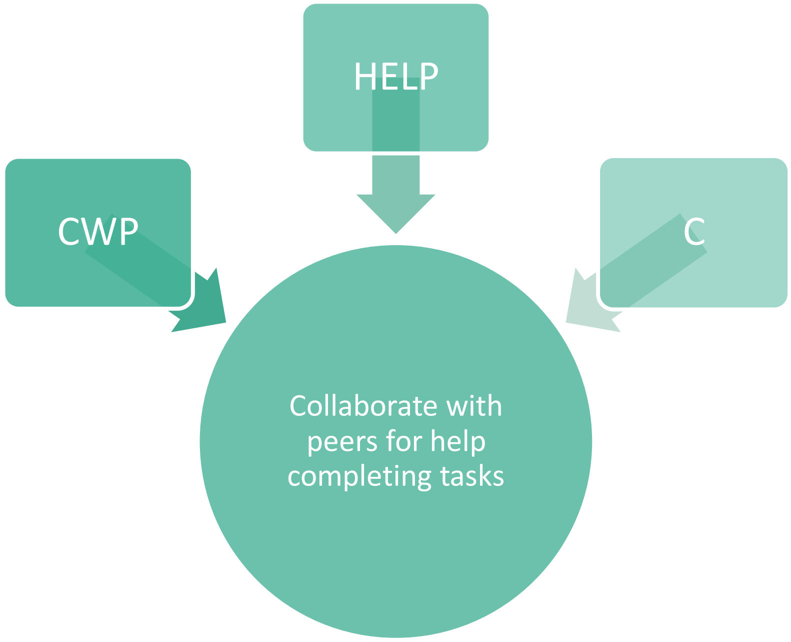


Figure 4.6 Constitution of general classroom obligation #3.

For example, at specific moments during lessons, students were asked to collaborate with each other for the purpose of

- Facilitating understandings of data-scientific skills and concepts;
- Completing tasks in table-groups; and
- Brainstorming ideas and responses.

Out of all observed episodes of students collaborating with each other, 45% accounted for student collaboration for the purpose of facilitating data-scientific understandings; 35% accounted for student collaboration for the purpose of competing tasks in table-groups; and 20% accounted for student collaboration for the purpose of brainstorming ideas and responses.

The directed activity of consulting with peers to complete tasks helped establish a pattern wherein students collaborated with each other for the purposes of task completion without Ms. Gellar's direction. The final two types of classroom activity that constituted the third general classroom obligation, collaborating with peers for help completing tasks, represent two distinct interpretations of "help." For instance, episodes coded "HELP" refer to students engaging in helping-acts with peers by explaining how to carry out solution methods and carrying out problem-solving methods together. This approach assumes an interpretation of helping as involving collaborative discussion and reasoning about what solution-methods to employ and how to carry them out. Episodes coded "C" refer to students engaging in helping-acts with peers by either copying each other's work/answers or doing each other's work. This second approach reflects an interpretation of helping as consisting of giving and acquiring answers and outputs to complete tasks. The following excerpt provides an example of these conflicting interpretations and intimates that the difference between the two is founded on students' understandings of the relationship between (1) disciplinary skills and concepts and (2) tasks. The exchange takes place as all students are working on their warm-up, written on the whiteboard as pictured in Figure 4.7:

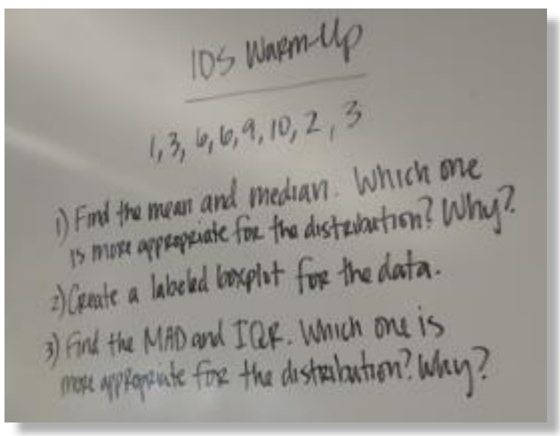


Figure 4.7 Students were asked to complete the warm-up pictured above at the start of class and with their group-mates.

At the opening of the exchange, Mervin disagrees with Antoine's calculation of the numerator portion of the MAD calculation they were introduced to the week prior; the conversation evolves thereon.

LINE 1	**MERVIN**	**You were right—[*whispers to himself*] 20, eight, then this—[*turns to Antoine*] eight divided by 20.**
2	ANTOINE	No, it's 22.
3	**MERVIN**	**22[?]—I got 20 [after adding—**
4	KAREN	Yes, I got 20 too, that's why I was [inaudible]—

5	**ANDRES**	**How'd you guys get 20, though?**
6	MERVIN:	I got 20 [*insistent*]. Where'd you get that [22] from? Four, three, four, two, four, five.
7	**KAREN:**	**[To Mervin] You know, everybody else got 22—**
8	ANTOINE:	Yes, that's why. I got 22.
9	**MERVIN:**	**[*Responds with incredulity*] I didn't get that. [*Looks over his worksheet and begins to read***
10		***out his calculation*] Look, okay, seven—no, wait…six…**
11	ANTOINE:	[*Nods, indicating that Mervin just proved himself wrong and Antoine right*] 22.
12	**MERVIN:**	**Oh…my bad.**
13	ANTOINE:	Because you'll get 2.7—
14	**KAREN:**	**Just put 22. Everybody has 22 [laughs]**
15	MERVIN:	Seven, six…[*recalculating*] I got 20 [again].
16	**ANTOINE:**	**No, it's 22.**
17	MERVIN:	[*Reviews his calculation again*] …oh sh** [*finds and erases his mistake*] …
18	**KAREN:**	**[To Antoine] After I do 22 divided by eight, right?**
19	MERVIN:	[*Passively as he erases*] Yes, it should be…
20	**ANTOINE:**	**2.7.**
21	MERVIN:	2.7.
22	**KAREN:**	**2.7 [*writes on her worksheet*]? Right? [*Antoine nods*]**

At the beginning of the conversation, Karen tells Antoine that, like Mervin, she, too, calculated 20 instead of 22 (line 4) halfway through her MAD calculation. Soon after, in lines 7 and 14, she encourages Mervin to "Just put 22" because "everybody else got 22." Mervin, however, is not convinced that Antoine's calculation is correct and does not seem concerned with Karen's reference to everyone else's calculation. He is more concerned with understanding the basis for the discrepancy between his and Antoine's response, and later with understanding how to properly execute the appropriate solution methods to complete the warm-up. Antoine works collaboratively with Mervin by checking his calculation and eventually telling him exactly how to construct at boxplot with the necessary components.

This exchange suggests that Karen's interpretation of "help" can be likened to copying precisely because she interpreted peer-collaboration as a means to an end: acquiring answers necessary to complete a task. Thus, for Karen, the purpose of learning and actualizing disciplinary skills and concepts was to facilitate her ability to find answers to completing tasks. However, once aware of the answers (i.e., 22 [line 2] and 2.7 [lines 20-22]), the value of reasoning through the task by reflecting on disciplinary understandings was diminished because the perceived ultimate end-goal was achieved. On the other hand, despite Karen's suggestion that

Mervin accept Antoine's answer and Antoine's insistence on the correctness of his calculations, Mervin's concerns remained insatiate. He was less concerned with finding the correct answers and more concerned with understanding the reasoning for Antoine's contestation of Mervin's response. Mervin's interpretation of "help" can be likened to exploring and explaining discrepancies in peer responses precisely because he interpreted peer-collaboration as a means of clarifying and supporting disciplinary understandings. Thus, for Mervin, processes involved in completing this task presented opportunities for him to reason with disciplinary skills and concepts both in carrying out calculations and in evaluating the viability of calculated outcomes. Unlike Karen, knowing the answers did not diminish his valuation of disciplinary understandings because he had not yet fulfilled the purpose of the task as an exercise in data-scientific reasoning. In the following section, I will begin to delve into the specifically data-scientific obligations that emerged through student and teacher participation in the classroom community.

Specifically Data-Scientific Obligations. Constituted by student and teacher participation in discipline specific activity, the IDS classroom emerged two specifically data-scientific obligations as follows:

1. Justifying data-scientific assertions when challenged by peers by referencing established problem-solving procedures, reference materials, or the teacher's data-scientific assertions
2. Producing data-scientific outputs by following stated computational steps

Specifically data-scientific obligation #1. The first specifically data-scientific obligation of justifying disciplinary assertions when challenged by peers by referencing established problem-solving procedures, reference materials, or the teacher's data-scientific assertions helped outline ways of engaging in data-scientific argumentation. Figure 4.8 below provides a visualization of the constitution of the first specifically data-scientific obligation.

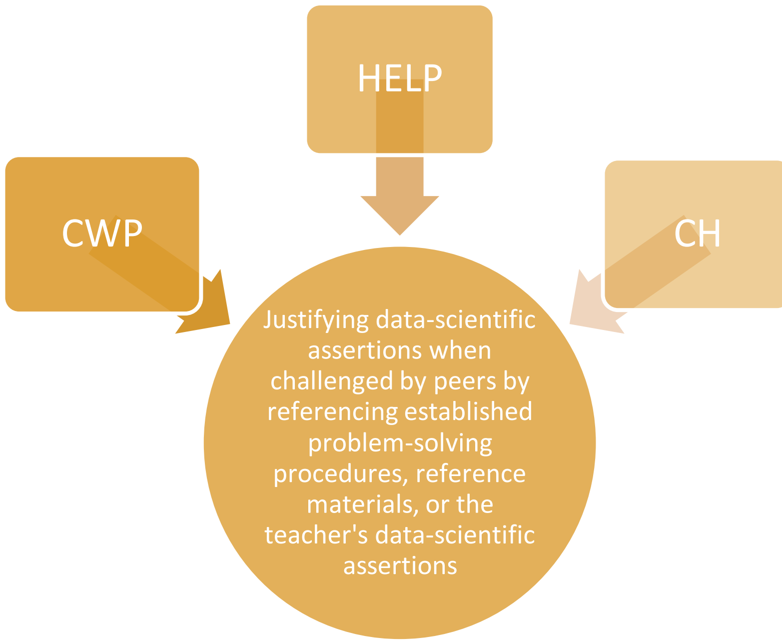


Figure 4.8 Constitution of specifically data-scientific obligation #1.

The first type of activity that contributed to the constitution of the first specifically data-scientific obligation was consulting with peers when directed to by the teacher to complete tasks ("CWP"). By delineating the purpose of peer-collaboration as concerned with facilitating students' abilities to complete tasks, directives to consult with peers helped establish a pattern wherein students turned to each other for help or to compare responses without being asked to by the teacher. While some students focused their attention on completing tasks by asking and giving peers the answers (see discussion of activity coded "C" in section "General classroom obligation #3" above), others sought to understand whether there were discrepancies between their responses and that of their peers and, importantly, why ("HELP"). As students explained solutions, solution methods, or reasons for selecting a particular solution method, their reasoning found basis in established disciplinary understandings and reference materials that captured them. Therefore, while the activity of consulting with peers to complete tasks created opportunities for students to copy answers under the guise of "helping" each other, it also provided opportunities for students to reason about data-scientific understandings as they worked to resolve discrepant solutions.

Inevitably, consulting with peers to complete tasks brought forth the third directly constituting classroom activity of challenging-acts, which I coded as "CH." Challenging-acts refers to episodes of student collaboration, both teacher-directed and student-initiated, where students challenged the legitimacy of other students' disciplinary understandings and skills. Accordingly, episodes where students challenged each other's mathematical calculations constitute 44% of all episodes coded as challenging-acts; episodes where students challenged each other's data-scientific assertions/responses make up 32% of all observed challenging acts; and episodes where students challenged each other's basis for the selection of solution-methods/ways of executing solution methods accounted for 24% of all coded challenging-acts. Additionally, these types of challenging-acts predominantly took place during lessons when students were in the classroom setting. This makes sense given that, as stated earlier during my discussion of the first general classroom obligation, students most often carried out calculations and collaborated with peers to complete tasks during lessons, and used RStudio as a tool to carry out commands including calculations when in the computer lab setting. As such, of all the episodes I classified as challenging-acts, 96% occurred in the classroom setting and 4% occurred in the computer lab.

Another significant finding regarding challenging-acts is that students challenged each other ("CH") five times more often than they challenged Ms. Gellar (I coded these challenging-acts distinctly as "CH-T"). When engaging in argumentation, students cited established disciplinary understandings and reference materials as the justifying basis for their calculations, disciplinary assertions, and selection and execution of solution methods, yet the basis for their own data-scientific understandings ultimately traced back to disciplinary assertions conveyed by Ms. Gellar. This means that Ms. Gellar's contributions to the classroom community's data-scientific understandings were perceived as legitimate and, thus, less likely to be challenged by students who regarded their teacher as the guiding authority figure in the classroom.

Specifically data-scientific obligation #2. The second specifically data-scientific obligation, producing data-scientific outputs by following stated computational steps delineated ways of producing data-scientific outputs. This disciplinary obligation was constituted by classroom activity that emerged from lines of authority demarcated by the first general classroom obligation; problem-solving standards established by the second general classroom obligation; and patterns in and purposes of peer collaboration delineated by the third general classroom obligation. Figure 4.9 below provides a visualization of the constitution of the second specifically data-scientific obligation.

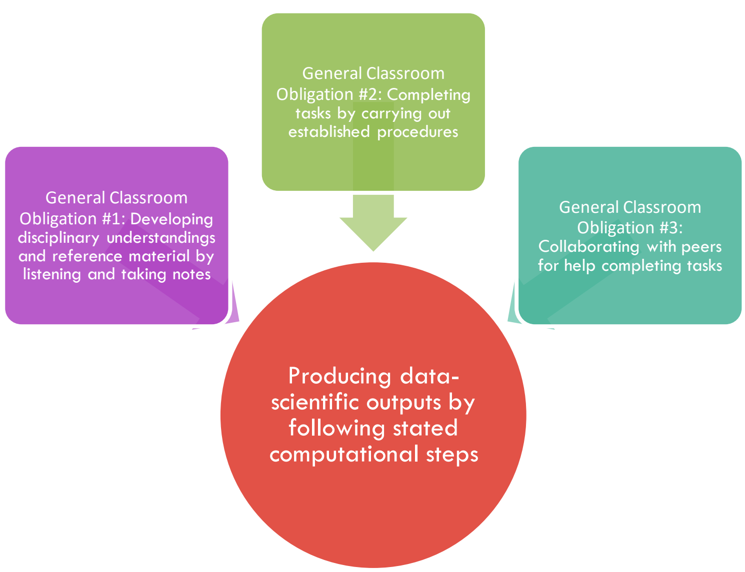


Figure 4.9 Constitution of specifically data-scientific obligation #2.

The first general classroom obligation, which formed the foundation of authority distribution in the classroom and positioned students as learners to be taught and the teacher/RStudio lab slides as the knowers responsible for teaching. Students engaged in task completion activities in ways that conformed to acceptable standards for science-doing in accordance with expectations conveyed by Ms. Gellar, the written curriculum, and the RStudio lab slides. In the computer lab setting students were heavily dependent on computational steps outlined in RStudio lab slides to complete assignments because completion required students to generate data-scientific outputs—including graphical representations, calculations, and probability simulations—by operationalizing coding skills developed concurrently as they worked on lab assignments. Because students had an incipient understanding of how to code, following procedures outlined in labs was necessary in lieu of alternative means and methods. Indeed, the majority, 64%, of procedural task completion took place in the computer lab setting.

There was one type of problem-solving activity that took place only in labs (see Figure 4.5). This activity, coded “TINK,” refers to experimenting, or tinkering, with features of RStudio to complete lab assignments when established procedures proved ineffective. Tinkering with RStudio accounted for a small portion, 13%, of all problem-solving approaches. While tinkering allowed students to experiment with RStudio, explore its features, and pursue lines of inquiry not addressed in lab assignments, the students who engaged in this activity only did so because of the following issues:

1. RStudio glitches that interrupted functionality, preventing students from working through lab assignments (two observed instances);
2. Student attempts to circumvent outlined procedures and streamline completion of lab assignments (two observed instances); and

3. Student coding errors (three observed instances) that failed to generate desired outputs.

Although the classroom community did not enact these approaches regularly, an analysis of student tinkering helps inform an understanding of how some students reconciled procedural task completion with technological problems and student errors. Additionally, the same types of issues that led some students to tinker with RStudio in experimental and exploratory ways led others to consult with peers for help completing lab assignments. Consulting with peers to complete lab assignments led to student engagement in helping-acts and copying just as it did in the classroom setting (see General Classroom Obligation #3). However, while 50% of all observed episodes of copying occurred in labs, a greater percentage of helping-acts (57%) involved discussing how to carry out procedures and reasoning about problem-solving collaboratively among peers.

An analysis of this type of activity also suggests that limitations posed by new forms of data-generating internet-enabled technology in the classroom temporarily disturbed established procedures for task completion by interfering with the effectiveness of computational steps. For example, when RStudio sessions began to glitch, lab procedures proved futile. Also, students were not immediately aware that their computers were malfunctioning because their limited experience with coding and RStudio led them to first assume that unanticipated error messages were due to typing the wrong command rather than due to technological issues. This initial assumption of student error coupled with the procedural rhythm of students' computer lab activity further moved students to seek help from others to address the issues preventing them from meeting specifically data-scientific expectations.

Now that I have presented and examined the general and specifically data-scientific obligations that students felt compelled to fulfill in the IDS classroom, I will discuss how these obligations informed the distribution of authority in the classroom and its implications for students' opportunities to exercise disciplinary and conceptual agency.

Authority distribution

The first general classroom obligation of developing disciplinary understandings and generating disciplinary reference materials by listening and taking notes speaks primarily to how authority was distributed in the classroom. Analysis of this obligation and constitutive activities indicates that authority was primarily distributed to Ms. Gellar. Her approach to guiding whole-group classroom discourse indicates that while she invited students to exercise agency by posing questions, the questions were usually very simplistic, arguably rhetorical, and served to check that students were learning what they were expected to learn rather than to engage students in co-constructing data-scientific knowledge or cultivating in-depth understandings of the principles that underpin data-scientific tools and solution methods. For example, when the class was engaging in the human boxplot activity, where students were asked to think of their collective heights as one data set and create a visual representation of their height distribution in the form of a boxplot, Ms. Gellar was responsible for introducing new concepts, directing the activity, and guiding whole-group discussion. Students' participation, on the other hand, consisted of following directions for carrying out the activity, copying lesson items when prompted to, and answering the teacher's questions. Below I offer an exemplar that illustrates how all members of the classroom community engaged in the IRE approach to classroom discourse and contributed to the co-constitution of authority in the IDS classroom. The exchange took place after Ms. Gellar walked students through the process of identifying the student that represented the median height in the class. Once Mervin was identified as having the median height, Ms. Gellar asked students to similarly identify the midpoint of each one of the groups to his left (taller-height group) and right (shorter-height group). While the all-female shorter-height group successfully identified the person that represented their midpoint, a disagreement emerged within the mostly-male taller-height group regarding how many people made-up their sub-group and, thus, who the middle person was. The excerpt indicates how these types of whole-group interactions written into the curriculum positioned the teacher as responsible for imparting data-scientific knowledge unto students and positioned students as responsible for learning the knowledge imparted unto them.

MS. GELLAR: By the way, how many people are in each half [to the left and right of Mervin] if we're not counting Mervin?
ANDRES: 19.
MS. GELLAR: 19, so if…we're trying to find the middle, how many people in is that? [*Students mumble faint responses*] So 19, so after the 9th person, that's going to be your middle. So… [*starts counting at one end of the all-female shorter-height group to arrive at the 'median' person*]

FEMALE STUDENTS: Christine.
MS. GELLAR: Christine, yes. [*starts counting at the other end of the mostly-male taller-height group to arrive at the median*] So, who's behind Kim? Giselle? [*students nod*] Okay. So, Giselle's the middle of this half. Okay, so, we first split [the class] in half and then we split each half in half, so how many pieces do we have?
STUDENTS: Four.
MS. GELLAR: Four, so each fourth is referred to as a quartile. Each fourth is referred to as a quartile, okay? So, I need Christine—I need you so I can measure your height.
[*Christine approaches the front of the class and Ms. Gellar marks her height on the poster paper*].
And I need Giselle…
[*Giselle approaches the board and Ms. Gellar marks her height on the poster paper*]
Okay, alright, so, what did we call each quarter? What was it called?
STUDENTS: Quartile.
DIEGO: Quartile.
MS. GELLAR: Quartile, okay. So, we call it a quartile because it is a fourth—quarter, *quatro*—right? Quartile.
ANDRES: Miss, do we write that [the contents of the projected "Quartiles" slide onto the classwork sheet]?
MS. GELLAR: Not quite…I'm going to give you a definition in just a second. This is quartile three [*labels appropriate height mark "Quartile 3 (Q3)"*], **or we can abbreviate as Q3, okay? This is quartile one** [*labels "Quartile 1 (Q1)"*], **we can abbreviate this as Q1. And this** [*points to "Median" already written next to Mervin's height mark*] **really, what would this be?**

JESUS: Quartile two?
MS. GELLAR: [*Nods*] **Quartile two. So, what percentage of people fall below quartile one?**
ANTOINE: Them [points to female students standing at the end of the lower-height group lineup].
MS. GELLAR: What percentage?
STUDENTS: 25.
MS. GELLAR: 25% of the people will fall below quartile one [*points to Q1 label*]. **What about the median? What percentage of people will fall below the median?**
ARMANDO: 50.
MS. GELLAR: 50. What percentage will fall below quartile three?
STUDENTS: 75.
MS. GELLAR: 75. They're quarters, right? So, 25% of you are shorter than this [*points to Q3*]. **50% of you**

are shorter than this [*points to median*]. 75% of you are—

DIEGO: Taller—

MS. GELLAR: Shorter than this…what percent fall between quartile three and quartile one?

STUDENTS: 50.

MS. GELLAR: 50, so 50% of you, half of the class, fall between this height [*points to Q3*] and this height [*points to Q1*]. 50% of you fall between quartile three and quartile one. So, if I want to describe the typical height, I can say—I don't know what height this is, but let's say that's 5'3", right? And Giselle, how tall are you?

GISELLE: 5'7".

MS. GELLAR: 5'7". So I can say, "Well, 50% of the class is between 5'3" and 5'7". I can describe your heights that way, okay? So, the next part of your classwork is a description of those quartiles. Okay, so go ahead and take a second, make note.

[Students copy description of quartiles projected onto screen earlier]

The excerpt captures the nature of questions regularly posed to students during whole-group discussions throughout the duration of the academic year and how they warranted narrow surface-level responses. In this case, all questions were satisfactorily answered with one-word (or number) responses, except for Jesus' response in line 36: "Quartile two", and did not require students to engage in complex reasoning. For example, the question leading up to Ms. Gellar's mention of quartiles (lines 12-13) asks students to determine how many "pieces" a boxplot consists of if students "split [the distribution] in half and then [they] split each half in half." While the elicited response provided a smooth transition into the concept of quartiles, this pattern of interaction did not provide opportunities for students to reason substantively about their data nor about new concepts, and instead required a very basic count of 'pieces' to render the anticipated and satisfactory student response of "4" (line 14). Additionally, by the time Ms. Gellar asked students what each quarter was termed (line 21), she had already explained that each quarter was referred to as a quartile, provided a working definition for quartiles, and projected a slide that articulated a similarly worded description of quartiles which remained on display as students answered her question. Therefore, responding to the question merely required that students either recall the term she had recently described or that they read the contents of the slide pictured in Figure 4.10.

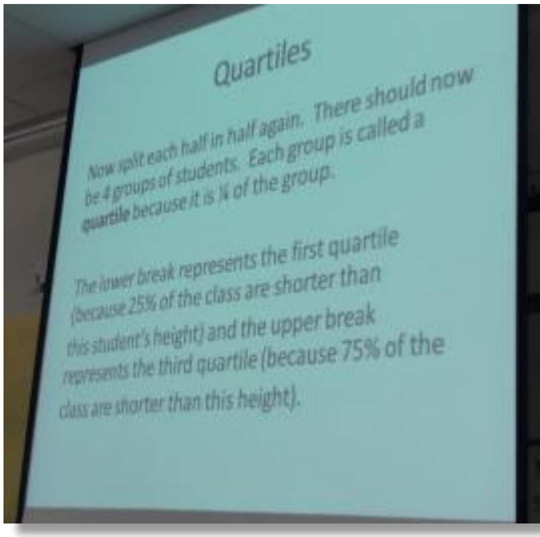


Figure 4.10 “Quartiles” slide projected during discussion of quartiles.

This demonstrates how whole-group discussion functioned primarily to check of student knowledge rather than as an opportunity for students to engage in the co-construction of knowledge. We see more examples of these types of checks as indicated in Table 4.2:

Table 4.2
Examples of Questions Posed During Whole-Group Discussion to Check Student Knowledge

Line	Question	Response
1-2	“[H]ow many people are in each half [to the left and right of Mervin] if we’re not counting Mervin?”	“19.”
30-32	“This is quartile one [*labels “Quartile 1 (Q1)”]*, we can abbreviate this as Q1. And this [*points to “Median” already written next to Mervin’s height mark*] really, what would this be?”	“Quartile two.”
34-37	“[W]hat percentage of people fall below quartile one? ... What *percentage*?”	“25.”
38-39	“What about the median? What percentage of people will fall below the median?”	“50.”
41	“What percentage will fall below quartile three?”	“75.”
46	“[W]hat percent fall between quartile three and quartile one?”	“50.”

Furthermore, the excerpt above also contains instances where the teacher was responsible for determining the legitimacy or correctness of student responses, not unlike traditional mathematics and science courses. For example, one instance occurs when Ms. Gellar interprets the human boxplot she drew using students’ heights (lines 44-47). After rhythmically stating “25% of you are shorter than this. 50% of you are shorter than this” She begins to voice a third assertion “75% of you are—” (line 45) at which point Diego chimes in “Taller—” (line 46). This is followed by Ms. Gellar’s correction of Diego’s assertion via the continuation of the pattern in which she had begun to voice her boxplot interpretation, “shorter than this.” It is important to note, however, that Diego’s assertion, while technically incorrect, acknowledged that heights could, inversely, be described in terms of the percentage of students that are *taller* than Q1, the median, and Q3. Thus, even though Ms. Gellar offered one way to interpret the boxplot, by no means was this the only way. I surmise that Diego’s offering of “taller” was a positional response informed his own height; and so, he was interpreting the boxplot as someone who was taller than (at least) 75% of all students in the class (Figure 4.11). Similarly, it is also possible that Ms. Gellar’s assertions about the boxplot were also positional.

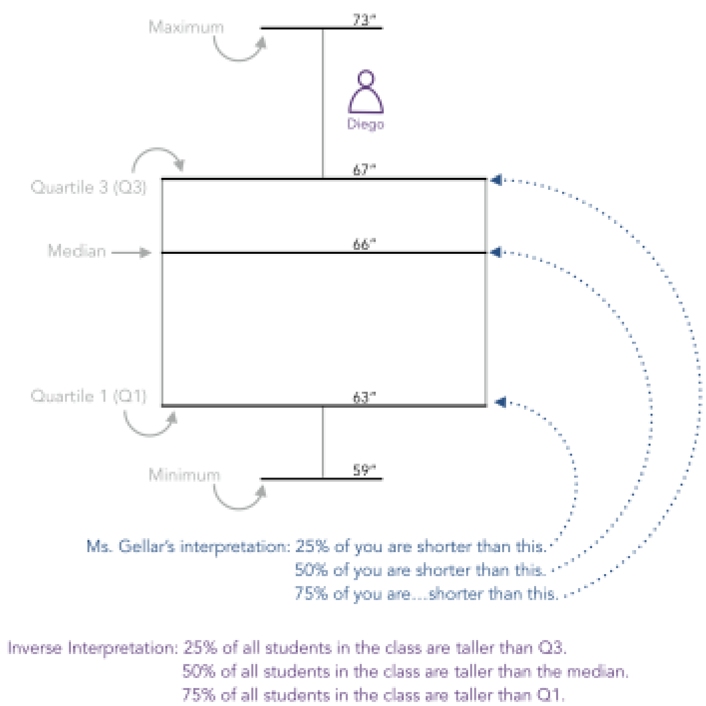


Figure 4.11 Positional interpretations of the human boxplot.

Here, lost on all was an opportunity to co-construct new and emergent understandings about the role positionality and subjectivity play in different stages of the data life cycle including stages of interpretation and graphical representation. Moreover, although Ms. Gellar was responsible for assigning and determining the purpose and meaning of tasks; determining the legitimacy of solutions and responses; and establishing standards for legitimate execution of solution methods, this was not the case when students were in the computer lab setting where RStudio lab slides determined what it meant to do data science in legitimate ways by delineating specific procedures students had to follow to develop the coding skills necessary to complete lab assignments. I will expound on data-scientific activity as it pertains to the computer lab setting during my forthcoming discussion of the second specifically data-scientific obligation.

Students contributed to the distribution of authority by demonstrating a willingness to participate in patterns of classroom activity, namely answering Ms. Gellar's questions, taking notes when prompted, and showing attentiveness when Ms. Gellar delivered lessons and when peers offered assertions during whole-group discussions. While authority in the classroom was mostly distributed to the teacher, a select number of students—regarded by their peers as having strong academic or mathematics backgrounds—gradually gained recognition among their peers as experts. As a result, these students gained some level of authority when it came to decision-making about the legitimacy of solutions and reasonableness of solution methods *among peers*. Their guiding principles for decision-making, however, were ultimately premised on Ms. Gellar's explicit and implicit conveyance of data-scientific knowledge and what it meant to engage in legitimate data science-doing in the classroom. Students routinely cooperated with Ms. Gellar's directives to meet disciplinary expectations and seldom challenged her data-scientific assertions. In fact, I only observed five separate episodes where students challenged Ms. Gellar's data-scientific assertions, compared to 32 separate episodes where students challenged each other's. Students' disinclination to challenge the teacher points to their recognition and acceptance of her authority in the classroom, suggesting that students held themselves accountable to the teacher.

I also believe that personal investments and conviviality influenced students' willingness to cooperate in the establishment and fulfillment of expectations that defined what it meant to effectively and legitimately engage in data science-doing. For instance, 40 of the 42 students enrolled in the IDS class were seniors and some were taking IDS to make up for a previously failed math course to graduate. Still others, including the two juniors enrolled in the course, were personally invested in earning high marks and viewed meeting the teacher's disciplinary expectations as essential for earning a good grade. Undeniably, those relying on passing IDS to graduate were invested in doing well in the course, or at the very least securing a passing grade. Additionally, Ms. Gellar and her students shared a respectful relationship and communicated with each other cordially; some students even joked with her lightheartedly, demonstrating their rapport. I believe that the interpersonal relationships that existed within the classroom and students' sentiments toward Ms. Gellar strongly influenced their willingness to partake in the classroom community by observing the rules for social and disciplinary engagement initiated by Ms. Gellar.

Opportunities for students to exercise agency. The first general classroom obligation rested on the assumption that the teacher was responsible for imparting data-scientific knowledge and that students were responsible for understanding and applying that knowledge. While students contributed to classroom discourse, it was mainly in the capacity of demonstrating their grasp of disciplinary knowledge, this included recalling data-scientific concepts and satisfactorily executing procedural and/or calculational steps to solve problems. Moreover, students were not positioned as co-contributors to data-scientific knowledge in the classroom because knowledge itself was treated as established, self-evident, and true. As such, students were not expected to participate in decision-making about the interpretation of tasks, legitimacy of solutions, and reasonableness of solution methods—this obligation rested squarely with Ms. Gellar and RStudio lab slides given the narrow distribution of authority in the classroom. In fact, all obligations and their constituting activities functioned to establish and maintain the authority of the teacher and lab slides as representatives of data-scientific subject matter by narrowly defining the ways students could legitimately contribute to the classroom. What constituted legitimate classroom contributions was also narrowly defined as learning disciplinary understandings to use them as the basis for making decisions about appropriate solution methods and how to execute them to complete tasks.

Engaging in various classroom activities that at their core were concerned with completing tasks afforded students opportunities to exercise disciplinary agency by doing things like matching solution methods to problems and following established procedures to problem-solve. However, this did not mean that students understood why certain solution methods were more appropriate for certain problems, nor that they understood the principles that underpinned disciplinary skills and concepts. Below, I present two excerpts from my classroom observations to discuss how issues relating to conceptual data-scientific understandings were obscured because students were able to produce the necessary solutions, responses, and outputs to satisfy task completion.

The excerpt that follows takes place during lesson six of the curriculum after students were introduced to several tools and concepts including measures of center, measures of spread, boxplots, and shapes of distributions in the weeks prior. They were asked to compare two data distributions of commuting times for a high school student, one depicting freeway commuting times and another depicting commuting times on surface streets. Students were asked to use measures of center, measures of spread, minimum values, and maximum values to construct an argument about which commuting method was preferable.

RESEARCHER: Can you tell me what measures you used for these two [commuting times boxplots]?

…

ROSELYN: Well, we used median for both of them because—

RESEARCHER: For both?

ROSELYN: [Nods]

RESEARCHER: Why?

ROSELYN: Because, the…

JESUS: They're skewed.

ROSELYN: Yes, they're skewed, pretty much.

RESEARCHER: They're both skewed?

ROSELYN: Yes.

RESEARCHER: And you used median, you said.

AMBER: [Nods]

RESEARCHER: Why did you use the median instead of mean?

ROSELYN: It's the one we have to use when the thing is skewed [smiles].

RESEARCHER: Okay. Do you know why?

AMBER No, [Roselyn and Amber smile and look at each other nervously] we just know.

RESEARCHER: What about measures of spread?

ROSELYN: We used IQR.

RESEARCHER: For both?

ROSELYN: Yes.

RESEARCHER: Do you know why you would use IQR?

ROSELYN: [Roselyn smiles but does not respond]

JESUS: It says right here… [reads from class worksheet] "If the distribution is skewed or has outliers, it is best to use the median as measure of center and the IQR as a measure of spread." So that's where we got it from. I don't know *why* but that's on our notes so…

RESEARCHER: Do you have any idea why?

JESUS: No, I don't know—

AMBER: We really don't [laughs].

In the excerpt above, we see how Roselyn, Jesus, and Amber use the problem-solving approach of appropriately matching solution methods to problems, and we also see the shortcomings of this approach. Roselyn makes it clear, in line 16, that her group's decision-making was not an exercise in developing conceptual understandings but was instead premised on notions of what she and her peers were *obliged* to do data-scientifically. This tells me that she viewed matching solution methods to problems as a legitimate problem-solving approach students were responsible for executing. In support of Roselyn's response to me, Jesus further corroborates the legitimacy of her perceived obligation by citing Ms. Gellar's assertions/lessons items recorded in his notes (line 25-27). Further, once he states that the reasoning behind what makes the median the most appropriate measure for a skewed distribution is unbeknownst to him and his group-mates, he again alludes to the legitimacy of Ms. Gellar's assertions when he says, "I don't know *why* but that's on our notes *so*…" (line 27) suggesting that since guidelines for selecting measures only mention matching, reasoning about why certain measures are appropriate is not necessary. Also evident in the excerpt is Roselyn's use of the term "thing" to refer to the "distribution," which stood out to me given that at this point in the semester students had been introduced to and privy to countless mentions of "distributions" and the day of this interaction was no exception. Her flippant reference to the distribution as "the thing" suggests to me that she, presumably, did not feel obliged to use discipline-specific terminology for her sake or for me (since she is speaking to me) because her data science-doing was an obligation she was fulfilling for Ms. Gellar.

The next excerpt takes place in the lab setting as students were working in groups of three on one of the three practicums of the unit. Practicums were projects that asked students to use the skills and concepts imparted through RStudio lab assignments to conduct data analyses and write a report including supporting data outputs used to complete the task (i.e., graphs, codes, interpretations, etc.). This practicum, entitled "The Summaries," asked students to select a data set from one of their data collection campaigns and compose a statistical question that compared two or more groups by including a (categorical) grouping variable—that is, a variable composed of multiple groups such as gender (male/female) or age-group (child, teenager, adult), for example—and an additional (numerical) variable. In the exchange below, Katie tells me about her group's progress.

RESEARCHER: Can you ladies tell me how it's going with the practicum? …where are you in the process?

[…]

KATIE: Well, right now we're just trying to figure out which plot looks better with our variables.

RESEARCHER: Okay, what is your [statistical] question?

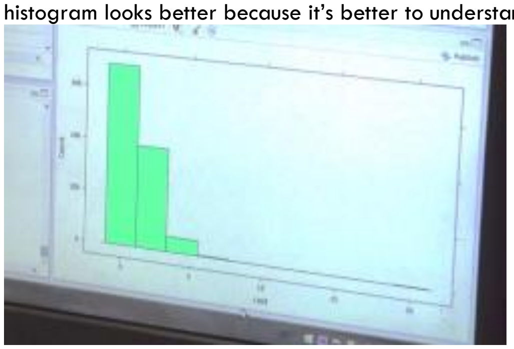
KATIE: [Reads from her computer monitor] "What variables affect the health level more, the cost or the sugar?"

RESEARCHER: "What variables affect the health level more, the cost or the sugar?"—Okay. So, what kind of visual are you creating?

KATIE: [Katie stares at her screen and smiles] I don't know, Miss [bemused]…[types a code into RStudio, rendering a histogram]

RESEARCHER: Can you tell me what you just pulled up—that visual—what does it mean?

KATIE: Well, right now, I'm just checking which one looks better, a histogram or dot plot and the histogram looks better because it's better to understand.



RESEARCHER: Okay, but can you tell me what I'm looking at?—what does that [histogram] mean?

KATIE: [Thinking]

RESEARCHER: What does this graph tell me about your question?

KATIE: It's telling me the cost, like, right here it just [circles cursor inside the bounds of the histogram]—telling me the cost.

RESEARCHER: It's telling you the cost of snacks… what does count mean? Count is what? What does that mean? ...[Katie stares at her computer screen, contemplating the question] I'm looking at your Y-axis. So on your X [axis] you have "Cost," right?—and then on the Y [axis] you have "Count"—what does that mean? What is that referring to?

KATIE: That, I do not know.

RESEARCHER: So I notice that the numbers on the count axis are 0, 100, 200, 300. What might that be referring to?

KATIE: Maybe how many people put that much, because there's like 500–something entries.

RESEARCHER: Hmm, okay.

KATIE: So, maybe the number of entries?

RESEARCHER: Okay, thank you.

As Katie works on creating an appropriate visual representation of the variables in her statistical question (lines 4-8), her words suggest that her interpretation of appropriateness is based on "which plot looks better," and although Ms. Gellar emphasized appearance/clarity as a factor in determining appropriateness, she also reminded students to think about other aspects such as types of variables (categorical or numerical), size of the data set, and the shape of its distribution. Therefore, while considering which plot looks better is important as it is directly related to clarity, what is problematic is that her consideration is not accompanied by a holistic conceptual understanding of what the plot conveys with regards to the data at hand. A comparison of Katie's group's question (lines 7-8) and the histogram she generated in RStudio reveal that although Katie thinks the histogram is "better to understand" (line 15), she has but a tenuous understanding of what it actually conveys as indicated in line 22, where she says it represents "the cost." Her interpretation demonstrates a surface-level reading of this histogram as self-evident, indicated by her response and her circling of the cursor to highlight the obviousness of meaning: the cost of snacks. When I ask her to elaborate on what it reveals *about* the cost of snacks, she says "That, I do not know" (line 28). At this point, I still had not realized that the histogram only depicted one variable and not the other two, health and sugar, included in her question. I assumed her Y-axis was related to either "health" or "sugar." Once I realized it was not, I asked her to explain what the Y-axis ("Count") meant rather than assume it's meaning, but she did not know (line 26-27). I found this problematic because it meant that her consideration of the histogram as a satisfactory and appropriate representation of the data was based only on the fact that the histogram included the cost variable. What it revealed about this variable, however, did not seem as important as the fact that it "[looked] better with [their] variables" (lines 4-5). Also, if "count" referred to the amount of entries (line 31) for the cost of snacks, this meant that the histogram only represented data for one of the three variables mentioned in her group's statistical question. It was unclear, however, if Katie realized this because of her mention of "variables" as plural (line 5). Of course, she could also have meant that after generating a histogram for all variables, she and her group-mates felt that it provided a clear representation for each one separately, except a close-up reading of her RStudio console indicated that she had not previously generated a graphical representation for the additional variables, and was instead working on a different task prior to generating this histogram for "cost." As such, I surmise that if Katie and her group-mates found this histogram appropriate for their question, they would have to generate two additional histograms and compare them separately which would have countered the purpose of using RStudio to streamline and improve their ability to analyze relationships among variables to understand what the data say.

Normative Data-Scientific Identity

In the episodes discussed above, we see two examples of students exercising disciplinary agency by selecting and carrying out solution methods in their efforts to complete tasks. What we do not see is their ability to reason with and about solution methods and the principles underlying their employment. While students demonstrated their potential to fulfill overarching classroom obligations of completing tasks and producing data outputs, understanding these types of activity in relation to their own personal and intellectual development were not necessary for the legitimate fulfillment of emergent classroom obligations. Furthermore, given the findings and analyses presented earlier, the normative identity of what it meant to do data science in legitimate ways as constituted in Ms. Gellar's classroom delineated specific ways of reasoning about disciplinary understandings, solution methods for completing tasks, conditions for peer collaboration, and standards for argumentation. Hence, legitimate data science doing consisted of:

- Listening and taking notes to develop disciplinary understandings and reference material
- Carrying out established procedures to produce data-scientific outputs and complete tasks
- Collaborating with peers to facilitate task completion
- Referencing established problem-solving procedures, reference materials, or the teacher's data-scientific assertions to justify data-scientific assertions when challenged by peers.

I view this normative identity as a *rubric* for what qualified as ideal data science-doing in this particular classroom, and as such, am not arguing that students uncritically or impassively abided by this constituted identity, but rather that for a number of reasons, they chose to cooperate in their constitution. In the next chapter, Personal Data-Scientific Identity, I will discuss how students felt about this “rubric” and how that informed the extent to which they came to cooperate in its constitution. While my analysis of classroom observations supports the notion that students did, in fact, cooperate in the constitution of classroom obligations, there were also ways in which students simultaneously resisted curricular assumptions about them and what they should learn. Because these sentiments regard students’ valuations of classroom obligations, I will reserve this discussion for the following chapter.

CHAPTER FIVE
Personal Data-Scientific Identity Analysis

In this chapter, I will present my findings as they pertain to students' valuations of classroom obligations and estimations of their peers' and their own data-scientific competency to understand the extent to which students came to identify with data-scientific-doing as constituted in the IDS classroom. Exit interviews revealed that students valued general classroom obligations as a necessary means of earning good grades, and valued specifically data-scientific obligations as key to acquiring the disciplinary competencies necessary to earn good grades. Students spoke of their own data-scientific competencies as consistent with the work of data scientists and thus based their estimation of themselves as successful data-science doers on their ability to carry out tasks and apply established disciplinary understandings. In this chapter, I will first discuss students' valuations of their classroom obligations as a whole, followed by a discussion of the learning rationale's that informed these valuations. I will then engage in a discussion of students' estimations of their peers' and their own disciplinary competencies, followed by an analysis of how students reasoned about data-scientific competency.

Students' Valuations of Classroom Obligations

When I asked students to tell me about what they learned in the IDS class, 10 out of the 12 student interviewees said they gained familiarity with data science-doing and specifically referred to learning the importance of carrying out specific procedures as they related to coding, calculating statistical measures, and generating data outputs in RStudio as instructed to in order to arrive at the *correct* solution or output. For example, Angie said that she learned that "if you don't do something exactly, you're not going to get the right result," indicating that in order to successfully complete tasks, she had to follow established procedures. She also expressed that she did not always have this understanding, and it wasn't until she saw her grade drop during the first semester of taking IDS that she realized that not following prescribed steps for lab completion was hurting her grade:

> In the beginning of the school year, when I would get error messages [in RStudio], I would get upset really fast and say, "screw it" and just put the code into my [lab assignment] even if it was an error. Now I've learned that you need to fix codes because they will affect the labs and if it's not correct, you need to go back and check again…Last year I got a "B" [in IDS] and it upset me because the last few weeks were labs so my grade kept dropping. So, this semester I took more time doing the labs and [would] do them at home. Even if I didn't get the codes right, I would just press F1 for more info…I've actually learned how to be faster with the codes because last semester I would have to go back and look at old codes but this semester I mastered them. I know that if I type "view" [the data set] will pop up or if I type "n row" it'll show me the variables [in a data set]. It's very useful knowing the simple codes.

This excerpt of my interview with Angie reveals that while she initially resisted fulfilling classroom obligations during the first semester, as evident in her refusal to resolve errors (lines 1-3) using established procedures such as typing codes exactly as they appeared in RStudio lab slides, her investment in a good grade eclipsed her frustration with error messages (lines 4-5). When she realized that not following lab procedures with exactitude was hurting her grade, she decided to modify her approach in efforts to secure a better grade during the second and final semester (lines 5-6). Thus, Angie's efforts to follow lab procedures to complete labs during the second semester were motivated by her endeavor to earn a top grade in the class after seeing her grade drop in the first semester.

Similarly, Diego and Kim, who had previously taken and failed IDS with Ms. Gellar, also shared that they learned from their mistakes the previous year and sought to meet disciplinary expectations in order to earn better grades the second time around. In the excerpt below, Diego shares how his experience taking IDS a second year differed from his experience taking it the first year it was offered:

> Last year I didn't really care about this class. It was confusing because it was about something we had never learned about. It wasn't like geometry or algebra that just continue what we have already learned. This year, I got the hang of it with the labs and it's pretty fun because I know what I'm doing…Last year I

wouldn't try and this year I actually cared because I didn't want to have a bad grade…it looks bad to have bad grades. It was an elective but I needed it for [the] math [graduation requirement]. Last semester I had a "B" and this semester I have an "A." … [Last year the labs] got confusing throughout the [year]. If you don't understand the last one you did, you won't understand this one. And if you don't do well, it's worth a lot of your grade. If you don't do well on the labs, it'll bring your grade down a lot…In class I'm more involved. I try to answer the questions the teacher asks the class. I turn everything in on time.

Like Angie, Diego also resisted observing classroom obligations the first time he took IDS because he did not care for it and found it confusing (line 1). In lines 3-4 he mentions getting "the hang of it" in the second year, which I interpret as a strategic move to engage in practices necessary to secure a better grade, although not necessarily to gain conceptual and more sophisticated data-scientific understandings. Given that classroom obligations delineated ways of gaining familiarity with data-scientific skills and concepts and applying these skills and concepts to task completion, getting "the hang of it" refers to perceiving the existence of disciplinary expectations and engaging in classroom activity that seeks to meet those expectations. Indeed, in lines 4-5 Diego clearly conveys that like Angie, he too was more invested in getting a good grade in the class and cared about how bad grades are perceived. Aside from his investment in earning a top grade during his second year in IDS (lines 6-7), passing the course had major implications for Diego's future as he needed the class to meet a graduation requirement. Thus, Diego felt that the stakes for meeting classroom obligations were high. Additionally, Diego offers a subtle critique of labs and the negative implications of not completing them by following established procedures based on his own experience failing the class when says "if you don't do well, it's worth a lot of your grade. If you don't do well on the labs, it'll bring your grade down a lot" (lines 7-8). By failing IDS the first time, Diego learned that understanding expectations and working to fulfill them was necessary for successful completion of the course. This is evidenced in line 11 when he alludes to participating in the initiation-response-evaluation method of whole-group discussion by answering “the questions the teacher asks the class” and completing assignments.

Furthermore, no students explicitly expressed discontentment with what was expected of them nor their efforts to meet classroom obligations, but four students spoke of their personal interests and capabilities as incompatible with data science-doing based on its constitution in the classroom. Diego expressed that data science required a lot of math and while he enjoyed math and found IDS fun (line 3 in the excerpt above), he felt he would get bored of it if he did it for a living. Angie, who expressed her desire to pursue a career in marketing said that data scientists worked in front of a computer all day and she preferred being out in the world. Both Dolores and Angie expressed this sentiment during one of our exchanges in the classroom. Both in class and during our exit interview, Dolores expressed that she ambitioned to become a dental assistant as she found this career path to be more in-demand compared to data science—a perspective that was undoubtedly influenced by her assertion that those in the medical field are better positioned to help society as opposed to someone sitting in front of a computer. Gerardo said that data science required too much memorization of codes and he found this particularly boring. Thus, while these students conveyed little interest in pursuing an education or career in the field of data science, they nevertheless cooperated with Ms. Gellar and participated in the fulfillment of classroom obligations.

Students valued the expectation that they meet classroom obligations to the extent that they were personally invested in earning a good grade in the class either to meet their short- or long-term goals. For example, seven out of the 12 student interviewees expressed their personal investment in performing well academically in service of achieving a successful career. For example, Angie valued the instruction she received in the IDS class because she felt it introduced to her to working with a computer, which she felt was an important familiarity to have for future careers. She was particularly interested in studying marketing to help her father achieve his dream of owning his own business, and so, she felt that what she was learning in IDS would help her with business-related decision-making. While it was evident that she valued IDS for helping her gain experience working with computers, she did not voice a valuation for data-scientific skills and concepts.

Nube, one of two juniors in the class, conveyed that earning good grades was important for her long-term goal of gaining acceptance to a four-year university, and subsequently medical school, to become either a cardiologist or a pediatrician. At the time of our interview, Glenda had been accepted to UCLA and endeavored to become either a physicist or a surgeon. Like Nube, Armando also spoke of his desire to become a pediatrician but simultaneously expressed a profound curiosity for technology, namely mobile technologies, and cited software development as his back-up plan. Andres, Adan, and Sandra shared that they were interested in pursuing careers in high-demand fields including computer engineering, electrical engineering, and software development, respectively. These students' long-terms goals of performing well academically in high school to facilitate their admission into higher education, and eventually into their desired career paths, played a highly influential role in their perception and valuation of classroom obligations and indicate that they viewed the course as instrumental to achieving other goals rather than as intrinsically valuable or personally enriching.

Moreover, while other students also expressed an interest in acquiring a postsecondary education and careers in STEM, their valuation of classroom obligations was informed by more immediate goals related to high school graduation. Diego and Kim, also endeavored to pursue higher education, but Diego's post-high school plans to go into law enforcement and Kim's plans to study astronomy rested on their ability to pass IDS in order to meet their graduation requirements. Dolores, spoke of her desire to become a dental assistant, and while she did not mention that she needed to pass IDS to graduate, she emphasized her dedication to educating herself, heeding her parents' warning, "if you don't get yourself educated, you're not going to be nothing in life." Dolores felt that part of preparing for her career path was to ensure that she earned top grades and graduated high school—symbols of her efforts to educate herself. Lastly, Gerardo did not express interest in pursuing a career in data science, and while he was still indecisive about his plans after high school at the time of our interview, he said that he was actively working with a military recruiter, exercising, and studying for the SBAC exam new recruits must pass to enlist in the Air Force. Another requirement for enlistment in any military branch is earning a high school diploma, thus Gerardo's observance of classroom obligations was presumably tied to his desire to either continue his education at a community college or enlist in the Air Force. It is also important to note that 11 out of 12 students mentioned that technology and computers are central to future high-paying in-demand jobs—a perception that many felt was influenced by their economics, history, and government teacher, Mr. Suarez who repeatedly spoke about this to those enrolled in his class, according to students. This means that students participated in the IDS class with an understanding of the value of working with computers but not necessarily with an understanding of the importance of cultivating data-scientific literacies for civic engagement and everyday life.

Students' Rationales for Observing and Valuing Classroom Obligations

Structural rationales for data-scientific learning. Students' valuations of classroom obligations as necessary to performing well academically and earning good grades to meet immediate and long-term goals were supported by structural and/or situational rationales for learning. Interviews revealed that eight out of the 12 students found participating in data-scientific activity structurally significant as a means of attaining or achieving something outside of themselves such as good grades, college acceptance, and peer acceptance. Outside of facilitating access to their respective educational and career goals, when asked how they would use the skills they developed in the IDS class in their personal lives, five students expressed that they did not perceive their skills as useful for their out-of-school lives. This means that while they found cooperating with classroom obligations to be structurally significant, their participation in the classroom was not motivated by a desire to gain mastery or expertise in data science, and classroom obligations did not provide opportunities for students to develop these interests because they emphasized students' roles as learners, reserving roles of beholders of disciplinary knowledge for the teacher and RStudio lab assignments. Due to the fact that classroom obligations narrowly defined data science-doing, students who did not already have disciplinary interests that were compatible with data science, or ambitions to pursue a career related to data science were not afforded opportunities to develop situational rationales for learning data science. This finding is worthy of serious consideration for STEM reform efforts that specifically set out to encourage the development of data-scientific thinking and identification with data science-doing among students who have been historically excluded from STEM education and careers because it suggests that while students with pre-existing interests in STEM benefitted from limited opportunities to grow these interests, those without pre-existing interests did not.

Situational rationales for data-scientific learning. Students who only viewed data science as consisting of a technical skillset, did not speak of disciplinary understandings as valuable outside of school, college, or data-oriented careers such as marketing, statistics, and computer science. Those that developed more nuanced understandings of the conceptual affordances of data science, however, viewed data-scientific skills and understandings as consisting of both a technical skillset and an evaluative mindset, thus these students were able to derive life skills from their experience taking IDS. Adan, Armando, Kim, and Sandra felt that the skills they learned in the IDS classroom helped them to not only develop a technical skillset but also push how they thought about mathematics, science, mediated representations of data, and decision-making and problem-solving in their everyday lives. For example, by taking IDS Armando realized that he routinely and unknowingly engaged in data science-doing at home to solve everyday problems related to budgeting and grocery shopping. He also mentioned using his now-bolstered tendency to think data-scientifically in other non-IDS/non-STEM classes:

> Instead of doing a basic interpretation of a graph, you can see that there's more behind it…for example, in another class, I had to do my college plan. It wasn't just about searching what school I want to go to and how far it is. What I did was look for different routes to get there, alternate routes, alternate modes of transportation, classes offered, transfer rate, what schools are better for what, and determine which schools would be better for my career choice.

In the excerpt above, Armando emphasizes that what he learned in IDS helped boost his ability to think critically about important decisions in ways that supersede surface-level understandings of data and information in general (line 1). What's more, he goes on to talk about how deeper-level thinking can be applied to solving problems and completing tasks that may or may not be directly related to data science such as choosing a university (lines 1-2). In lines 3-5 he demonstrates how deeper-level thinking allowed him to engage in informed decision-making about important life choices.

Similarly, Adan shared that taking IDS helped him realize that he, too, would engage in data-scientific thinking to make informed decisions about everyday life outside of the IDS classroom. He said that while he would research things before arriving at a conclusion or solution, taking IDS solidified his feelings about the importance of making informed decisions like choosing a college or buying a house by conducting research, but also provided him with specific tools and techniques to do so. Kim, a professed lover of math, said that what she learned in IDS helped shift her thinking about math:

> Ever since elementary, I really loved math. I'm in trig right now and have 104% and…it's really easy but I always thought math was 1 + 1 or…a = a…. Math is always right, you can't prove math wrong. But doing data science, I can see that math isn't always 1 = 1. I'm not saying 1 = 1 is wrong, but there's a lot of ways to express that, like 1 = .5 + .5—like when we answered a statistical question, there were many ways we could answer one question, or many ways we could ask the question to get the same answer. It's like they say, "there's many roads to Rome." I think it helped me broaden my mind about math and the concept of mathematics combined with computer science and how it all connects.

Kim was very confident in her math competency, and actually cited pride in her academic work as the reason for failing IDS her first year taking it because she dared not turn in incomplete or late work as it would not accurately reflect her self-perceived mathematical competence. She valued her experience taking IDS because it helped broaden her scope of what it meant to do math and what constituted legitimate mathematical responses and conclusions (line 7). Lines 4-7 demonstrate that not only did she develop technical skills necessary to write appropriate statistical questions as defined by Ms. Gellar, she also gained conceptual understandings by thinking about the purposes of engaging in activities and completing tasks. For Kim, the ability to draw conceptual connections (line 8) between data science as a field and data science as a concept proved highly valuable for her own personal development. Moreover, Sandra felt that taking IDS, overcoming challenges, and being regarded as a highly competent data science-doer by her peers allowed her to develop personal characteristics that were transferable to different aspects of her life. To this effect, she offered, "[I learned] to not give up—keep trying, find different ways to arrive at solutions. To be open to different methods and not just stick to one approach to solving issues.” In our interview, Sandra also spoke about the versatility of technical skills she gained in IDS as illustrated in the excerpt below:

> Data science will make a good impact with any major because when it comes to gathering information, that’s pretty much what any job will require. Having knowledge of data science and knowing how to interpret information with computers you’ll have a faster way to find real-world solutions now that everything is going toward tech.

Her assertions demonstrate that her experience taking IDS allowed her to develop nuanced understandings of the technical and conceptual affordances of data science. She points out that the skills she developed would be useful for conducting research in different academic disciplines (“major” in line 1) and real-world problem-solving. Below, Table 5.1 provides a breakdown of these responses and classifies the nature of students’ skills based on how students described skills and their applicability.

Moreover, Adan, Armando, and Kim spoke of more personally meaningful rationale’s for wanting to perform well academically. While all students conveyed that they were motivated to fulfill classroom obligations as a means of achieving career goals that required high academic performance, Adan, Armando, and Kim also spoke of their desire gain “access to experiences of mastery and accomplishment” (Cobb & Hodge, 2010, p. 185). Despite the fact that classroom obligations narrowly defined data-science doing, students like Adan, Armando, Kim, and Sandra

Table 5.1 Student Views on the Applicability of Data-Scientific Skills to Their Personal Lives

Use of data-scientific skills in your personal life?

No	*Yes*	*Skill Classification*
Angie		Technical
Antoine		Technical
Diego		Technical
Dolores		Technical
Gerardo		Technical
Glenda		Technical
Nube		Technical
Andres		Technical
	Adan	Life
	Armando	Life
	Kim	Life
	Sandra	Life

who expressed having disciplinary interests that were compatible with data science and endeavored to pursue a career related to data science were afforded opportunities to bolster their situational rationales for STEM learning in IDS. Additionally, Adan, Armando, and Kim, in particular, spoke of a life-long passion for disciplines central to data science: mathematics and computer science. In the following chapter I will provide analytical portraits of Armando and Kim as examples of students who engaged in data science-doing in ways that indicated developing critical data-scientific understandings. For now, I will proceed by discussing how students' valuations of general classroom obligations reflected their perceptions of obligations as obligations to themselves.

Student views of classroom obligations as obligations to themselves. All student interviewees expressed either structural or situational rationales for valuing classroom obligations. Ultimately, all students were invested, to varying degrees, in meeting these obligations because by doing so they were also working toward their own efforts to earn good grades. While they were all personally invested in performing well academically in the IDS classroom, their motivations for doing so differed as discussed above.

I also found that while students with situational rationales for valuing classroom obligations were motivated to perform well academically in order to earn good grades, they were also motivated by pre-existing valuations of mathematics, in the case of Adan and Kim, and computer science, in the case of Armando, as captivating and personally meaningful. Thus, while these students' also valued classroom obligations as obligations to themselves, they valued them in service of their personal disciplinary interests and intellectual curiosity, and were, in fact, building on existing and profound disciplinary interests. This finding supports the notion that future iterations of programs that seek to inspire interest in the field of data science among students from non-dominant groups should endeavor to provide opportunities for students to develop conceptual understandings in service of exercising conceptual agency in the classroom as a way of fostering identification with data science-doing among all students, including those who do not already have pre-existing interests in STEM generally. A useful measure of whether students are beginning to identify with data science-doing is whether they are fulfilling classroom obligations as obligations to others or obligations to themselves for their own personal and intellectual enrichment (Cobb et al., 2009; Cobb & Hodge, 2010).

Students' Estimations of Their Peers' and Their Own Data-Scientific Competency

The majority of students whom I interviewed, with the exception of Gerardo, expressed that they perceived themselves as successful data science-doers in Ms. Gellar's classroom. Their assessment of their competency and how they defined success was based on their perceptions of what it meant to do data science and what they perceived to be the work of data scientists based on their experiences in the classroom. For example, according to Glenda, data scientists are responsible for gathering and analyzing data. She went on to describe the skills she gained in the class as consisting of developing disciplinary understandings (i.e., learning about measures of center and spread) and technical skills (i.e., learning to use RStudio to generate graphical representations) that enabled her to do the work of data scientists as she perceived it. Moreover, Nube said that the work of data scientists involved researching topics by creating statistical questions and going through the data cycle to draw conclusions for decision-making. Furthermore, she described the skills she gained in the IDS class as consisting of those necessary to analyze data for decision-making. According to Nube, her skills included developing statistical questions that anticipate variability in responses; administering surveys to collect data; and using RStudio to understand the data and generate graphical representations. Because students' perceived skills and understandings were generally consistent with their perceptions of legitimate data science-doing, their estimations of self-competency in data science were positive and generally high. However, being critical of data and data artifacts, telling stories with data, and examining interesting patterns and exceptions in the data did not figure into student articulations of what it meant to do data science. Additionally, while they spoke to learning how to carry out solution methods and procedures for data analysis in RStudio, they were not encouraged to be critical of those processes nor of the data itself. This is problematic because while they gained experiences carrying out data-scientific inquiries, they did not gain experiences that encouraged criticality toward the politics of data, data collection, issues of privacy and surveillance, and cooptation of user data by large firms.

Even though Ms. Gellar only identified five of the 12 student interviewees as demonstrating high academic achievement in her class, 11 of them expressed to me that they had high estimations of their data-scientific competency, which I believe can be attributed to their observance and cooperation in the fulfillment of classroom obligations. A relevant finding here is that out of the 11 students who held high estimations of their data-scientific competency, 10 of these students attributed their estimation of self-competency to their ability to carry out established solution methods or follow established procedures for task completion. Relatedly, nine of these students attributed their estimation of self-competency to their ability to recall disciplinary understandings. For this reason, how students perceived their data-scientific competency is not a testament to whether they accurately understood disciplinary skills and concepts as intended by the teacher and the curriculum. Instead students' estimations of themselves as data science-doers were concerned with how they came to understand what it meant to do data science in accordance with their experiences as members and participants in the classroom community. Thus, students felt that they were developing data-scientific competency as long as they went through the motions of the class and actively worked to meet perceived expectations.

Angie and Diego's accounts are useful for understanding how students reasoned about their peers' and their own data-scientific competencies as closely related to personal effort. Angie and Diego both spoke about initially resisting fulfilling classroom obligations out of difficulty and frustration and attributed their growing data-scientific competency not to the development of conceptual understandings, but instead to their decision to follow directions for problem-solving and task completion. In telling me about how she was performing much better in the second semester of IDS, Angie cited her decision to follow lab procedures as instructed in RStudio labs as the key to her success in the class. Diego's perception of what constituted "success" in IDS included knowing how to carry out established procedures to arrive at the correct solutions and generate appropriate graphical representations. In this way, students viewed data-scientific competency as directly related to following procedures for task completion rather than gaining conceptual data-scientific understandings. There was an apparent overarching sentiment among students that achieving anything, including data-scientific competency, was only a matter of effort. This was positively reinforced when students like Angie and Diego followed directions for task completion and problem-solving, resulting in improved grades regardless of their level of conceptual understandings. Even students who expressed no interest in pursuing careers in STEM, or specifically data science, said that they still believed they could become data scientists if they so desired as this was only a matter of effort. Thus, the difficulties that students faced were perceived as within their control. Upon realizing the negative effects that their resistance was having on their academic achievement, Angie and Diego's valuations of what was expected of them changed. The shift in their valuations and their decision to cooperate with classroom obligations were primarily driven by their attempts to avoid the negative repercussions that resisting classroom obligations had on their grade.

Furthermore, just as overcoming difficulties was viewed as within their control, so was achieving facilities. Indeed, both Angie and Diego felt that they achieved some level of data-scientific mastery by recalling established disciplinary understandings and following established procedures for task completion. Altogether, students conveyed that they possessed the capability to fulfill classroom obligations and viewed difficulties and shortcomings as having to do with their own effort rather with factors outside of themselves. They viewed themselves and their peers as fully capable of performing well academically, suggesting that their estimations of competency were not concerned with what they knew, but rather their willingness and motivations for fulfilling classroom obligations. While this view can prove empowering, it disregards the creativity, improvisation, and higher order thinking necessary to push the field of data science forward and shape new and ever-shifting ways of thinking and working with data in the 21st century.

Extent of Student Identification with Data Science-Doing

Taken together, students valuations of their classroom obligations and estimations of their peers' and their own data-scientific competencies indicate that they viewed classroom obligations as obligations to themselves in service of achieving their educational and career goals. Thus, their commitments to gaining data-scientific understandings did not necessarily indicate interests or commitments to the field itself. While this dynamic can be expected in traditional mathematics and science classrooms (Boaler & Greeno, 2000; Hull & Greeno, 2006), one must be mindful of the fact that IDS was not a traditional mathematics or science classroom and was instead supposed to inspire motivated student interest in data science as a way to achieve equitable outcomes for non-dominant students in STEM. Some students cooperated in the constitution and fulfillment of classroom obligations in strategic attempts to pass the class, earn good grades, or gain familiarity with technical skills that they believed were of relevance to a future driven by technology and computers. Significantly, some students also cooperated in the constitution and fulfillment of classroom obligations because the learning experiences they were gaining supplemented compatible disciplinary interests in mathematics and computer science. These students viewed the IDS class as an opportunity to develop their disciplinary understandings, gain skills relevant to their interests, and satisfy their personal intellectual curiosities.

My analysis of students' personal data-scientific identities revealed nuanced motivations and rationales for student participation in the classroom community and the extent to which they came to identify with data science. It also revealed that while students were willing to participate in the fulfillment of classroom obligations, their estimation of the usefulness of data science to themselves, their families, their communities, and society as a whole were important factors that motivated their curiosity and interest in pursuing postsecondary study and careers in data science. While students felt that they gained technical and/or conceptual skillsets, their perception of the applicability of these skills to the real-world influenced their valuation of data science as a field and as a concept. This is an important finding for educators, curriculum writers, researchers, and policymakers invested in equity-oriented STEM reform because it bespeaks that narrowly defined expectations for student learning and classroom participation do not provide opportunities for students who do not already have preexisting interests compatible with data science to develop those interests. Conversely, this instantiation of a STEM reform effort provided opportunities for students with interests in mathematics, computer science, and technology to expand their scientific understandings. Indeed, these students felt that taking IDS helped them develop more than technical skills and disciplinary understandings and instead allowed them to gain life skills by applying data-scientific thinking to problem-solving and decision-making in their everyday lives.

In the following chapter I will take a closer look at the classroom participation and interviews conducted with Kim and Armando as two students who began to develop what I term Critical Social Data-Scientific (CSDS) identities. While I do not believe that their development of CSDS identities came to full fruition during their time taking IDS, I will discuss factors that contributed to these students' opportunities to reason in personally significant ways with and about data for everyday problem-solving and decision-making.

CHAPTER SIX
Developing Critical Social Data-Scientific Identities: The Case of Kim and Armando

In this chapter, I will provide an in-depth analysis of Kim and Armando's exit interviews as well as their participation in the IDS classroom to discuss how and why they began to develop Critical Social Data-Scientific (CSDS) identities, unlike the majority of their peers. I focus on these two students because their participation and perspectives revealed that they were beginning to reason conceptually with and about data in ways that indicated 1) in-depth reasoning about relationships between data science skills and concepts; 2) thinking about data and data artifacts in real-world context; and 3) thinking critically about aspects of data collection and analysis. Indeed, I argue that these three criteria constituted their developing CSDS identities. In an effort to contextualize Kim and Armando's learning rationales, educational aspirations, and their approach to data science-doing in the classroom, I will provide an analytical portrait of each followed by a discussion of how they engaged in each of the three criteria and opportunities that influenced their ability to do so. I believe that taking a closer look at Kim and Armando as case studies will enable a nuanced understanding of particular mechanisms that enabled them to engage in personally meaningful and enriching data science-doing within and outside of the classroom. While I do not argue that Kim and Armando fully developed CSDS identities, I do argue that they began to cultivate significant understandings and ways of reasoning that, provided the appropriate scaffolding and support systems, could be supported in future iterations of equity-oriented STEM initiatives that seek to bolster student identification with STEM-doing, particularly within data science.

Introducing Kim

When I interviewed Kim toward the end of the second and final semester of IDS, she was well on her way to passing the course and was considered by Ms. Gellar as one of the top students in the class. Our conversation revealed that Kim's academic identity as a high-achieving student was a very important facet of her identity as an individual. Kim was the youngest of four siblings and while she would not be the first to graduate high school, she spoke of pursuing a college education as a daunting task that none of her older siblings had been able to see through:

> I'm going to go to WCC [Western City College] for a year [after high school] …I'm going to bust my butt trying to [transfer to a university] because I have really bad experience with college…. [N]ot me personally, but in my family—no one has graduated from college in my entire family, my extended family. No one has graduated college.

This was Kim's second time taking the course and our exit interview revealed that not only was she determined to perform well academically and pass the class, but she was also on a mission to prove to herself and her family that she would graduate high school and be the first in her family to persist through higher education. Kim spoke with regret and disappointment of having failed IDS her first year taking it and felt that doing so undermined a very personal promise she made with her sister to prove to their mother that they would achieve what their brothers had been unable to:

> When my brothers didn't make it to college, and my sister was a senior [in high school], she was like, "You and me have to go to college. We have to prove to my mom that the girls can do it." But then when she dropped out, she was devastated. I mean she loves her baby, but she wished that she could have stayed in school and not get pregnant. And now I'm just like—I don't know. And messing up in high school with my grades, I am just like [*pensive*]... I was the smart one. I was always the smart child, and they had a lot of hopes and dreams for me but I was like [*trails off*]... And so, I'm going to WCC for a year because I got really, really high scores on my placement test, so I don't need those extra years. [Counselors at WCC] said that if I can really pick up the pace and take Summer and Winter classes, that I can be out of there by a year and a half, at most. And I was like, okay, well, I'm going to do it. I'm going to do it because I want to transfer and I want to be the first in my family to get my bachelor's degree. I want to be that person.

Undoubtedly, failing IDS ushered in an internal struggle for Kim wherein she tried to reconcile what it meant to be "the smart one" while simultaneously "messing up" her grades in high school (lines 5-6). She refused to concede that failing IDS necessarily meant that she did not understand data-scientific skills and concepts, citing elsewhere in our interview that her non-passing grade was a result of a personal decision to not turn in incomplete or late work, and thus did not speak to her academic ability. The excerpt above conveys the personal significance that passing IDS and graduating high school had not only for Kim, but also for her family who supported the constitution of her strong academic identity. She viewed her failure to pass IDS, and the threat that it posed for her high school graduation and ambitions for a higher education as a letdown to herself, her family, and the “hopes and dreams” they held for her (line 6-7). Her strong desire to “be that person” (line 12) in her family to earn a bachelor’s degree drove Kim to try her best during her second year taking IDS. Her cooperation in the constitution and observance of classroom obligations was central to her ability to perform well academically. In fact, being one of the top students in IDS in its second year of implementation also served to vindicate Kim’s strong sense of academic prowess as it helped demonstrate that getting a better grade was well within her academic capabilities and a matter of turning assignments in when due.

In addition to viewing her success in IDS as a means of uplifting her mother and meeting the academic expectations that her family had for her, performing well in IDS was also a way of pursuing her own pre-existing passion and curiosity for science and mathematics which had proved therapeutic and perspective-altering at an emotionally painful moment in Kim’s childhood.

> I think [my desire to study astronomy] kind of originated when I was little, like when I was 7 or 6 years old, probably. I had a really sucky life. All throughout middle school, I was bullied; and I always remember being outside, especially at night, I would look and stare at the moon; the moon is my favorite astronomical thing out there... I would look at the moon, I would look at the stars, and I’d be like, “Why am I in pain?” and I don’t mean that in like asking God, “Why are you doing this to me?” but I mean it like, “Why am I allowing myself to be put down by everything that they say or do, if I am just a small speck in the universe?” And that really helped me. “I don’t care what you say, because what you say doesn’t matter anyway; not even to yourself, because you are nothing compared to everything that is out there.”

> And I really want to explore what is out there. I want to not only explore the world, but explore every possibility that is out there; something that could make a difference in the world.

For Kim, reflecting on the stars, the moon, and outer space as a means of coping with the emotional and psychological distress she endured by being bullied helped shift her existential perspective in ways that enabled her to see beyond her immediate lived reality. Thinking of herself as a “small speck on the universe” proved therapeutic (lines 7-8) because it expanded the limits of reality, life, and possibility through scientific reflection. What’s more, I believe that Kim’s experience of overcoming being bullied and the critical role that science played in doing so cemented her view of science and scientific discovery as capable of helping others and making “a difference in the world” (lines 10-11).

Conceptual reasoning about relationships between data science skills and concepts. In our exit interview, Kim credited learning to work collaboratively with her peers as a boon to her ability to not only understand concepts, but to develop understandings regarding the purpose of tasks, the function of data-scientific tools and skillsets, and how they all related to larger data-scientific concepts. The following excerpt drawn from our interview is very telling in this regard:

> In math in general, the subject math, every math class that I had, I would be almost done and everybody would ask me questions, everybody would come to me. And when I have a question, I'd have to ask the teacher. But here, I put my ego aside and actually ask a student to help me so that…they can have the experience of being able to explain something, not only the teacher. [Mr.] Suarez, [my] history teacher, he’s always saying “What's 1 + 1?” And everybody's like “2.” And he's like “Why?” And [everybody says], “because 1 + 1 is 2.” Because like why? It's like because that's what we were taught. That's what we were conditioned. But we never know how to explain things. We just say because that's how we were told, because of this. We don't understand the concept of it. And here, I understood it because the fact that I understood it in the beginning so I can explain it to other people in different ways, so that they can understand it too. So, if I'm lost about something, I want to ask them their opinion so that I can see how they do it, and then probably incorporate that into my work. And they can have the opportunity to explain something that probably they knew how to do it, but they were kind of iffy on it, and probably if they explained it to me, they're half-explaining it to themselves too and then they can have a better understanding of that subject too.

Despite expressing strong confidence in her mathematical competency (lines 1-3), the excerpt indicates that Kim was not as well versed in collaborating with others and explaining or verbalizing her reasoning. She offers that IDS provided an opportunity for her to seek out the help of her peers as well as offer it, and by so doing enabled her to develop communication and reasoning skills. She shares that “the experience of being able to explain something” to others allowed students to partake in the type of privileged conceptual reasoning typically reserved for the teacher (lines 3-4), and that she felt that students should also be able to share in the experience of reasoning in ways that help establish conceptual understandings. Based on her own experience as someone whom other students often consulted for help, explaining often required her to explain the “why” of things. Indeed, as I posited earlier in Chapter 4, while students were able to share their responses, answers, and solutions with me, they typically did not have a clear understanding of “why” certain solution methods were appropriate for certain problems. Thus, while they were actively developing disciplinary understandings that enabled them to exercise disciplinary agency in the classroom, they had fewer opportunities to develop conceptual understandings and exercise conceptual agency. This was due to the narrow distribution of authority in the classroom. Unlike the majority of the students in the class, Kim possessed some level of authority, not within the larger classroom community, but among her peers. Here, we are able to hear directly from Kim, who came to be recognized as a legitimate data science-doer, that approaching peer-to-peer help and collaboration as an exercise in articulating ideas and providing explanations helped her develop deeper data-scientific understandings—even for someone who had previously taken the course and was already familiar with data-scientific concepts imparted in the course. She also critiques traditional teaching approaches that position students as receivers rather than co-creators of knowledge, and argues that such practices condition students to accept disciplinary assertions rather than develop deeper conceptual understandings (Cobb et al., 2009).

On several occasions Kim demonstrated that she understood *why* a particular solution method was appropriate for solving a given problem as well as the relationship between skills and concepts. For example, when students were asked to analyze two data distributions for surface streets and freeway commuting times to determine which commuting method was best, I asked Kim to explain what she needed to know in order to make a determination:

RESEARCHER: What do you want to see? …How will you determine the best method? What will you be looking for?

KIM: We're looking for the shortest average commute time, whether it be on the freeway or surface streets…whichever has the shortest commute time will be the best way to get to school.

RESEARCHER: Is there a measure that would tell you which is the shortest commute time?

KIM: The number that we get for the measure of center which gives you the average… [looks through the worksheet] …for surface streets, the typical was 33 minutes but for the freeway, the typical was 35 minutes so you would think that surface streets would be a quicker way to get to school.

RESEARCHER: Okay. What about measures of spread? Would that influence…

KIM: Yes, because [the] typical just tells you what the average is, but the measure of spread means—it's the possibility of you getting [to your destination in] about this [much] time, more or less. For example, if the measure of spread was two minutes then you can get to school in a matter of two more or less minutes than what the average is and that can really affect the data itself…. For example, it takes 35 minutes to get to school if you take the freeway, but the measure of spread is one minute, then you would get to school between 34 minutes or 36 minutes, but if the measure of center for streets is 33 minutes but the measure of spread is 10 minutes, then there's a possibility that you can get to school in 43 minutes or 23. So the [spread] can really affect [your commute time].

While Kim does not specify which measures of center and spread she used during this particular exchange, she told me earlier during this lesson that her group calculated the median of the distribution as a measure of center and the IQR as a measure of spread because the distribution was skewed. She added that using these measures enabled them to more accurately identify the balancing point and variability in a skewed data set. In the excerpt above, she conveys an understanding of what measures of center and spread actually measure and what this means in the real world (lines 7-8). Even though this assignment was couched in the real-world context of commuting to school, Kim extended it with examples provided in lines 16-21 and demonstrated that she understood how one's typical commute time might be affected given low or high variability. By doing so, Kim indicated that she not only understood how to determine appropriate solution methods and why they are appropriate for solving particular problems; she also indicated that she understood the meaning of measures of center and spread in the context of commuting times.

Data-scientific reasoning in real-world context. In the IDS classroom, Kim demonstrated her desire to use data-scientific skills and understandings to address problems that she believed affected people in the real world. In the excerpt below, we see Kim use a lab assignment and her disciplinary skillsets to shed light on issues of unemployment and homelessness among American military veterans. Our exchange took place while students were in the computer lab setting working on a practicum which asked them to develop and answer their own statistical question using data from one of their own data collection campaigns. When introducing the assignment, Ms. Gellar encouraged students to come up with a question they found interesting. I approached Kim and her groupmate, Diana, to ask if they had decided on a question. Our exchange was as follows:

RESEARCHER: Diana, have you all figured out a question?
DIANA: Ummm—sort of and kind of. We're going to figure out the average of [*looks and points to Kim's computer monitor. Struggles to articulate their question*]—
KIM: The average employment status of veterans and non-veterans.
RESEARCHER: Veterans and non-veterans?
KIM: Yes [*nods in affirmation*].
RESEARCHER: What data set are you using?
KIM: American Time-Use Survey…
RESEARCHER: So, why did you choose to use the time-use and not the—why did you choose this data set?
KIM: This one has more variability because it's not just on this class.
RESEARCHER More variability in terms of what?
KIM: In terms of the number—the total number of entries, the number of variables…
RESEARCHER: So, I see it has—wow—more than 12,000 entries. Okay, so what is your question? Do you know your question?
KIM: "What is the average employment status of a veteran compared to a non-veteran?"
RESEARCHER: Is that [the question]? —Diana, you had said that you had "kind of" figured it out…
DIANA: Mhm [*nods and points to Kim*].
RESEARCHER: Okay, so why do you find that interesting? Well—what led you to ask that question?
KIM: There's a lot of problems going on about how people are saying that we need to help our veterans out because they're unemployed or they're living on the street. But right now, I was looking at the first couple hundred entries and most of the veterans, if they were unemployed they were married or if they were not married they were employed. So, it

didn't seem like a problem in the first couple hundred entries, but I don't know. There's 12,000—I'm not going to look through every single one so I'm probably just going to make a graph to see that.

RESEARCHER: That sounds really interesting. I'm excited to see what your group comes up with.

KIM: Thank you.

Kim approached the assignment as an opportunity to address something that she felt was a real issue in society, and thus, sought to use her data-scientific understandings and technical skills in RStudio to shed light on an issue that she deemed relevant to her world (lines 21-27). In effect, Kim was thinking about the data at hand within the real-world context of unemployment and homelessness. Furthermore, her question also sought to address issues related to justice, opportunity, and equity for military veterans who return to bleak outlooks once reintroduced to civilian life (lines 21-22). The questions posed by other students also involved thinking about data in real-world contexts, but the extent to which they dealt with real-world issues was limited to mentions of variables (ex., calories, sweet or salty flavor profile, cost) and names of snacks that exist in the real-world. Additionally, the majority of students approached the practicum as they did other RStudio lab assignments and prioritized task completion rather than deeper-level engagement and meaning-making with data. By using the practicum as an opportunity to shed light on social issues, Kim demonstrated her inclination to use data science to develop understandings that could potentially help others, or as she put it in our interview, "make a difference in the world." She also used the practicum as an opportunity to examine whether the data supported a perceived social phenomenon (i.e., "people are saying" in line 21).

Given Kim's academic endeavors and passion for science and mathematics, this assignment provided an opportunity for her to bring her out-of-school understandings into the IDS classroom for further analysis, allowing her to engage in data science-doing in ways that supported, enriched, or elucidated her understandings of the world beyond the classroom. Thus, Kim was able to engage in a personally enriching process of meaning-making with and about data. Despite the affordances of working with the American Time-Use Survey data, including the fact that it likely contained more variability than the smaller in-house data sets generated by students in the class (lines 9-13), Kim soon realized that she was using the wrong data set, as the assignment clearly specified that students were to pick from two data sets collected during students' participation in the Stress/Chill campaign and the Food Habits campaign. When I revisited Kim and Diana, our exchange unfolded out as follows:

RESEARCHER: And how's it going over here?
KIM: We had to change the question because it was the wrong data [set] [*laughs nervously*].

RESEARCHER: What data were you using?
DIANA: Outside source
KIM: Yes, an outside source.
RESEARCHER: And you were supposed to use the data for your class?
KIM: Yes. So, we changed the question to, "Do healthier foods cost more than unhealthy foods?"

RESEARCHER: So why did you decide to switch your—I mean you still have [the student] time-use [data set], why did you decide to go with food habits [data set]?
KIM: Well, we were trying to find something that could connect to something…in our society, I guess you could say. And for this question, it is somewhat a problem because people are saying that we're not healthy because we're poor, but [that's] because healthy foods cost more. So, we asked a question that could potentially prove or disprove that—

DIANA: We're trying to show the different healthy levels of different types of snacks and how much they cost. So, depending on the type of snacks and their costs is depending on the types of healthy levels that they have…we're trying to compare whether or not the healthy snacks are more expensive or the unhealthy snacks are either more expensive or cheaper…

RESEARCHER: What were your thoughts before—what was your hypothesis? Kim, you were saying you were trying to connect it to…
KIM: Something in society.
RESEARCHER: Yes, so what is that—you said that people think or people say—
DIANA: She had thought that…the healthier the snacks, the more expensive they are and that the—m

RESEARCHER: Do you see that in society? Is that something you see?
DIANA: Yes [nods unhesitatingly].
RESEARCHER: Where do you see that?
DIANA: A salad will be like $7 when you can buy a Big Mac for three bucks—three-four bucks.
RESEARCHER: Or a cheeseburger for a dollar.
DIANA: Or a cheeseburger for a dollar, exactly.
KIM: It's even simpler than that. A Gatorade can cost like 69 cents and a water [bottle] can be like $1.20, $1.50.

Evident in this second exchange is that Kim, again, approached the practicum as an opportunity to address an issue that she perceived to be of relevance to society. In lines 13-14, she explicitly stated that the intent behind the question was precisely to ask a question that "could connect to something…in our society." Like her previous approach, Kim took an issue relevant to her everyday life and brought it into the classroom for further analysis. Regardless of whether the data supported or contradicted such claims (lines 6-7), using data in this capacity and for the purpose of understanding phenomena in her own world held the promise of enriching Kim's understandings of the world and issues of socioeconomic disparity within it through data science-doing.

In this second exchange, unlike the first, Kim inserted herself and her community into the process of meaning-making with data. She did so by arguing that claims that posit that "we're not healthy because we're poor" fail to acknowledge disparate access to affordable healthy foods (lines 14-16). This shows that Kim was critical of and sought to tackle the widely held misconception that low-income families consume fast and unhealthy food at disproportionate rates (Chandler, 2015; Vikraman, Fryar, & Ogden, 2015; Philip, Rocha, & Olivares-Pasillas, 2017). Additionally, in lines 32 and 35-36, both Diana and Kim provided examples drawn from their own experience to support Kim's claim that healthy food is more expensive and, thus, less accessible to those living in poverty. It should be noted, here, that Kim did not use her perspective as a definitive response to her statistical question, but instead used it to lay the contextual groundwork that informed the purpose of her data analysis and her motivation for engaging in it. This I surmise according to lines 16-17 where she indicates that data analysis will help her "prove or disprove" both what "people are saying" and what she understands as a contributing factor to unhealthy eating for those living in poverty. Moreover, by saying, "we asked a question that could potentially prove or disprove that," Kim conveyed that her question was not specifically designed to address the claims she presented, but acknowledged that her findings could potentially shed light on the matter of concern by allowing her to infer meaning from her data analysis and resultant outputs. This is a significant finding because it demonstrates that Kim's approach to data science-doing as it pertained to completing the practicum served as an exercise in inferential thinking. What's more, her approach to completing the assignment imbued the exercise with personal signification, making it consequential for life beyond the classroom.

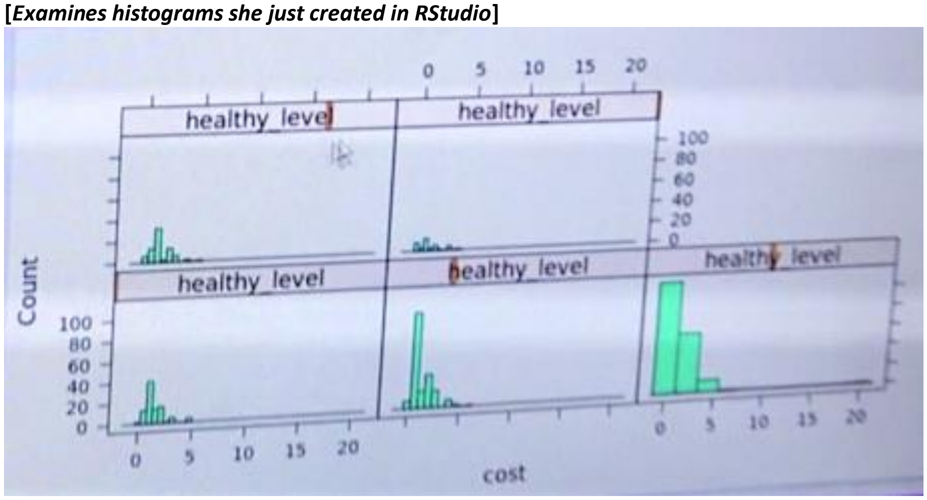
This exercise provided unique affordances to Kim that routine assignments and exercises did not. For example, students had an entire class period, lasting 90 minutes, to come up with a question, analyze the data, and generate graphical representations to be included in the completed assignment. The exercise was open in nature to the extent that students were provided the time and flexibility to make their own decisions with regards to what questions to ask and what methods to use in order to answer their question. Given, they had to share their question with Ms. Gellar before carrying out the practicum in order for her to approve it, thus, questions had to conform to established standards for what constituted an appropriate statistical question. Still, while most students settled on questions on the basis of ease and feasibility of completion, Kim's question primarily sought to address real-world issues.

Kim's ability to further develop her budding CSDS identity was first hampered by the assignment's specification that students use one of their in-class data collection campaigns. Presumably, she had yet to share her question with Ms. Gellar for approval when she first told me that she was interested in looking at the relationship between veterans and issues of homelessness and unemployment. While Kim was able to come up with a new question that was still socially-relevant and mounted a critique of social disparities, her decision to use the more expansive American Time-Use Survey data set because it contained more variability went unacknowledged by the larger classroom community even though Ms. Gellar explicitly told students on multiple occasions that variability was something they should look for in a data set.

Furthermore, Ms. Gellar told students that they could choose from one of two in-house data sets, but when I asked other students why they selected to work with the Food Habits data set, all of them expressed that the alternative was not truly an option as there were numerous missing entries. Thus, while students were outwardly presented with two options, they realistically only had one. Therefore, even though students were encouraged to exercise agency by selecting a data set and carrying out their own mini-study by designing a statistical question, their opportunities to exercise agency were framed within established bounds for data science-doing as determined by the course.

Thinking critically about aspects of data collection and analysis. The openness of this practicum, coupled with Kim's efforts to understand what her class' snack data revealed about larger issues relating to poverty, unhealthy eating, and affordability of healthy foods, led Kim down a path where she was compelled to turn a critical eye to aspects of data collection. I present the following excerpt to show how Kim's criticality toward aspects of data collection began to develop once she produced data outputs to answer her statistical question:

KIM: [***Examines histograms she just created in RStudio***]



RESEARCHER: Can you tell me what you're looking at? —or what I'm looking at?

MS. GELLAR: [***Announces that students have 10 minutes left to work on the practicum. She extends the due date for the assignment from today, Friday, to Monday afternoon.***]

RESEARCHER: What do you read? When you look at that [visualization] what do you see?

KIM: [***Actively thinking through the exercise***] **Each one of these [boxes] is a healthy level, but**

instead of showing you the number…it gives you this [orange] line—depending where it is, like, this is [healthy level] one [*points to box on the bottom left*], this is two, this is three, this is four, and five.

RESEARCHER: Ohhh, okay, I see.

KIM: These numbers [on the vertical axes] show how many people input that much data for each [healthy level], but I'm not sure about cost, because I don't think anything cost $20 or $5, so there's no point to putting this range here, making each one of these [prices] harder to see…but I can see that nothing cost more than five. [*Points to boxes for different healthy levels*] this hardly goes past five, this doesn't go past five, doesn't go past five. A little bit here [*points to shortest column in the histogram at the bottom right*] but that's just part of the graph.

RESEARCHER: Does this graphic support your hypothesis?

KIM: [*Pensive*]

RESEARCHER: What do you think? What do you ladies think, do these histograms support what you thought you'd find?

DIANA: I kind of think it's within the middle…

KIM: [*To Diana*] Well, I think you say that because [the histogram for healthy level 3] is taller, but that just means that more people [input entries for healthy level 3]. What we really need to be looking here is for the range but the thing is that here, in the third healthy level, somebody paid 20 bucks for something… [*looks at the histograms on her monitor pensively*].

Based on my interaction with Kim, she grew pensive and conflicted when she realized that the histograms did not necessarily support a hypothesis premised on her own lived reality. The histograms indicated that snacks purchased by students for all healthy levels were within the $0-$5 price range (lines 14-15) and that students entered more data for snacks with healthy levels 2 and 3 compared to snacks with healthy levels 1, 4, and 5 (lines 2, and 25-28). In an effort to reconcile what she expected the data to show and what it actually showed, Kim began to ponder the legitimacy of the data itself, stating "I'm not sure about the cost because I don't think anything cost $20 or $5." She then went on to comment on how the presence of a $20 outlier affected the rendered visualization and impaired her ability to read and interpret the histograms (lines 15-16). Our exchange came to an end as the class was almost over and Kim continued to think about the meaning of the histograms. Evidently, if all snacks were within the same price range on average and for all healthy levels, this meant that the data did not support what both Diana and Kim experienced in their daily lives. I discern, based on Kim's pensiveness and the contentions she raised regarding the legitimacy of entries for the "cost" variable, that she did not genuinely believe in the accuracy of the data before her. I am not arguing that the data was inaccurate because it failed to support Kim's out-of-school understandings, nor do I hold that the data somehow invalidated what Kim believed to be true, but I believe that Kim's growing skepticism, given more time and probing questioning by other members of her classroom community including her peers and Ms. Gellar, could have led to a powerful awareness of how the process of data collection can elucidate and simultaneously obscure certain aspects of the phenomenon under measure. Here, I am specifically referring to the fact that the "healthy level" of snacks was arbitrary.

During the practicum, students were reminded that healthy level 1 meant that snacks were "very healthy" and healthy level 5 meant snacks were "not very healthy." Thus, when inputting snack data, students were tasked to rate, on an individual basis, the level of healthiness for each snack entered throughout the day during a given period of data collection. Therefore, there was no particular measure or criteria outlined in the curriculum to help students establish a healthy level for snacks. This means that while one students could have determined that a snack bar that contained granola was very healthy (level 1), another student could have rated the same snack as not very healthy (level 5) if they felt that the presence of granola alone did not signify a healthy snack given that many of these bars also contain lots of sweeteners, sodium, simple carbohydrates, and fats.

Moreover, while Kim asked a question that she hoped would shed light on real issues relating to healthy eating and food costs, the contents of the data set were not representative of students' actual "Food Habits" as the name of the data set suggested. During the first year of implementation of the IDS curriculum, this data collection campaign was referred to as "Snack Habits" as it specifically sought to measure students' consumption of snacks. For the second year, the name was changed to "Food Habits," however, students were explicitly told to collect only data on snacks and to exclude full meals. This means that the "Food Habits" data set did not contain the type of data necessary to address Kim's question, "Do healthier foods cost more than unhealthy foods?"

Moreover, while Kim could feasibly have answered her question by overlooking the fact that the data set only included data for snacks and not food in general, doing so would have undermined the foundational motivation that gave rise to her question in the first place. I believe Kim was ultimately likely to prioritize task completion regardless of her criticality toward the data given the following perspective shared during our interview where she alluded to the negative consequences of not completing assignments, "This year I'm more of like 'You've got to finish [the assignment], you've got to finish it,' because I know what's going to happen if I don't." By treating the Food Habits data set as static and legitimate, students became responsible for ensuring that their statistical questions conformed to the contents of the data and were, thus, answerable. Kim's desire, and ultimate inability, to shed light on social concerns within the bounds of dominant classroom expectations and obligations provides a nuanced account of how narrowly defined tasks and criteria for doing data science upheld epistemological exclusion within an instantiation of STEM reform and hampered her ability to continue to think richly and critically with and about data. While I do not believe that this experience ultimately hurt Kim's motivation to think data-scientifically in her everyday life, her desire to use science and math to help society, and her aspiration to pursue an education and career in STEM, it is important to recognize that it did not validate and support the continued development of her data-scientific criticality and the accompanying potential to yield new insights and establish conceptual and sophisticated data-scientific understandings. And so, even though Kim gained experience learning to work with data, she was not afforded the opportunity and authority to reflect on and transform how she and others were thinking of data as static and true. Kim's conceptual data-scientific understandings enabled her to discern that something was not making sense, but without questioning the validity of the healthy level variable and reflecting on what the data actually represented ("snacks" as opposed to "food"), she was unable to arrive at a clear understanding of the data and of her question.

The excerpts discussed above indicate that Kim began to develop a CSDS identity within the bounds of acceptable data science-doing as determined by the classroom community and as imparted by classroom obligations. This means, however, that her developing CSDS identity was constricted by the bounds of acceptable data science-doing as it played out in the classroom and that students had little say in the transformation of existing data-scientific understandings and the creation of new knowledge. Armando's developing CSDS identity, on the other hand, occurred wholly outside of what was constituted as legitimate data science-doing. Indeed, his process of meaning-making with data took place in opposition to dominant expectations imparted by classroom obligations. In the following section, I will discuss opportunities and mechanisms that gave rise to and simultaneously limited Armando's development of a CSDS identity.

Introducing Armando

When I interviewed Armando, he had been identified by Ms. Gellar as one of the lower achieving students in her IDS classroom. I found this odd because throughout my classroom observations and interactions with Armando, he demonstrated active participation in the classroom and engaged interest in understanding data-scientific skills and concepts. Our interview revealed that Armando had a passion for technology and the medical field and was extremely motivated to pursue a future in either computer science or pediatrics. Armando's interest in pediatrics emerged when he began attending MSHS as the school encouraged students to pursue careers in the medical field. He expressed that his desire to care for children stemmed from his belief that “kids are the future of the world” and that caring for them made him feel like he was making a contribution to larger society. Due to the fact that MSHS is a medical-themed high school, students had the opportunity to participate in internships with local healthcare providers, an opportunity that proved influential for Armando:

> I have the internship at St. Juliana Medical Center here in [the local city of] Stanwood and I have been getting a lot of insight into what the medical field is. Sometimes, they would allow me to work with the children there. That is what made me want to become a pediatrician…. Because I have the most hours working there as a volunteer and I have a better understanding than the others, I am the volunteer leader.

Armando’s desire to help those around him, particularly children, could be attributed to his position as the oldest of three children and his role as a resource to his parents for household budgeting and informed everyday decision-making. For example, when I asked him how he might use the skills he learned in IDS in his everyday life, he shared the following:

> [Say you] want to go buy groceries [and] you're on a budget. You might want to go online to each store that may be nearby and then check their pricing on the item that you want to get, to check which one would be less expensive… Also, calculating the distance it takes from your house to the store [is important]. Gas—calculating how much gas you're going to be using up to get to that location. [Asking,] “Is it better to go to a further location that is less expensive than to go to a closer one that is more expensive?” ... That example [is] one that I deal with most often with my parents. They want to shop somewhere because it's less expensive than a place that is closer, but if you think about it, you're not only going to spend…money [on groceries]. There'll be your car, the usage of

your car, the gas, your tires, your engine [to consider]—so all of that goes into play. If you think about it, you're either spending the same or a lot more.

This excerpt is a testament to Armando's important role as a resource within his family and contextualizes an understanding of his desire to help those around him as he continues on his educational and career trajectories. Furthermore, our interview revealed that his parents expressed admiration for Armando's resourcefulness and problem-solving skills, particularly when it came to technology. In fact, even though Armando described becoming a pediatrician as his career "Plan A," he shared that before attending MSHS he ambitioned to become a computer programmer. His fascination with and passion for all-things-technology began when he was seven years old when his parents bought him his first desktop computer. Armando recounted how encountering problems with his computer helped establish a deep-seated and lasting curiosity for understanding the way technology works:

I was about seven years old when I had my first desktop [computer]. That desktop…would constantly have problems whether it would be inside the computer or in the software… My parents didn't know much about technology—at that time, I did not either, but it was easier for me to understand more of what was happening. Because of that and also because we didn't have the money to go [get the computer repaired] every single time there's something wrong with it… For example, the power bank of the computer, it would fry out. The fuse would just crack in the middle so you either had to replace the whole unit or you would fix the fuse. Most of the times, it's harder to fix the fuse so you just have to buy another one. Like I said, we didn't have the money to go buy parts or get it repaired. I actually had to go in there, open it up, see how it works. [I would] go on the computer and find out what would be wrong with it—what is the cause [of the problem]. From there[on], I would try to fix it.

In the excerpt above, Armando speaks about the circumstances that set the stage for his technological curiosity. He alludes to generational differences in familiarity with technology that enabled him to begin working with computers at the age of seven when he notes that even though neither he nor his parents knew much about computers, it was "easier" for him to understand his computer's functionality and the issues that impaired it (lines 2-4). In lines 4-6 he acknowledges that another major catalyst for attempting to fix computer hardware and software issues was the fact that although his parents managed to buy him a computer, they could not afford to pay for repairs or to buy new computer parts. Thus, socioeconomic circumstances in his home played a critical role in what eventually became Armando's tendency to tinker with technology as a means of understanding how it works. When I asked him how his parents felt about his taking apart a computer they, presumably, struggled to buy he said his parents were "happy" and "proud" of his ability to do so. They found this fascinating because technology was something new to them and they viewed computers as a possible profession for their son. And so, Armando's parents were encouraging of his tinkering and curiosity about technology.

More opportunities for Armando to tinker with computers arose when his father developed a friendship with a computer repair shop owner. The shop owner invited Armando to visit his shop and offered to teach him how to troubleshoot and otherwise fix computers:

> In that shop, was the first time that I built a computer with the spare parts that he had laying around. I built a computer and I programmed it. Once I [had] the operating system going, I had to download all these drivers. Sometimes, because we didn't have the internet drive, I would have to work through it to try to get that driver from the system.

Thanks to Armando and his parents' resourcefulness he was able to gain experiential knowledge of computers that proved profoundly meaningful. He said he spent Summer and Winter breaks learning from the shop owner and experimenting with spare computer parts to the extent that he eventually began to repair computers brought in by customers—what he referred to as a true "test" of his knowledge and understanding. These experiences helped shape how Armando approached learning about and with computer-technologies by supporting and rewarding his genuine desire to understand *how* computers and other form so technology work.

The out-of-school experiences that Armando brought into the IDS classroom shaped his participation and the extent to which he came to observe and value classroom obligations. I argue that while Armando, for the most part, cooperated in the constitution and observance of classroom obligations, less-structured activities and assignments, coupled with his preexisting passion for technology and curiosity for how things work, enabled him to begin to develop a CSDS identity in Ms. Gellar IDS classroom. Unlike Kim's developing CSDS identity, Armando's developing CSDS identity occurred outside of narrowly defined ways of data science-doing.

Conceptual reasoning about relationships between data science skills and concepts. Before Armando stepped into the IDS classroom for the first time, he had already gained valuable out-of-school experience tinkering with computers and was keen on understanding how things worked. During our interview, Armando spoke briefly about his experience in a computer science class where students were tasked with animating a cat on a website. He said, "we would take a code and move it into the console and press 'Begin' and the cat would move, but doing that didn't help us understand the code." He added that while the teacher clearly specified what they could do with codes, Armando challenged the teacher's narrow delineations of coding because in his view, they did not foster deeper understandings:

> I would go more [in-depth and tell my teacher,] "With codes you can do other things not just this. Right here, you're not actually learning the code or seeing the code. What [you are] doing is just telling it what to do."

While Armando did not offer the same criticism of his IDS class, he described RStudio lab assignments similarly when he said that before entering the IDS classroom for the first time he "thought [coding in IDS] would be easier for [him] but the coding [in RStudio was] a lot different than software development." He compared coding in RStudio with coding for software development by positing that coding in RStudio is used for information retrieval, whereas coding for software development is meant to program a computer by inputting algorithms to tell it how to operate.

Armando's desire to understand how things work was also evident in his participation in the classroom, namely in his challenges of Ms. Gellar's data-scientific assertions. These challenges emerged as Armando sought clarification of concepts or struggled to understand the conceptual reasoning that founded her data-scientific assertions. Classroom obligations held that authority was principally distributed to the teacher and to the RStudio lab assignments, thus, her data-scientific assertions were not often challenged, but out of the five challenging acts I observed during the second semester of IDS, three of these were initiated by Armando. His challenges to Ms. Gellar were never attempts to undermine her role as the authority figure in the classroom nor were they personal attacks, but were instead proffered as clarifying questions. Challenges should be interpreted as students' efforts to gain conceptual understandings of data-scientific skills and concepts.

For the remainder of this section, I will focus on an instance of Armando's challenge of Ms. Gellar's data-scientific assertions about chance and chance outcomes in the context of his efforts to deepen his understanding of data-scientific skills and concepts. This challenging-act took place when Ms. Gellar introduced a new theme—probability. During the lesson on probability, students were provided with an "enduring understanding," which described probability as "[a measure of] the long run frequency of occurrence for chance outcomes." Ms. Gellar asked students to think of synonyms for "chance" in their groups; synonyms included "possibility," "luck," "odds," "random," and "coincidence,"—all of which were legitimated as appropriate responses. After sharing their synonyms, Ms. Gellar asked students to explain what the statement "that just happened by chance" meant. Students offered the following responses:

Roselyn: **It was just a coincidence**
Jesus: If it happens again, it might not be the same outcome
Pheobe: **It was unplanned**
Gerardo: Unanticipated

Ms. Gellar, in turn, synthesized these responses by stating that if something happens by chance, that means "there was no intention in what happened—it just happened…nobody made a decision that that will happen." Given these collective understandings of the meaning of chance and chance outcomes, the class was asked to think of situations where outcomes could be attributed to chance. The following excerpt documents the exchange that led up to Armando's challenge of Ms. Gellar's assertions about the meaning of chance:

MS. GELLAR: **Let's see who can come up with the most creative [example].**
DIEGO: Gambling?
MS. GELLAR: **Gambling. Okay, can you be a little more specific? What kind of gambling?**
DIEGO: Slot machine.
MS. GELLAR: **Slot machine. Okay. [*Looks for new volunteer and nods at Roselyn*]**
ROSELYN: The gender of the baby.
MS. GELLAR: **Gender of a baby—ooh, I like that one! Nobody chooses what kind of child they have, right? —that's true.**
ARMANDO: Receiving a gift?
MS. GELLAR: **Receiving a gift [*ponders*]—somebody made a decision about your gift.**
ARMANDO: But you don't know it.
MS. GELLAR: **You don't know, but somebody made the decision, right? So, the gift was a decision.**
JENESSA: [*Inaudible*]
MS. GELLAR: **Jenessa's asking, "What do you think about horse racing?" Do you think that that [has a] chance [outcome]?**
STUDENTS: [*Several students say "Yes"*]
JESUS: **Unless it's rigged!**
MS. GELLAR: Yeah—you might not be able to tell, right, who's the winner—I just saw a couple of other hands—Carlo, did you have one?
CARLO: **[Playing] *lotería*?**
MS. GELLAR: Lotería…yeah because that's just based on whatever you get, right? [*Looks for new student*] Diana.
DIANA: **Bowling?—**
DIEGO: No, that's skill.
MS. GELLAR: **Yeah [*agrees with Diego*]—that's based on something. Yeah. [*Points to Jesus*]**
JESUS: A raffle?
MS. GELLAR: **A raffle—okay. Yeah. Any others? Mervin.**
MERVIN: Let's say you're holding a gun and what if you get a misfire?
MS. GELLAR: **Say you're holding a gun and there's a misfire? Uh…you don't know that's going to happen, I guess.**
ARMANDO: [*Raises hand after Ms. Gellar agrees that a misfire is a chance outcome*]
MS. GELLAR: **[*To all students*] Wait, we're having a discussion so you should be listening to the person that's talking and at the moment that's Armando.**
ARMANDO: As long as you can manipulate the situation is it chance?

MS. GELLAR: **No.**

ARMANDO: No?

MS. GELLAR: **So as long as the situation can be manipulated in some way it is not chance.**

ARMANDO: So, like he said, the gun…he didn't decide to pull the trigger.

MS. GELLAR: **Yeah, so, there is an intention in whether I'm firing the gun or not, but I think he's referring to the fact that it's a misfire—that's unintentional. So, that's a good point, right? The fact that he's firing a gun is not by chance. But the fact that there's a misfire was the chance [outcome]—well, actually last year [in IDS], I don't know if you [Diego and Kim] remember this, but we talked about Russian Roulette, you guys know what that is? Yeah—that's similar. Somebody's choosing to fire a gun but you don't know whether the bullet is going to come out or not.**

Of the nine proposed examples of situations where an outcome might be due to chance (including Ms. Gellar's example of Russian Roulette), Armando's and Diana's were deemed as not consisting of outcomes that were due to chance. Below (Table 6.1), I offer a breakdown of these proposed examples in an attempt to understand the logic behind Armando's challenge to the assertion that receiving a surprise gift is not a chance outcome.

Table 6.1

Situation where outcome is due to chance	**Outcome**	**Was outcome intended by primary person affected?**	**Was example legitimated?**	Reason for legitimation or delegitimation
DIEGO: Slot machine gambling	Symbol combination	No	Yes	Player does not manipulate outcome
ROSELYN: Sex of a baby	Specific sex	No	Yes	Parents do not select gender
ARMANDO: Receiving a gift unexpectedly	Gift	No	No	Someone intended to deliver gift
JENESSA: Horse racing	Winning horse	No	Yes	Winner is not specifically intended
CARLO: Lotería	Cards drawn	No	Yes	Cards are shuffled/not intended
DIANA: Bowling	[Unclear]	[Unclear]	No	Game of skill not chance
JESUS: A raffle	Winning ticket	No	Yes	Winner is not specifically intended
IRVING: Holding a gun, accidental misfire	Misfire	No	Yes	Misfire is not intended
MS. GELLAR: Russian Roulette	Bullet fired	No	Yes	Cylinder is spun, chance of gunshot unknown

Given established understandings of the meaning of "chance," I argue that Armando initially believed that his example was appropriate because the individual primarily affected by the outcome did not orchestrate the occurrence of the specific outcome (line 11 in the excerpt above). In this regard, his example of unexpectedly receiving a gift qualified as a situation where the outcome was due to chance. To put it another way, unexpectedly receiving a gift from someone would qualify as a random or coincidental occurrence. However, Ms. Gellar argued that because someone intended to deliver the gift, the outcome was not due to chance (lines 12-13). While Ms. Gellar's reason for invalidating Armando's proposed example might make sense in a conventional statistics course, it essentially closed off engaging in rich and nuanced whole-group discussion about a data-scientific concept within the reform-based and equity-oriented IDS classroom. Moreover, ensuing examples blurred established definitions for chance outcomes as the rubric used to judge his example was not consistently applied to others. This also occurred with Diana's example of bowling in line 24. It was unclear to me whether she was referring to winning a bowling match, hitting a pin, or rolling a strike and Diana did not have a chance to elaborate as her response was immediately dismissed by Diego as a matter of skill (line 25), an assertion that was supported by Ms. Gellar (line 26). It is worth noting that even though horse racing can also be thought of as a game of skill for both the jockey and the racehorse, when Jenessa proposed this example (lines 14-15) it was approved as involving a chance outcome (lines 17 and 19-20). Where Diana did not challenge the dismissal of her example given the approval of a peer's example that in ways paralleled hers, Armando challenged the dismissal of his when Mervin's example of a misfire was accepted.

Mervin's proposed example in lines 30-31 was premised on the idea of someone "holding a gun" and unexpectedly experiencing a misfire. Ms. Gellar mistakenly thought he was talking about *firing* a gun and subsequently experiencing a misfire (line 42). In Armando's view, Ms. Gellar's dismissal of his example as not consisting of a chance outcome did not make sense given her assertion that although the person directly affected by the misfire in Mervin's example intended to shoot but did not intend to experience a misfire. I surmise that based on Armando's professed personal curiosity and years of tinkering with computer parts and mobile technologies to understand the logic behind their functioning, his constant reasoning about how things work outside of the classroom translated to reasoning within the classroom about what qualified as a chance outcome. Thus, Armando believed that both his and Mervin's examples involved unanticipated outcomes for the individual directly affected preceded by intent by that individual or someone else to, in Mervin's case, fire a gun or deliver a gift, in Armando's case.

I present this example, not to make a definitive statement about chance outcomes nor to propose re-determinations for proposed examples, but instead to demonstrate how Armando's desire to understand how things work played out in the classroom setting. At first, he tried to understand the logic that defined a chance outcome, followed by his proposal of an example that was consistent with that logic. His challenge of Ms. Gellar's data-scientific assertion represented a manifestation of his efforts to understand the data-scientific notions of probability and chance. However, because his attempts to develop conceptual understandings were positioned as challenges to how authority was distributed in the classroom, and therefore occurred outside and in opposition to narrow delineations of data science-doing, opportunities for him and his classmates to develop conceptual understandings about data-scientific concepts, in this instance of chance and probability, during whole group discussion were hampered. Nevertheless, Armando's attempts to develop conceptual understandings about data-scientific skills and concepts persisted because they were an extension of his out-of-school tinkering, problem-solving, and sense-making practices.

This exchange illustrates the importance and value of recognizing students as capable contributors to the development of new ways of thinking and to the conceptualization of traditional statistical concepts within data science. Necessarily, rash judgements of student responses and sense-making practices allow little room for the development of new knowledge in the classroom and impair both students' abilities to make meaningful disciplinary connections and the teacher's ability to truly understand the non-traditional epistemological contributions and practices that students bring into the classroom. I believe that addressing this shortcoming is a matter of supporting the development of teachers' interpretive power—that is, their ability to avoid misinterpreting student sense-making and meaning-making processes "when a discourse practice does not conform to a teacher's expectations of what an explanation or argument looks and sounds like" (Rosebery et al., 2016, p. 1574). Supporting teachers in their constant development of interpretive power in the STEM classroom is about acknowledging and embracing the ways of knowing of non-dominant students. These are processes that can be easily overlooked or dismissed if measured against "historically privileged ways of knowing, talking, seeing, and acting shaped by European American practices and values" (Rosebery et al., 2016, p. 1574; Warren & Rosebery, 2011; Bang et al., 2013; Seiler, 2013). Thus, when it comes to addressing issues of equity and access to STEM fields for students from non-dominant groups, it is necessary to consider the differential ways in which students might be making sense of new concepts and how established disciplinary understandings might take on new signification given non-traditional epistemological orientations of students.

Data scientific reasoning in real-world context. On several occasions during our interview, Armando spoke about the general usefulness of data-scientific skills for everyday problem-solving and informed decision-making. In my introduction of Armando above, I mentioned that he often helped his parents make important decisions and problem-solve in the home. Due to the fact that he saw the versatility and usefulness of data-scientific skills for everyday decision-making and problem-solving in the real-world, he felt that his peers, too, should care about data science:

> Yes, they should because it really does help. It helps you get more organized with things. There's many obstacles that we face every day, so [data science] can help, maybe not solve the problem, but it will make it easier for you to understand the problem, and then help you solve it.

He went on to say that his family and community should also care about developing data-scientific understandings because of their importance and usefulness for informed decision-making, personal organization, and problem-solving. Armando spoke of data-scientific reasoning as a versatile and widely applicable way of thinking critically about problems and everyday situations. Indeed, he relied on this form of reasoning as he participated in the IDS classroom. In the next section, I will discuss how Armando's approach to reasoning for problem-solving and decision-making led to thinking critically about aspects of data collection and analysis.

Thinking critically about aspects of data collection and analysis. What ultimately emerged as Armando's criticality towards aspects of data collection and analysis, and, thus, as part of a developing CSDS identity, began, like challenges to Ms. Gellar's data-scientific assertions about chance outcomes, as attempts to develop conceptual understandings of data-scientific skills and concepts.

A key moment in Armando's developing CSDS identity in this respect occurred in the computer lab setting. Students were given an opportunity to analyze data from the Titanic, which included variables for passenger names, survival, sex, class, and fares paid. In the classroom lesson that introduced the Titanic, students were told that fare rates were indicative of wealth as those that paid higher fares were wealthier than those that paid lower fares (see Figure 6.1). When Ms. Gellar introduced the Titanic lesson, she asked students to share what they knew about the sinking of the ship and explicitly told them that she was not asking about the film portrayal but rather about the actual historical disaster. Some students still responded with “Jack died” and other recollections from the film. She reminded students that they were going to be looking at actual passenger data from the Titanic and that there was a widely held belief that the wealth of passengers played a role in their survival, and so they would be using fare price as a proxy for wealth to determine if survival was influenced by wealth through statistical analysis using RStudio.

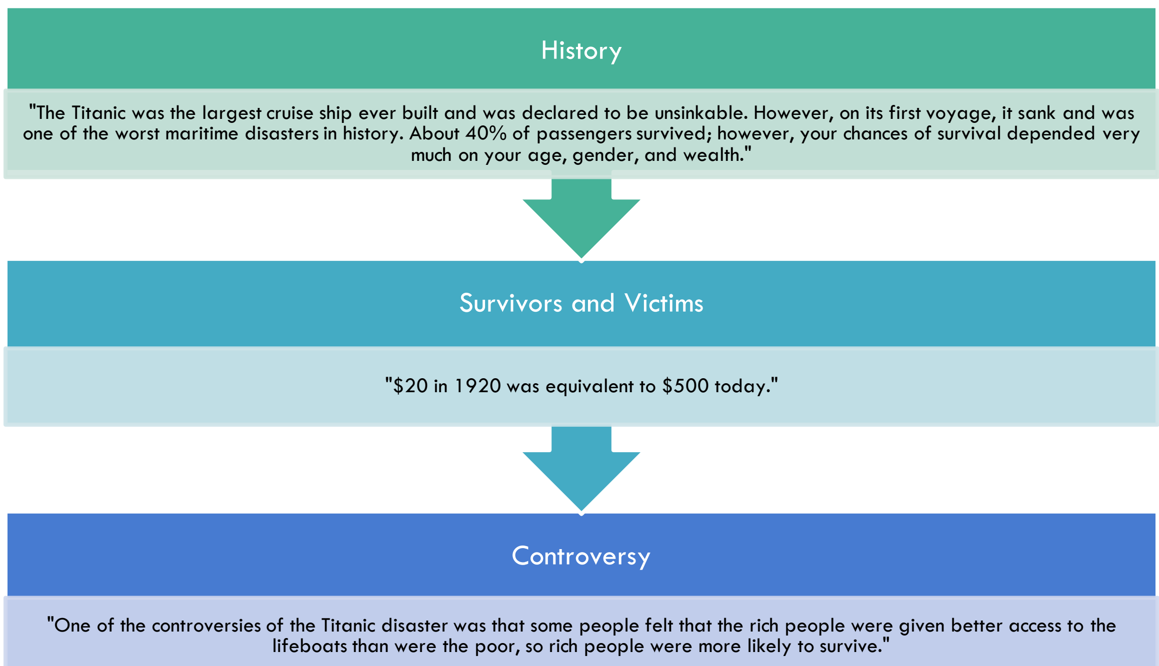
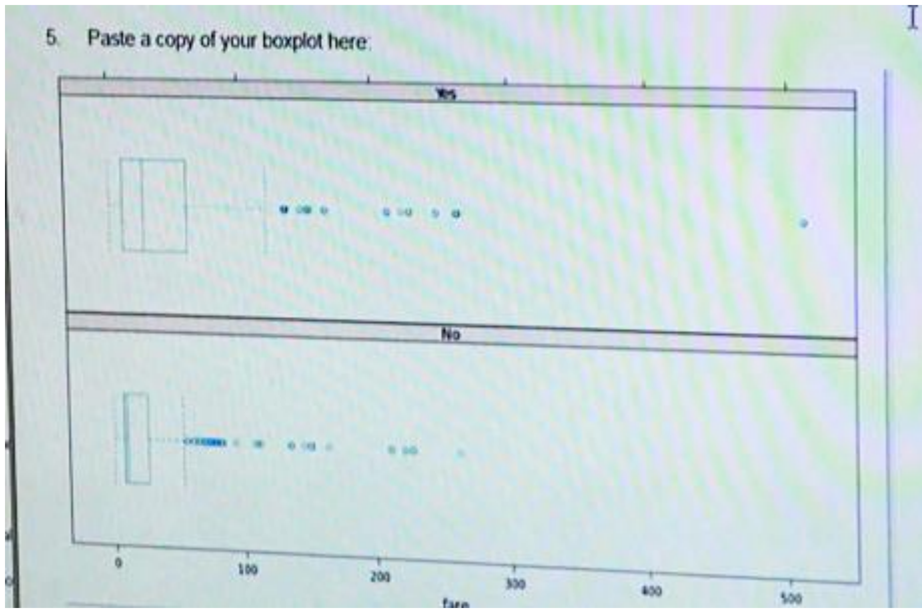


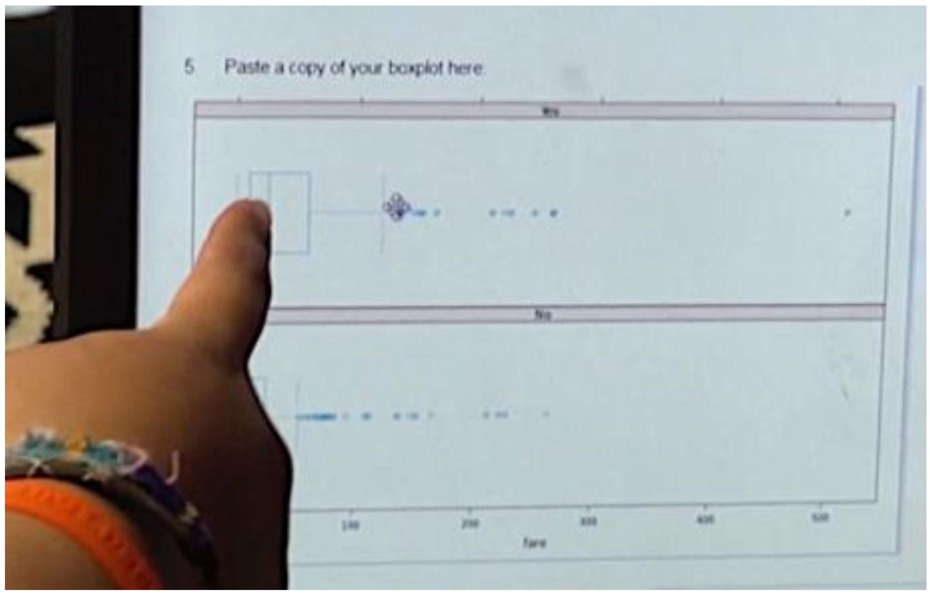
Figure 6.1 Lesson items projected in the classroom setting prior to student analysis of Titanic data set in the lab setting.

The excerpts I am about to share are part of one extended interaction, involving Amber, Katie, Armando, and myself, that took place while students were working through the Titanic Shuffle Lab where they were asked to analyze the data to determine if wealth (fares paid) played a role in survival. I highlight this exchange because it captures Armando’s growing interest in understanding what the data reveal, which led to his questioning of the meaning of variables in the data set. Our exchange began when I approached Amber, Katie, and Armando’s group and asked if they could interpret the meaning of a boxplot they produced using the Titanic data set (see line 4 below) to answer the lab question, “Based on the plot, do you believe richer passengers were more likely to survive? Explain why and describe how certain you are of being correct.”

Line 1	**AMBER:**	**The boxplot [we generated] it doesn't even show anything about it being a rich person or a**
2		**poor person. It just shows a boxplot and who basically survived or not.**
3		
4		
5	RESEARCHER:	So, what does the boxplot tell you then?
6	**AMBER:**	**It’s just basically the fares. People—how much they paid and then the passenger who**
7		**survived or not—whether they survived or not. That’s it.**
8	RESEARCHER:	I’m trying to read your boxplot. So, the top one is “Yes”—[does] that “Yes” [refer to] people
9		who survived?
10	**AMBER:**	**Yes.**
11	RESEARCHER:	And the bottom is people who di—
12	**AMBER:**	**Didn’t survive, and how much they paid.**
13	RESEARCHER:	Can you help me understand how to read that boxplot?
14	**AMBER:**	**I don’t know how to read the boxplot, honestly.**

This excerpt is consistent with my finding that students felt compelled to complete the task at hand even in the absence of comprehension of the task or of data outputs that they themselves produced. Here, Amber exercised disciplinary agency by using established solution methods to render a boxplot but she lacked the conceptual understandings necessary to derive meaning from it. I invited her on multiple occasions to exercise conceptual agency by helping to highlight different aspects of the boxplot (lines 8-9 and 11) and then asking for her interpretation given her understanding of those aspects (line 13), but she resigned herself to her inability to interpret the boxplot in line 14. I then invited her groupmates, Katie and Armando, to provide their interpretations in lines 15-16 below:

RESEARCHER:	**[To Amber's groupmates, Katie and Armando] Do any of you know how to interpret that boxplot?**
KATIE:	[*Calls to Armando*] Armando?
RESEARCHER:	**Like, what is it saying?**
AMBER:	See, I don't know.
ARMANDO:	**"Yes"—what is that? People who survived?**
KATIE:	Yes, and then that's the fare [*points to numbers on the x-axis*].
RESEARCHER:	**Can anybody tell me what that…**
ARMANDO:	So, what this is saying is more people survived than people that died, but only 40% survived [*baffled*].
KATIE:	**Isn't this the median?—**



ARMANDO:	[*To the researcher*] No, this [*referring to median Katie pointed out*] is the median.
RESEARCHER:	**Okay.**
KATIE:	And then this is the IQR, Then those…—
ARMANDO:	**[That's] the range… I forgot how to read boxplots**

Amber's deference to Armando, who sat two seats over, is telling because it indicates that she held a high estimation and valuation of his data-scientific competency despite Ms. Gellar's identification of Armando as one of the low achieving students in the class. Thus, she turned to him, rather than to Katie who sat between Armando and Amber, for help answering my question. It quickly became evident to me that Armando had yet to interpret the box plot himself (line 20) and his initial interpretation demonstrated that he did not understand what it was conveying (lines 23-24). The boxplot indicated that the median fare paid by survivors was higher than the median fare paid by non-survivors. It also indicated that range of fares paid by survivors was greater that of non-survivors, which meant that among survivors of the Titanic were those who paid higher fares. Further, with Katie's help, Armando begins to recall the components of a boxplot and struggled to derive meaning from it until he, like Amber, resigned himself to his inability to interpret the boxplot in line 30. Had one of these students been able to interpret the boxplot as instructed to in previous lessons, they would have realized that unlike Amber's claim in lines 1-3, the boxplot contained all the information necessary to answer the question at hand. However, in light of the fact that no one seemed to understand what the boxplot conveyed, I asked the group how they were going to proceed with the lab assignment.

RESEARCHER: So how are you all going to determine whether wealth had anything to do with survival rates?
ARMANDO: I don't know. We're trying to figure that out because like Katie said, it didn't matter… who was rich or poor because they were all trying to take out women and children first.

RESEARCHER: Well, that's your hypothesis, right?—
ARMANDO: Or at least that's how the story was told.

In the absence of the data-scientific skills and conceptual understandings necessary for Armando to make sense of the data visualization he and his peers produced, Katie and Armando began to rely on knowledge derived outside the classroom to make a claim about survival on the Titanic. Armando alluded to the film portrayal of the Titanic when he said, "at least that's how the story was told" as a way of supporting the validity of his claim that rescue efforts sought to save women and children first and not the wealthy (line 37). Unable to draw meaning from actual data on passengers of the Titanic, we see the fictional portrayal take precedence in his and Katie's assertion of factors that influenced survival (lines 33-35).

RESEARCHER: I've spoken to some [students] who think that survival had to do with wealth and you're the first person who thinks the opposite—well, not the opposite but thinks it didn't have to do with wealth.
ARMANDO: [*After taking a moment to think*] Maybe it could have been [due to wealth]. Maybe…
RESEARCHER: I'm not trying to convince you one way or another. I just want to—
ARMANDO: Well, thinking about it, okay, yes. Well, the story—the way it was told was that women and children were supposed to be taken out [of the Titanic] first. But what if it was only the rich…what if it was their kids on first before the poor kids or women?

When I brought up the fact that other students believed the data supported the claim that the wealth of passengers played a role in their survival (lines 38-40), Armando grew pensive as he began to think about the simultaneous feasibility of both claims: that gender and age as well as wealth played a role in survival. Armando's consideration to modify his assertion of factors that affected survival from "women and children" to *wealthy* women and children, indicated to me that he was actively establishing conceptual connections between concepts defined in class and real-world phenomena, and, thus, reasoning about how a subset "works" in the real world. Still, unable to derive meaning from the actual data, Armando began to consider how an examination of the intersection of multiple variables (wealth, gender, and age) might provide a more complex portrait of survival on the Titanic than the indecipherable boxplot.

In an effort to reconcile this new possibility that wealth, gender, and age—together—played a role in survival, Armando began to manually explore the data by scrolling through the data set.

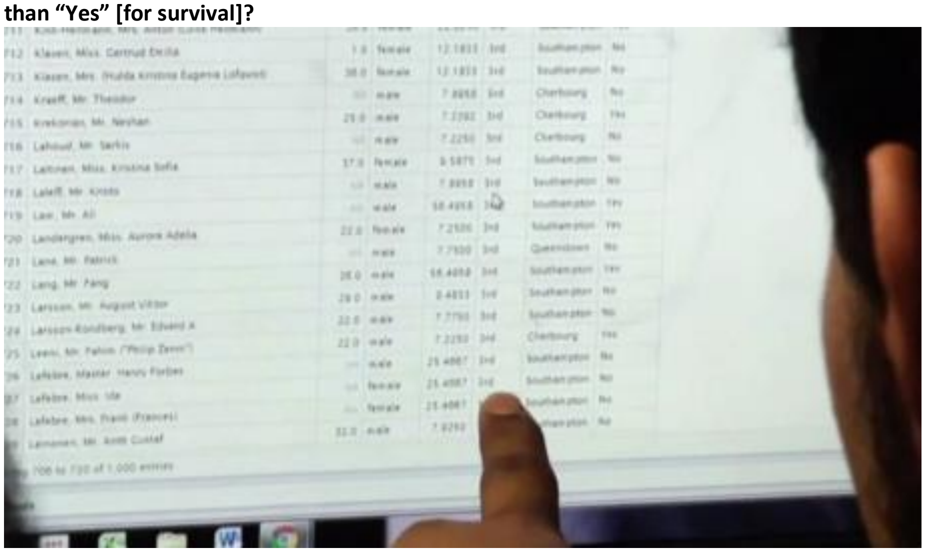
KATIE & ARMANDO: [*Begin to scroll through all entries in RStudio*]
KATIE: Oh [*surprised*], this is organized in order, by class.
ARMANDO: There was a thousand people on the ship?
KATIE: That's what is says…[*inaudible*]
ARMANDO: Alright, let's go back...look, there's some people that paid less and they're in first class.
KATIE: Remember, $20 back then was $500 today.
ARMANDO: I know, but there's people that are in first class and paid $134.
KATIE: Let's go to second class.
ARMANDO: That was $3,000 in today.
KATIE: Well, second class [fares] started getting lower but there's more in the [$20 range].

Considering that Katie and Armando initially valued understandings of the sinking of the Titanic based on a fictional portrayal, it is noteworthy that their manual exploration of the dataset imbued it with greater significance. We see an indication of this when Armando asked if there were 1,000 passengers on the ship, to which Katie responded in the affirmative, "that's what it says" (line 50) suggesting that it must be true if indicated in the data. Furthermore, by engaging in a manual exploration of the data set, which professional data scientists, computer scientists, and statisticians may consider highly inefficient at this stage of analysis when a boxplot was already generated, Armando began to develop an intellectual curiosity about the data while simultaneously developing criticality around the meaning of variables.

Armando's burgeoning curiosity about survival on the Titanic indicates that his rationale for engaging in this meaning-making learning process with data was becoming increasingly situationally significant. This means that the data was taking on greater personal relevance and was compelling him to engage in a more *critical* reading of the data and variables precisely because he was invested in understanding their significance and meaning. For example, he noted that the classification of first class made little sense given that one first class passenger paid a fare of $134 (line 54) while others paid amounts including $26, $39, $50, and $76. While students were asked to use fare price as a proxy for wealth, here, it seems it made more sense for Armando and his group to use the variable "class" as a more direct indicator of wealth than fare. Still, low fare prices for first class passengers brought about some degree of confusion for Katie and Armando. Also, had they based their understanding of wealth on fare prices as the lab instructed, a similar confusion would have ensued as some people who paid low fare prices were classified as first class. What we see here is that these students' deviation from the normative ways of doing data science was contributing to the cultivation of a CSDS identity that, while existing in opposition to legitimate ways of doing data science in this classroom, ushered in positive and motivated engagement with data analysis, which I believe is key to developing strong student identification with data science-doing and interest in data-scientific inquiry. In the following excerpt, Katie and Armando continue with their "rogue" approach to data science-doing:

ARMANDO: **If you look at it—look! —there's a lot of males that didn't survive… [*reads entries for "Survived" variable for first-class passengers*] "no," "no," "yes" …—Oh! They have ages too—up here.**

KATIE: Oh, some of them do, not all of them. Look at that one.
ARMANDO: [Are] these all adults?
KATIE: There was a "0.9" all the way at the top.
ARMANDO: That was a baby… [*continues to examine the data*] 17-year-old male—What would they have considered kids [back then]?
KATIE: Wait, there's one [age] four—check if they survived.
KATIE & ARMANDO: [*In unison*] Yes.
KATIE: Look, mostly a lot of females survived. I'm guessing there was like [*inaudible*].
ARMANDO: Okay, now let's look at the females that *did* survive.
KATIE: That one didn't.
ARMANDO: This one didn't [*referring to another passenger*]. Okay, if we go all the way down, this would be, what? —Third class? Third class females didn't…if you notice, there's more "No" than "Yes" [for survival]?

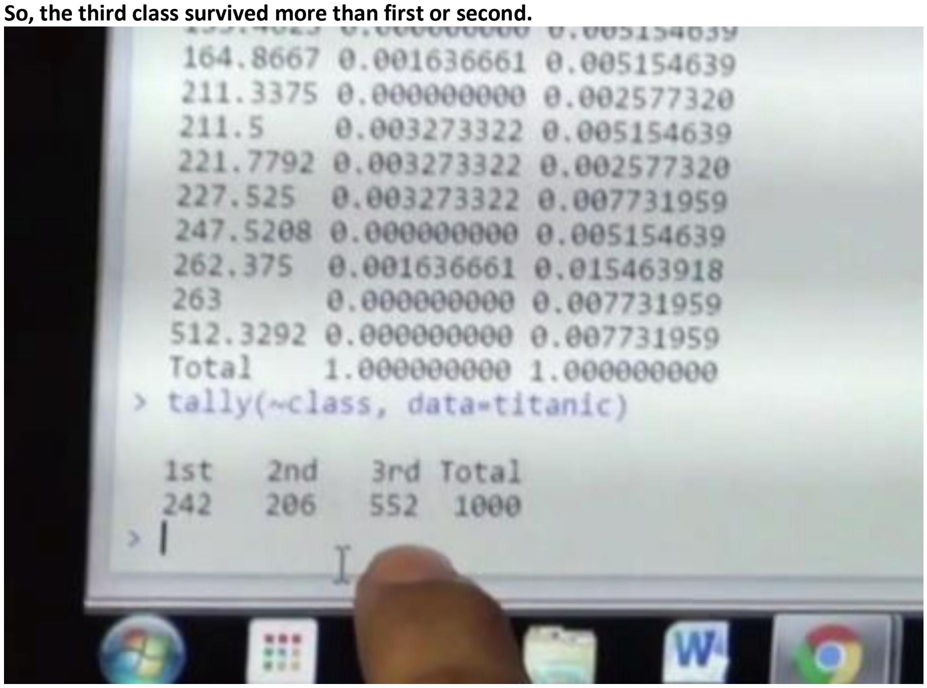


KATIE: Yes.
ARMANDO: Look—look at all of these...There, you see most of third class didn't survive.

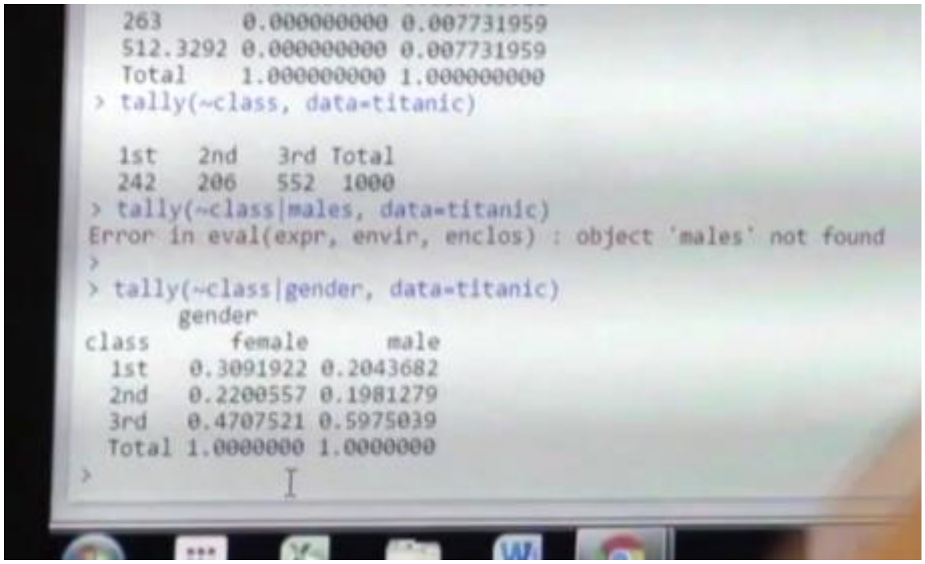
Armando was noticeably excited to see that the data included ages of passengers as he and Katie made a number of observations that, while not a formal part of their statistical analysis, contributed to their ability to make-meaning out of the Titanic data. Armando and Katie's manual exploration of the data, while arguably inefficient, led them to make data-based observations that they found very interesting including the notion that males in first class survived at lower rates than females in first class (lines 58-60 and 69), and that third-class passengers, both male and female, were less likely to survive than first-class passengers (lines 72-74). Again, here we see them engaged in a type of motivated data-analysis that stands wholly outside of what was constituted as normative, legitimate, and effective in this classroom. Their approach, however, cultivated opportunities for Armando and Katie to think richly with and about data and gain a deeper understanding of the data when compared to their production and [non]interpretation of the boxplot.

After developing data-based understandings of survival on the Titanic, Armando began to consider how he might operationalize an intersectional analysis of three variables (class, survival, and gender) in RStudio:

ARMANDO: Let’s do this: I’m going to separate them by class.
KATIE: How are you going to do that?
ARMANDO: Same thing we did here [earlier in the lab].
KATIE: And what [variable] did you [tally by]? “Survive”?
ARMANDO: Uhuh… Actually, let’s do that. Let’s tally it.
RESEARCHER: Can you tell me what you’re trying to pull up [in RStudio]?
ARMANDO: At this point, I’m not even sure what I’m tr—I’m just trying to see how many people survived for each class.
KATIE: So, is that survived or di—
ARMANDO: So, the third class survived more than first or second.



KATIE: But there were more “No”’s, that’s what I don’t get.
ARMANDO: Whoa, [*gets error message*] what did I do?
KATIE: Just put—
ARMANDO: It was gend—It’s supposed to be gender [*types in new tally code*].



RESEARCHER: And what is this that you pulled up? This looks different.

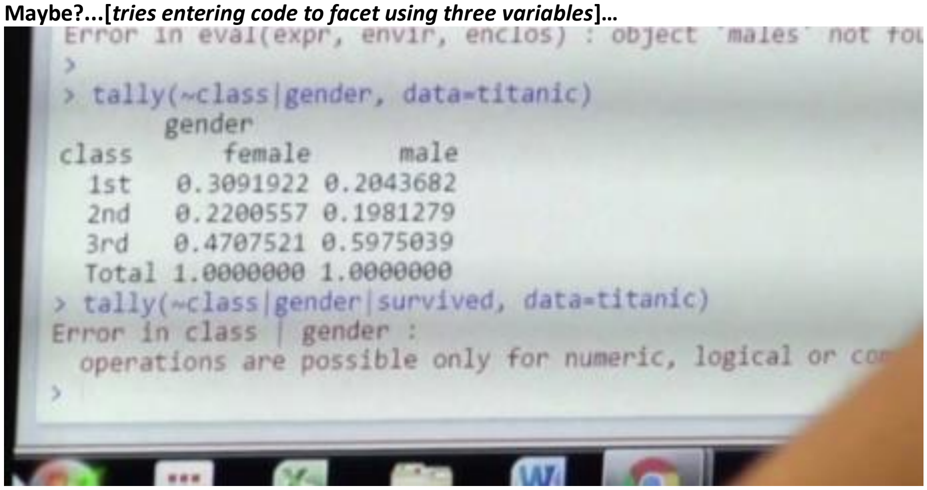
ARMANDO: It's separated by class and then gender. 30% of females in first class survived…Oh wait—I didn't even put that… [*to Katie*] Do you remember what was the code to…to check…two different variables? Do I need to facet again? Okay, I want to check the class, gender, and the survival.

KATIE: Umm… here, let me check [my coding history]—because I don't think it's going to…

ARMANDO: I mean, if I facet again, you think it'll work?

KATIE: I don't know.

ARMANDO: Maybe?...[*tries entering code to facet using three variables*]…



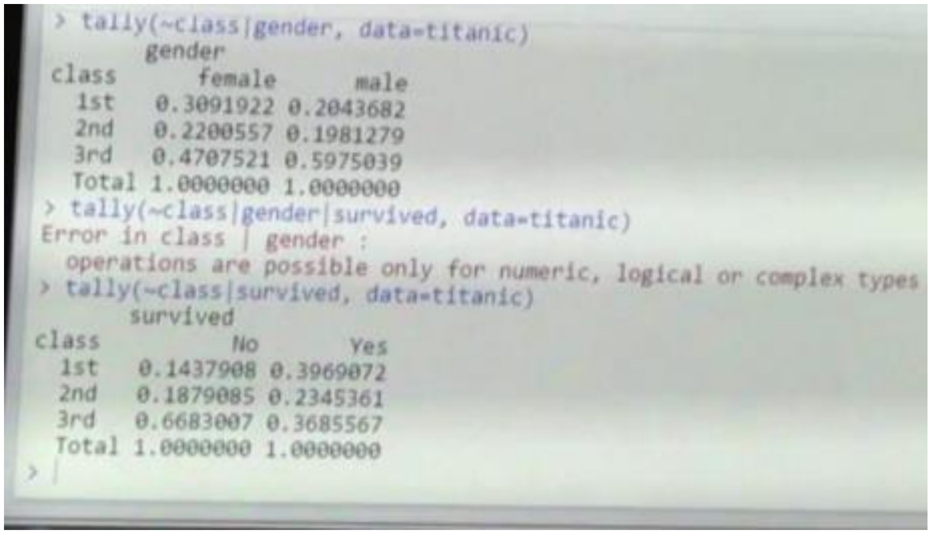
No [it didn't work].

KATIE: Because, we have done it, right? Like when we did three variables? [*Scrolls through her coding history*]

ARMANDO: Yes, but I don't know what was the code.

KATIE: I don't know…

ARMANDO: What if I do [*produces new tally*]



```
> tally(~class|gender, data=titanic)
       gender
class      female      male
  1st   0.3091922 0.2043682
  2nd   0.2200557 0.1981279
  3rd   0.4707521 0.5975039
  Total 1.0000000 1.0000000
> tally(~class|gender|survived, data=titanic)
Error in class | gender :
  operations are possible only for numeric, logical or complex types
> tally(~class|survived, data=titanic)
       survived
class         No       Yes
  1st   0.1437908 0.3969072
  2nd   0.1879085 0.2345361
  3rd   0.6683007 0.3685567
  Total 1.0000000 1.0000000
>
```

…Okay, but I want to get—by gender.

KATIE: I have the [flashcards] cards [that contain the codes we've worked with] at home. I can bring them next time if you want.

ARMANDO: Yeah, bring them.

RESEARCHER: Armando, so what does that [tally] tell you?

ARMANDO: It tells me the percentage of the people that survived or died, separated by class.

RESEARCHER: Can you interpret what it means?

ARMANDO: 14% of first class died and 39% of first class survived. 18% of second class died and 23% of second class survived. 66% of third class died and 36% of third class survived. The thing is I wanted to have it this way, but separated by gender.

RESEARCHER: You know, I just noticed something. Where it says "Total," it says "1"—

ARMANDO: Uhuh [yes].

RESEARCHER: So, I think…out of those who did not survive, 14% were first class, 18% were second, [and] 66% were third. Do you think that's correct?

ARMANDO: So, it's 14, 18, and 66%?

KATIE: She's saying that probably all of these numbers add up—like, how you know it's a "1"—it's probably [*referring to*] 100% here and then 100% here.

ARMANDO: Oh, okay. I see.

KATIE: So, she's saying, like, 39, 23, and 36 will all equal 100, but those are all the people that survived in each class.

ARMANDO: I see what you're saying now.

RESEARCHER: Do you agree [that that is what the tally is conveying]?

ARMANDO: I would have to go in deep into this.

The exchange above makes evident that Armando was actively thinking through, struggling, and improvising an analysis utilizing the tools and limited skills at his disposal (lines 87-90). However, the disconnect between what he *wanted* to do and what he *could* do impeded his analysis during a crucial moment in his development of a CSDS identity when his personal investment in the Titanic data set had peaked. His final tally, which clearly indicated survival and non-survival percentages for all classes, contained the necessary information for him to make a data-based claim about wealth and survival. Notwithstanding his ability to answer the lab question that prompted this exchange, he remained unsatisfied with the contents of the tally because he could not figure out nor recall the code necessary to visualize three variables in one tally. At this point, it became clear to me that Armando was no longer concerned simply with completing the lab assignment, but was instead intent on gaining meaningful understandings of the data.

Unfortunately, given that Armando's emergent approach to analyzing the Titanic data did not fit the strictures of the RStudio lab assignment, his high-level thinking remained undiscerned, unrewarded, and invisible to the classroom community. Furthermore, he ultimately failed to interpret the boxplot he and his peers produced and did not develop the understandings necessary to explain, in his assignment, how the boxplot supported his assertions about whether wealth played a role in survival on the Titanic. Thus, it comes as no surprise that Armando was identified by Ms. Gellar as one of the low-performing students in her IDS classroom because his desire to understand how technology works and his efforts to develop conceptual understandings were incompatible with narrowly defined standards for data science-doing. This is concerning given that with Armando we have a student captivated by and intent on learning with and about technology in personally fulfilling ways, yet his ability to achieve his ambitions of a career in STEM were compromised by his "low" academic performance in the IDS classroom.

Armando's developing CSDS identity stood outside of and deviated from the narrowly defined ways of legitimate data science-doing as constituted in the classroom. It is outside of normative demarcations of data science-doing that he began to develop an identity that was critical of data, the relationship and meaning of variables, and was personally invested in tinkering toward an intersectional analysis of the data. It is noteworthy that the emergence of his developing CSDS identity outside of and in opposition to normative data science-doing in the

IDS classroom did not negatively affect his self-valuation as a data science-doer. This I surmise based on his

1. Growing interest in and captivation with the data conveyed in his tone as he made observations;
2. Persistence in generating a telling aggregate of the data; and
3. Continued tinkering with what he knew despite an awareness of his limited recollection and knowledge of codes.

While this is promising and provides hope for his likelihood to pursue an education or career in data science, the *continued* development of his CSDS was compromised because it existed outside of what he came to understand as legitimate data science-doing.

Additionally, like many students in the class, this was Armando's first time engaging in and learning specifically about data science, therefore what was eventually constituted as legitimate data-scientific activity became representational of data science as a field in the absence of complimentary, contradictory, or corroborating experiences with data science. For this reason, I argue that reform-oriented efforts to cultivate the necessary critical literacies for democratic participation in our information and technology saturated society must be systematic in their efforts to cultivate personally meaningful CSDS identities that are in constant conversation with students' mathematical identities and with their out-of-school lives and lived experiences. Furthermore, Ms. Gellar's identification of Armando as a low-achieving student contrasted with his genuine curiosity for developing conceptual data-scientific understandings underscores the importance of adopting an intersectional information literacy and critical social theory/critical theory of education lens that supports critical social perspectives and validates students' meaning-making practices within math-centric disciplines that can easily default to traditional and narrowly defined standards for legitimate STEM-doing.

CHAPTER SEVEN
Toward Critical Data-Scientific Literacy for Equity-Oriented STEM Reform

Kim and Armando stood out as students who began to develop strong conceptual data-scientific understandings consisting of criticality toward aspects of data collection and analysis. Based on my analysis of their approach to data science-doing, I propose a framing of the development of critical social data-scientific understandings in Ms. Gellar's class as consisting of (1) developing in-depth understandings of the relationship between data science skills and concepts; (2) reasoning with and about data in real-world context; and (3) expressing data-scientific criticality. While I do not hold that IDS generally fostered the development of critical social data-scientific understandings among all students in Ms. Gellar's IDS class, I was privy to student activities and interactions that suggested the existence of opportunities for students to develop conceptual understandings. In this regard, my observations revealed the promising finding that students engaged in activities and made assertions that suggested that they experienced relatively equal opportunities to begin developing conceptual understandings and disciplinary understandings. However, narrowly defined ways of data science-doing capped continued development of conceptual understandings and instead emphasized and fostered disciplinary understandings. Figure 7.1 below outlines the types of actions and assertions that differentiated the development of critical social understandings, conceptual understandings, and disciplinary understandings in Ms. Gellar's IDS class.

While normative ways of doing data science in the classroom were conducive to student's abilities to exercise disciplinary agency, opportunities for them to explore and further develop conceptual understandings were generally not readily supported. There were numerous instances where students demonstrated that they were beginning to develop conceptual understandings, but these were rarely fleshed out or brought to fruition.

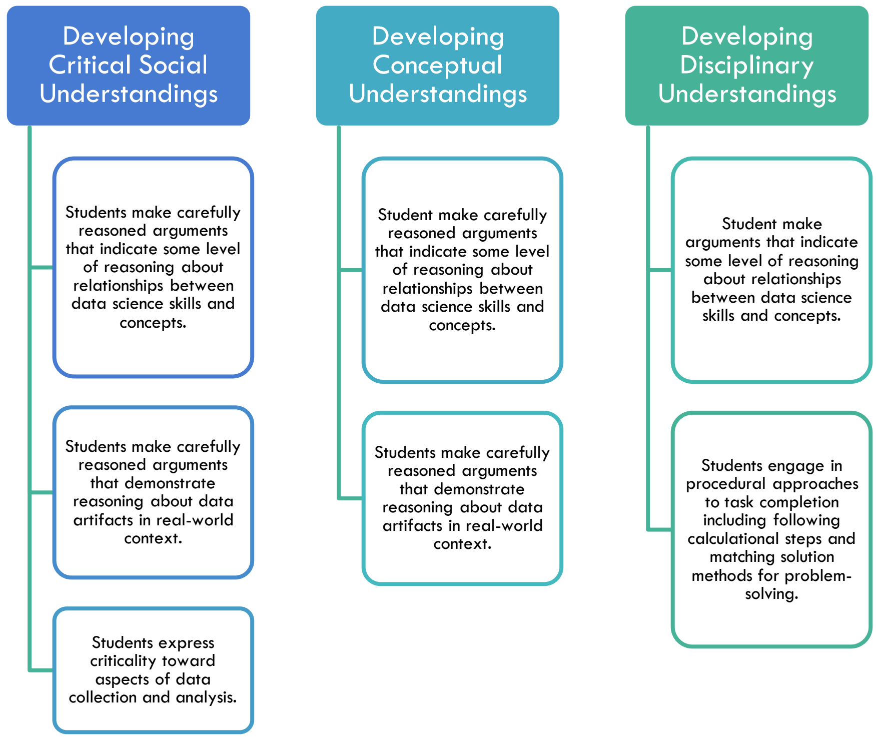


Figure 7.4

Therefore, while opportunities, indeed, existed for students to develop these understandings, several mechanisms ultimately impaired their ability to continue to do so in personally meaningful and enduring ways. In order to examine these mechanisms and the nuanced nature and interaction of the affordances and limitations of IDS as an instantiation of STEM reform efforts, I will dedicate the remainder of this chapter to discuss, first, the affordances of my intersectional theoretical framework for analyzing the developing data-scientific identities. I will then discuss both the limitations and unique affordances of Ms. Gellar's IDS class followed by my proposed extensions to Cobb et al.'s (2009) analytical scheme as a contribution to the field of education, namely to efforts aimed at increasing the presence of non-dominant groups in data science-related fields through the development of strong data-scientific student identities. Lastly, I will discuss recommendations for future iterations of data science-related programs for equity-oriented STEM reform.

Affordances of an Intersectional CST and CTE Framework for Analyzing Developing Data-Scientific Identities

My intersectional framework allowed me to analyze not only how students were developing data-scientific literacy, but also pushed me to think carefully about the ways that IDS supported or limited students' opportunities to think data-scientifically about themselves, their community, and society at large. This consideration proved immensely helpful for contextualizing the extent to which students were viewing and participating in the course as an exercise in higher order thinking about the usefulness of data science and data-scientific thinking for everyday life. It was also useful as an accountability lens for understanding how IDS, as an instantiation of equity-oriented STEM reform, interpreted and pursued equity in the classroom. While Ms. Gellar's IDS course provided an introduction and entry point for students who have been historically excluded from participation in STEM education and careers to gain exposure to data-scientific skills and concepts, statistical methods of inquiry, and coding experiences, my intersectional analysis underscored the fact that epistemological inclusion is complicated and difficult, but incredibly necessary. This is particularly the case for students who come from communities, cultures, ethnicities, and gender groups whose lived experiences and perspectives are often overlooked and/or misinterpreted in traditional school settings.

Addressing Limitations

Pace. The curriculum consisted of four units, but Ms. Gellar was only able to cover units 1 & 2 in both years of implementation. Additionally, on average, other school sites where IDS was implemented also did not get through to Unit 4. This is significant because all units were designed to build up to the fourth and final unit, thus there were concepts that were left unaddressed and connections left unmade in earlier units precisely because they were addressed in later units. Furthermore, while the course was designed to lead up to concluding projects that synthesized the “big picture” of IDS, students did not have the opportunity to engage in these activities or lessons. I believe this impaired their ability to see the larger purpose of IDS and its relevance to their daily lives.

Structure of lessons and labs. Despite the fact that the course was promoted as “inquiry-based,” the questions teachers were advised to pose to students, in accordance with the curriculum and professional development, did not require that students engage in deeper-level thinking about data-scientific concepts. In fact, the questions Ms. Gellar posed during whole-group discussion were consistent with those suggested in the curriculum. Furthermore, discussion topics, student responses and contributions, purposes of tasks, and learning objectives were all pre-determined, leaving little room for new and emergent understandings.

Authority distributed to technology. One of the more striking findings in my observations was the level of authority that was distributed to technology in the classroom. While RStudio lab assignments were authored by individuals in the Energize project, when students carried out these tasks they interacted not with individuals, but with the technology itself. This means that there was no opportunity for students to engage in discussions with the technology about lab content and purpose, or to pose challenges to assertions made within labs. Lab assignments were unchangeable and dictated when and to what extent students could engage in their own interpretation of tasks and solution methods. The limitations of this aspect of labs were clearly evident in my observations and interviews with Kim and Armando. Despite their investment in understanding data-scientific concepts and their application of coding skills to understand deeper meanings of data, the narrowly defined ways of coding within IDS and the highly structured direction for lab assignment completion hampered rich thinking with and about data, which went against IDS reform goals. The functionality and use of RStudio itself could also not be challenged as the coding component and the mechanisms through which it was achieved were unchangeable aspects of the curriculum. Existing scholarship cautions against the perception of new technologies as value-free (Selwyn, 2016; Selwyn, 2015; Couldry, 2013; boyd & Crawford, 2012) and its uncritical adoption into the classroom (Philip, 2017; Philip et al., 2013). This is problematic for its implications on student's abilities to exercise agency in the classroom and develop conceptual understandings in the lab setting because it feeds into the current popularity of new data-generating technologies and Big Data Hubris that function to uphold positivism (Couldry, 2014; boyd & Crawford, 2012) and the belief of technology as rational (Standaert, 1993; March, 2006; Khalifa et al., 2014).

Additionally, while the curriculum required that students have access to computers in order to satisfy the coding component, access to properly functioning technology was as issue from the beginning and ultimately resulted in the class having to meet at the computer lab for lab days. This affected the learning-rhythm that was established in the classroom as students could not sit and face each other in groups like they had in the classroom and as was directed in the curriculum. I do not doubt that this was an issue encountered at other school sites, as all were part of the same large, overenrolled, and under-resourced school district.

Block scheduling. On a structural level, block scheduling at MSHS compounded the effects of the pace and structure of the curriculum. Given that students met as a class three alternating days one week and two alternating days the following week, cycling in this pattern for the entire academic year, any missed lessons due to testing, high school senior activities, teacher and student absences, professional development days, and holidays dramatically extended the duration between lessons, posing challenges to students' ability to retain and build on successive disciplinary understandings. For example, if Ms. Gellar's IDS class met on a Thursday, they would not meet again until Monday of the following week, thus they would go three days without continuing to build on the previous lesson. What is more, if that Monday happened to be a pupil-free day, students would not meet again until Tuesday for a shortened period of time since Tuesdays were reserved for professional development for teachers and, thus, consisted of shorter school days for students. On several occasions, Ms. Gellar attempted to mitigate the compounded effects of scheduling by providing students with supplemental lessons and/or scaffolding following a large break. While this allowed students to recollect previous lessons and gain additional understandings, it also meant that she had to postpone teaching lessons from the curriculum, which made keeping up with the already fast-paced curriculum more challenging.

Student placement. A noteworthy challenge to the implementation and goals of IDS was also posed by placement of high school seniors in a course designed for sophomores and juniors. As I stated in Chapter 5, many of the seniors enrolled in the course viewed their participation in IDS as instrumental to their high school graduation or post-high school ambitions. While these were respectable ambitions, their motivation to perform well in the class did not mean that students were necessarily motivated to gain conceptual understandings and develop data-scientific learning identities for personal enrichment. Additionally, the fact that the overwhelming majority of students in the class were seniors meant that students were also absent due to standardized testing, college entrance testing, AP testing, and senior activities which added the effects already posed by other factors that impacted the flow of the course and its ability to maintain a cohesive and consistent learning rhythm from lesson to lesson.

Taken together, the pace of the curriculum; structure of lessons and labs; authority distributed to technology; block scheduling; student placement in the class all influenced students' abilities to develop strong conceptual understandings of data-scientific skills and concepts. While some of these factors, particularly block scheduling and decisions regarding student placement, were outside of the purview of curriculum writers and top-level IDS decision-making, these proved highly influential in the eventual learning outcomes for students enrolled in the class and the benefits they were able to draw from taking the course. For this reason, it is necessary that future iterations of IDS take these findings into account in their determination of how best to attune and implement similar initiatives in schools or out-of-school community spaces.

Capitalizing on the Unique Affordances of IDS

A starting point. Kim provided a clear articulation of the value and power of IDS as an introductory course when she said,

> This class really helped me figure out where to start if I ever do want to pursue this field. It gives the introduction part and it sets off the path, because I know that's the most intimidating part: where to start in this whole vast technology world.

Indeed, some students shared a similar and powerful sentiment. In this respect, Ms. Gellar's IDS class was immensely useful for introducing students to a starting point—a point of demystification for coding and data-scientific inquiry for students who had little-to-no experience working with data and coding. IDS also helped demystify the loaded term "coding"—a word closely associated with the work of hackers, representing a highly specialized skillset not typically attributed to working-class youth of color who are among those historically excluded from education and careers in tech-related fields. While I believe that the RStudio lab assignments were overly structured, the step-by-step instructions included in RStudio lab slides allowed all members of the classroom community to gain experiences coding, and thus helped demystify coding as an exclusive practice. Regardless of whether students ultimately liked the coding component or felt compelled to pursue a career that involves coding post-IDS, their ability to code in RStudio allowed them to understand coding as an *attainable* skillset. In so doing, it provided students with an alternative view of coding as a learnable skill within their reach. Thus, in order to inspire interest and curiosity in data science and build on the affordances of IDS as a starting point, it is essential that we address the limitations described above as I believe they stifled rich learning opportunities and students' abilities to identify as data science doers. This unique affordance of IDS cannot inspire data-scientific educational and career aspirations among students if they are not afforded opportunities to develop strong conceptual data-scientific understandings and exercise conceptual agency in the classroom.

Emphasis on collaborative problem-solving. Another one of the unique affordances of Ms. Gellar's IDS class was its promotion of peer-collaboration. Although some students utilized instances of collaboration to simply copy or give others answers, still some students benefited from collaborating because it allowed them to not only know the answer, but also practice explaining certain responses, solutions, and rationale's behind the use of solution methods. This means that for some students, helping others functioned as an exercise in understanding the purpose of tasks and developing conceptual data-scientific understandings. For those students who were regarded as highly competent by their peers, this helped motivate them to gain disciplinary understandings as they anticipated that peers would approach them for help. Sandy attested to this by saying that while she did not initially regard herself as highly adept at data science-doing, her peers' gradual and increasing tendency to come to her for help inspired her to gain understandings necessary for her to help them. This means that IDS' emphasis on peer collaboration was not only valuable for including students in the dissemination of knowledge, but also served to inspire motivated engagement among some students as they sought to purposefully develop data-scientific competencies.

Expanded views of science and mathematics. Future iterations of IDS and other data science-related initiatives have the potential to pose a powerful challenge to the manifestation of the nature-culture divide (Bang et al., 2012) in the field of data science. For Bang et al. (2012), this refers to existence of a binary that obscures the interdependence and interconnectivity of science and culture, functioning to uphold positivist views of science as epistemologically neutral. Similarly, I posit that the data-culture divide adopts a parallel positioning of the study of data and culture as mutually exclusive. In order to challenge positivist continuities between traditional school science and data science, the study of data must not only be couched in real-world scenarios but also thought of as a versatile and widely applicable conceptual skillset. My interviews revealed that by taking Ms. Gellar's IDS class, several students gained understandings of data science as "far-reaching" and applicable to diverse facets of their everyday lives. The question, then, is how do we build on these understandings? How can we fortify data science programs so that we begin to shift our thinking of data science in ways that challenge universalist treatments of this burgeoning field? I believe these efforts must include youth in the planning and design phases. Without the perspectives and insights of those targeted by these programs—that is, youth of color and young women—we cannot effectively address their concerns and interests as they pertain to educational and career endeavors, the usefulness and benefits of data-scientific thinking for everyday life, and the potential to use data science for civic engagement.

Toward Progressive Reiterations of IDS

In the year 2000, Calabrese Barton & Yang published their account of a case study of Miguel, the young Puerto Rican father and clandestine herpetologist who loved science and gained personally meaningful experiences with out-of-school science, only to have his "enthusiasm for science and nature…neither acknowledged formally by his teachers nor cultivated in school" (p. 872). About 15 years later, I saw parallel manifestations of the culture of power, particularly for Armando who expressed an enthusiasm and affinity for all things technology and, like Miguel, gained personally meaningful scientifically rich learning experiences out-of-school only to experience discontinuities between scientific doing and thinking inside and outside of school. For Miguel, this lead to his internalization of problematic notions of what constituted legitimate sources of knowledge and science-doing. I would like to believe that both Kim and Armando's tendency to think data-scientifically to make decisions and solve problems in their everyday lives will continue beyond high school and that they will persist in their pursuit of sating their scientific curiosity, be it in the field of data science or another field in STEM. It is extremely imperative and necessary, however, that programs aimed at exposing students to new fields recognize that for many of these students, exposure is not enough and an introductory course is more than that. For students who do not have a data scientist parent or relatives who can speak to them about computer science, or friends who can introduce them to diverse applications and functions of STEM fields, courses like IDS become exemplars of what data science *is*. Thus, these initiatives possess immense power in shaping initial impressions of a complex and continually developing field. At the same time, I recognize that the American educational system itself is undergoing dramatic shifts in light of changes to Common Core standards and the political assault on the American public education system, but these changes and challenges must be treated as impetus for improving science education reform efforts. This necessitates collaboration, deep reflection, and evolving methodologies and approaches for STEM reform.

Specifically, with regards to data science-oriented STEM reform efforts, we educators, researchers, curriculum writers, policy makers, and anyone invested in the pursuit of equity in STEM for historically underrepresented groups must challenge ourselves to address observed shortcomings while capitalizing on the affordances and promise evident in data science programs. It is in this vein of professional collaboration toward improved and equitable quality education that I discuss some of the observed hindrances and unique affordances of Ms. Gellar's IDS course.

Contributions to the Field

Cobb et al.'s (2009) framework for analyzing the mathematical student identities that students develop in a mathematics classroom was immensely useful for mapping students' development of data-scientific identities within a math-centric course like IDS. My application of the framework, however, revealed the need to expand the analytical scheme to include aspects of learning and student identity unique to data science. Thus, my application of Cobb et al.'s (2009) framework allowed me to begin to develop a specifically data-scientific analytical scheme to account for the differential learning settings and technological components present in IDS as an instantiation of equity-oriented STEM reform. Firstly, when analyzing the general classroom obligations that students feel compelled to fulfill in order to meet expectations of what it means to legitimately do data science as constituted in the classroom I propose modifications to the Cobb et al.'s (2009) scheme as indicated in Figure 7.2. My proposed modification begins by extending the analysis paid to authority distribution in the classroom to include consideration of the extent to which students are given opportunities to contribute to the creation of disciplinary knowledge in the classroom. Also, after applying the original framework to my analysis, I realized that authority distribution to teachers was not an explicit focus of analysis.

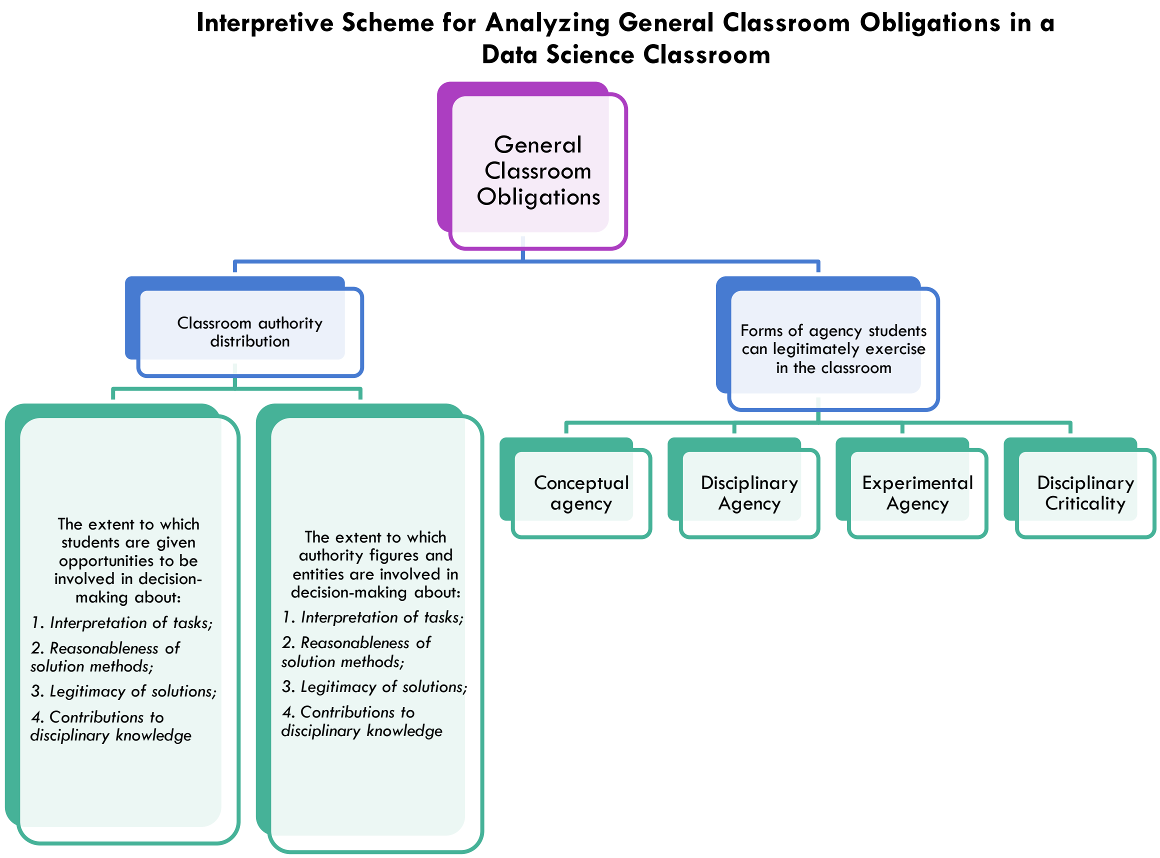


Figure 7.5 Cobb et al.'s (2009) scheme adapted for analysis of general classroom obligations in a data science course, extended.

Given the central role that epistemological inclusion must play in STEM reform efforts, I find it necessary to also analyze the extent to which teachers hold authority in the classroom. Additionally, because data science involves the use of new learning technologies in the classroom, my proposed modifications to the original scheme include incorporating an analysis of how authority is distributed to these technologies because, as evident in this dissertation, ceding unchecked authority to technology can prove problematic for students' abilities to exercise conceptual agency when they employ the use of technology to engage in data science-doing.

Additionally, I also propose extending an analysis of the forms of agency students can legitimately exercise in the classroom to include experimental agency and disciplinary criticality. I define experimental agency as students' abilities and willingness to (1) experiment with new learning technologies for purposes of exploration; and intuitive use of technology for decision-making and problem-solving in the classroom; and (2) to explore alternative solution-methods when recommended approaches prove ineffective or inefficient. While Kim and Armando exercised experimental agency in the computer lab setting by adopting alternative approaches to data analysis through manual exploration of data and tinkering with codes, these were not legitimate practices rewarded in the course. However, an equity-oriented IDS classroom should allow and reward students for taking initial disciplinary understandings and applying them in unanticipated ways if they contribute to students' abilities to engage in meaning-making with data and fundamentally nurture conceptual understandings. These approaches should not emerge outside of legitimate data science-doing, but instead be significant and legitimate forms of agency embraced in the classroom.

Similarly, disciplinary criticality must also be embraced as part and parcel of student participation in a course designed to introduce them to an ever-evolving field. I define disciplinary criticality as consisting of what I have referred to earlier in my dissertation as "challenging acts" wherein students contest the conceptual rationale behind others' disciplinary assertions for the purpose of developing deeper data-scientific understandings. Challenges should be understood as opportunities to develop new disciplinary understandings and validate students' disciplinary assertions as legitimate contributions to data science. Students' abilities to exercise disciplinary criticality in the classroom is a direct extension of authority distribution—if a student is allowed to partake in decision-making in the classroom, then they will also be able to partake in the contestation of disciplinary assertions and the development of new data-scientific knowledge.

Furthermore, I also found that students' valuation of data science as a field and data-scientific thinking for everyday life were affected by their views of the usefulness of data science for improving society and helping others. Students' inability to see the potential of data science to improve their lives and that of their families and communities affected their development of motivated engagement in data science-doing. Thus, I propose that an analysis of the personal identities that students develop in data science-related courses include serious examinations of students' understandings and valuations of data science and data-scientific thinking for real-world problem-solving, decision-making, and civic engagement (Figure 7.3 below). Educators, researchers, curriculum writers and policy makers should not only apply these schemes in their analysis of implemented programs and initiatives, but also take each component of the scheme into careful consideration during the design phase. I do not hold that my adoption and modification of Cobb's scheme is wholly complete and unchangeable, just as I do not believe that my findings from Ms. Gellar's IDS class can speak to outcomes in all data science course and programs. However, it is also necessary to acknowledge and account for the powerful findings and insights gained through my observations and interviews with students who represent the largest ethnic and cultural minority in the country and who are among those who have not had equitable opportunities to enter STEM fields.

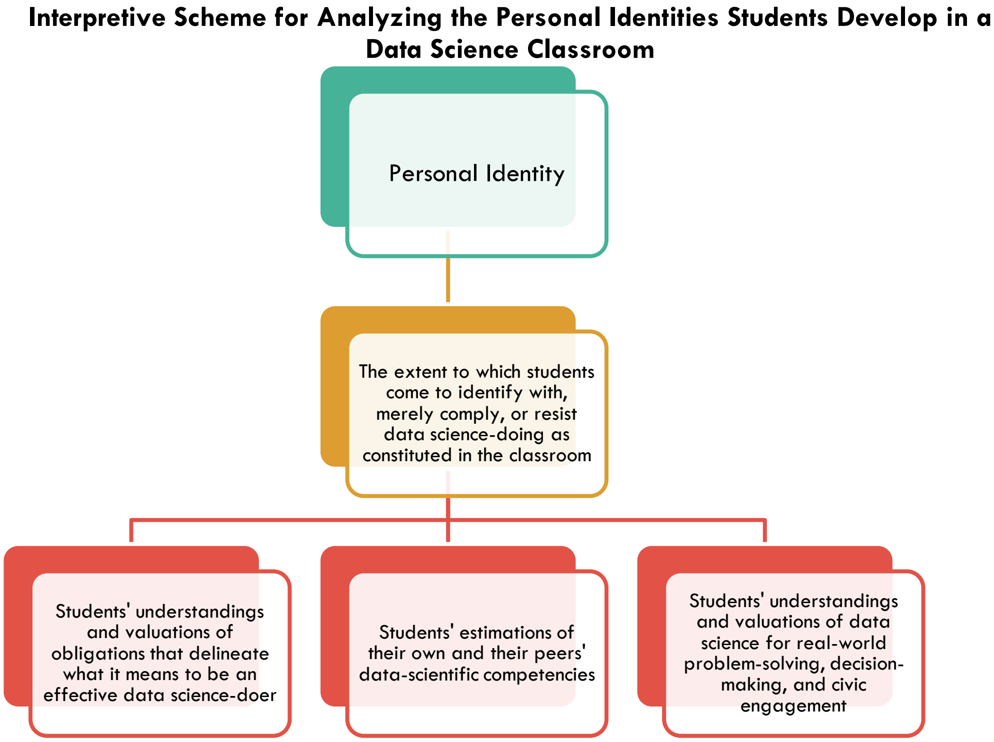


Figure 7.6 Cobb et al.'s (2009) scheme adapted to analyze students' personal data-scientific identities, extended.

Recommendations for Future Instantiations of Data Science for Equity-Oriented STEM Reform

During my time at MSHS, I sought to think of my position in Ms. Gellar's IDS class as that of a learner. Despite my credentials, I approached interactions with students as unique opportunities to learn about students, to understand their thought process, and to value their contributions—be they in the form of assertions, explanations, or clarifying questions. By doing so, I was able to gain more intimate insight into why students participated in the constitution of obligations the way they did as well as what accounted for their motivation to learn about data science, or their lack thereof. Based on my experience interacting with students in Ms. Gellar's IDS class and my analytical findings, I offer the following recommendations and reflection questions for future iterations of IDS and other data science-related programs and initiatives.

Consider messaging beyond explicitly stated learning goals. Given explicitly stated learning goals, how and what are student learning regarding what is expected of them? How do these expectations align with a humanistic approach to equity-oriented STEM reform in data science? What do teaching approaches convey to students regarding the forms of agency they can legitimately exercise in the classroom?

Reflect on the significance of introductory data science courses for historically excluded populations. In the absence of numerous and diverse opportunities to learn about data science and the wide-ranging applications of data science for everyday life, introductory data science courses, programs, and initiatives must be understood as exemplars for a new and rapidly developing field. This imbues introductory experiences with much representational power, thus, it is necessary to understand how these initiatives represent the field of data science and the versatility of data-scientific thinking. Understanding how data science is represented in introductory courses is important as it holds implications for students' motivated engagement in data science-doing.

Involve youth in the design of programs aimed at improving educational and career potentialities. Without seeking out and accounting for the lived experiences, educational and career ambitions, and disciplinary interests of youth, STEM reform initiatives cannot adequately address the needs of students. Students must be allowed to participate in the creation of programs aimed at helping them achieve quality learning experiences in STEM. Without considering students' perspectives, we will remain blind to the unique contributions of their meaning-making processes and how they can bolster equity goals of STEM reform efforts like IDS. Most importantly, in line with arguments made by Bang et al. (2012) regarding what counts as knowledge in traditional school science, the difficulty of pursuing an equity-oriented data science project lies in confronting the long-standing and ideologically rooted settled expectations (Harris, 1995) for what counts as legitimate scientific knowledge and who can contribute to its constitution, as well as what counts as legitimate science-doing and who can do it. Involving youth in programs that target them and their communities is essential for disrupting settled expectations of what constitutes legitimate scientific knowledge and who can contribute to its constitution. Moreover, involving students in the design process can help STEM reform efforts understand, and necessarily address, why some groups convey limited engagement even within reform-oriented spaces (Murrell, 1999; Lubienski, 2002). In the case of students in Ms. Gellar's IDS classroom, reasons for limited engagement varied, but salient among them was the belief that data science was not beneficial to, relevant to, or applicable in everyday life beyond the course itself.

Reflect on possible manifestations of the culture of power and the data-culture divide in the new field of data science. Reflecting on the ways in which the culture of power (Delpit, 1998; Calabrese Barton & Yang, 2000) pervades the sciences and extends into budding and rapidly developing fields like data science is imperative for a truly equitable shift in access to and diversity in STEM. This requires honesty, humility, and a deep understanding of the culture of power as persistent and eventually-emergent unless there is a conscious and explicit effort to challenge the myth of science and data science as objective. This is consistent with calls to action in existing scholarship aimed at reframing narrowly defined science as natural and factual (Brickhouse, 1994; Stanley & Brickhouse, 1994; Calabrese Barton, 1998; Calabrese Barton & Yang, 2000; Bang et al., 2012). A lack of reflection, here, will undoubtedly allow the culture of power and the data-culture divide to persist even within educational projects with stated goals of equity and social justice mirroring programs critiqued by Brickhouse (1994) as shortsighted, though well-intentioned.

Adopt life-long learner identities among educators, researchers, curriculum writers, and policy makers. Learning to think of ourselves—as educators, researchers, curriculum writers, and policy makers—as life-long learners is necessary for positioning ourselves and students as simultaneous learners and educators. This is fundamentally a matter of reflecting on authority distribution in the classroom. By no means should this be interpreted as my proposal to devalue the specialized training and contributions of teachers in the classroom, but embracing the notion that youth, particularly those from non-dominant groups, bring unique perspectives and meaning-making practices into the classroom can prove immensely beneficial for developing new disciplinary understandings and positioning students as valued contributors to scientific knowledge (Bang et al., 2012). By valuing students' contributions, students learn that they are not only responsible for learning established disciplinary understandings, they also learn that they are valued contributors to STEM knowledge systems, fundamentally challenging traditional approaches to teaching and learning that position teachers and beholders of knowledge and students as receivers of knowledge.

Conclusion

This dissertation has been a social justice project aimed at contributing to the improvement of social justice and equity efforts in education, namely those aimed at reforming STEM education as we know it and cultivating critical data-scientific literacies among non-dominant groups. To do so, I carried out a systematic analysis of the identities that students developed in Ms. Gellar's IDS course as an instantiation of STEM reform efforts to reveal the complex nature and factors that shaped learning outcomes for students from a predominantly Latino classroom, school, and school district. I argue that IDS afforded several unique affordances to students' opportunities to learn with and about data, but that a complex set of factors generally impaired students' abilities to develop strong data-scientific identities and identities as STEM-doers via the development of conceptual understandings and exercising of conceptual agency. I strongly argue that IDS has much to offer students who would not otherwise gain experiences learning to work with and think critically about data, but must actively work toward addressing limitations and capitalizing on the affordances that emerged during the implementation. I believe that only by disrupting settled expectations of authority in the classroom and challenging positivist views of science and scientific knowledge as objective and static can we contribute to the improvement of STEM education and career opportunities for Latino students and inspire motivated learning with and about data science, thus inspiring the cultivation of strong learning identities among groups traditionally underrepresented in data science-related sciences and STEM in general.

APPENDIX A
IDS Curriculum Overview, Units 1 and 2

IDS Curriculum Overview, Unit 1			
Themes	Data Are All Around	Visualizing Data	Would You Look at the Time?
Overview	Students learn about the importance and ubiquity of data in their everyday lives. They learn to handle and organize data using data storage structures. Students first collect and manage data manually and then transition into using statistical analysis software, RStudio, to learn about data distributions. Students begin analyzing and visualizing their own data collected through the mobile phone Energize app.	Students learn about the role that data collection methods, visual representations, distributions, and technology play in their ability to analyze and interpret data in real-world terms. Students expand their use of RStudio by creating data visualizations for comparison and analysis.	Students learn that claims and reports can be evaluated using data summaries. Summaries of categorical and numerical data allow students to see patterns in the data. Using RStudio, students learn to create tabular displays of data, calculate frequencies, subset data, create new categorical variables from numerical variables, and clean data to make it readable in RStudio.
Common Core State Standards	*S-ID: Summarize, represent, and interpret data on a single count or measurement variable.* *S-ID 1*: Represent data with plots on the real number line. *S-ID 2*: Use statistics appropriate to the shape of the data distribution to compare center (median, mean) of two different data sets. *S-ID 6*: Represent data on two qualitative variables on a scatterplot and describe how the variables are related.	*S-ID: Summarize, represent, and interpret data on a single count or measurement variable.* *S-ID 1*: Represent data with plots on the real number line. *S-ID 3*: Interpret differences in shape, center, and spread in the context of the data sets, accounting for possible effects of extreme data points. *S-ID 6*: Represent data on two qualitative variables on a scatterplot and describe how the variables are related.	*S-ID: Summarize, represent, and interpret data on a single count or measurement variable.* *S-ID 5*: Summarize categorical data for two categories in two-way frequency tables. Interpret relative frequencies in the context of data (including joint, marginal, and conditional relative frequencies). Recognize possible associations and trends in the data. *S-ID 6*: Represent data on two qualitative variables on a scatterplot and describe how the variables are related.
Skills & Concepts	• Data • Data set • Data trails • Privacy • Data collection, organization, representation • Data collection campaigns • Numerical and categorical variables • Data cycle • Statistical questions • Participatory sensing • Data analysis and interpretation • Dotplots	• Data distributions, shape, variability • Data visualizations • Clustering • Typical value • Algorithm • Input, output • Histogram • Participatory sensing • Shapes of distributions • Using the Dashboard, PlotApp • RStudio basics • RStudio basic commands • Computer syntax • Reading and interpreting multi-variable scatterplots, bar plots	• Evaluating statistical claims • Two-way frequency tables • Cleaning data names, categories, and strings in RStudio • Relative frequency • Marginal frequency • Joint frequency • Conditional relative frequency
Campaign	*Food Habits*: Students collect data about their snacking habits using a mobile phone app designed by Energize.	*Food habits*: Students continue data collection campaign.	*Time Use*: Students monitor the amount of time they devote to activities such as sleeping, studying, eating, and partaking in media.

IDS Curriculum Overview, Unit 2

Themes	What is Your True Color?	How Likely is It?	Are You Stressing or Chilling?	What's Normal?
Overview	Students learn that a number of different measurements are useful for making sense of large amounts of data; these include measures of center and measures of spread. Students learn to find and understand numerical summaries about the data.	Students learn that probability simulations help determine expectation of events and that probability measures the long run frequency of the occurrence for chance outcomes. Probability can be approximated via mathematical calculation or through simulations in RStudio.	Students learn that permutations of data provide a model that shows us how the world behaves if chance is the only reason for differences between groups or variables. Students also learn to determine if outcomes occur by chance or design by analyzing simulated probabilities vs. real ones.	Students learn that the normal curve describes many real phenomena, as values tend to cluster toward the center and less so away from the center. Students learn to overlay a normal curve to a histogram to informally determine if data are normally distributed and to estimate probabilities using RStudio.
Common Core State Standards	*S-ID: Summarize, represent, and interpret data on a single count or measurement variable.* S-ID 2: Use statistics appropriate to the shape of the data distribution to compare center (median, mean) and spread (interquartile range, standard deviation) of two or more different data sets. S-ID 3: Interpret differences in shape, center, and spread in the context of data sets, accounting for possible effects of outliers.	*S-CP: Understand independence and conditional probability and use them to interpret data.* S-CP 2: Understand that two events A and B are independent if the probability of A and B occurring together is the product of their probabilities and use this characterization to determine if they are independent. S-CP 9: (+) Use permutations to perform [informal] inference. **This standard will be addressed in the context of data science.*	*S-IC: Understand and evaluate random processes underlying statistical experiments.* S-IC 2: Decide if a specified model is consistent with results from a given data-generating process, e.g., using simulation.	*S-ID: Summarize, represent, and interpret data on a single count or measurement variable.* S-ID 4: Use the mean and standard deviation of a data set to fit it to a normal distribution and to estimate population percentages. Understand that there are data sets for which such a procedure is not appropriate. Use calculators and RStudio to estimate areas under the normal curve.
Skills & Concepts	• Subsets • Relative frequency • Measures of center • Measures of spread • Deviation • Comparing distributions • Five-number summary • Custom functions • Statistical questions • Comparison statements • Converting dotplots into boxplot	• Probability • Simulation of random events using RStudio • Sample proportion • Chance • Events • Independent & dependent events • Biased probability • Sampling with/without replacement • Compound probabilities • Two-way tables • Calculating probability	• Chance differences for categorical variables • Inference for categorical variables • RStudio "'do' loops" • Simulations • Randomness • Shuffling • Compound probabilities • Merging data sets • Stacking vs. joining • Answering statistical questions re: merged data	• Normal curve • Normal distribution • Standard deviation • Z-scores • Empirical rule • Shuffling in RStudio • Normal probability
Campaign	*Personality Color*: Students complete the Personality Color survey that will collect their data about their personality styles.		*Stress / Chill & Personality Color*: Students collect data on how they feel throughout the day, if they are alone, with others, what they are doing, and	

APPENDIX B
Student Exit Interview Protocol

STEM Identity

1. Let's say you're trying to explain what a data scientist is to a very young child in your family. Can you tell me, in detail, how you would describe a data scientist to them?
2. What type of work does a data scientist do?
3. Can you tell me what types of data science skills *you* have developed in IDS?
4. Has IDS shifted your thinking on what it means to be a mathematician, statistician, scientist, or computer scientist?

Self-Identification with STEM Identity or Science-doing

5. Do you want to work and/or study in fields related to science, technology, engineering or math? Please explain.
6. Do you think you could be a data scientist if you wanted to? Why or why not?

Relevance/Real-World Importance

7. "Data science, computer programming, and coding are important for future careers." Does this statement motivate or entice you to study or work in fields related to data science?
8. How [can/will/might] you use the skills you learned in this class in your personal life?
9. How have your views of yourself and the world changed after taking engaging in data science?

"What does this really mean?"

10. What will you do with what you've learned in IDS?
11. Do you care about data science?
    a. Should *I* care about data science?
    b. Should your *peers* care about data science?
    c. Should your *family* care about data science?
    d. How about your *community*, people that live in your neighborhood, should they care?
    e. What about the *children* in your community, should they care?
    f. In your opinion, *who* should care?

Looking Ahead

12. Can you tell me about your academic or career plans after high school?
    a. What steps have you taken so far to achieve this goal?

2nd Year IDS Students

A. Can you tell me how your experience in IDS this year was different than your experience in IDS last year?
   a. Why do you think you did better in the class this time?
   b. Why do you think you didn't do as well last year?

APPENDIX C
Sample Lesson Worksheet Handout

Name:____Block:_____________Date:_________________

Unit 2
Theme 1: What is Your True Color?

Enduring Understanding

- Statistics enable us to ____
  ____. Numerical summaries capture important elements of a ____________________________.
- Measures of center, also known as measures of_____________________________________, show the tendency of __________________data to_____________________________________
  ____.
- Measures of spread, also known as measures of ____________________, show how much the quantitative data is ____.
- Measurements of the ____
  _________________________________ and the _______________________________________
  ________________ within the data can provide insightful indicators about the data.

Lesson1: What is Your True Color?

Objective: Students will collect data that might tell them about their personality type, and will understand how to subset their data.

Vocabulary

- Subsets

Essential Concept

The 'typical' value is a value that ____, even though we know that not all members of the group _______________________________________
_________________________________.

True Colors

How will you use the data from the *True Colors Personality Test*?________________________________
____

1

Personality Test

What is the stated objective of the personality test?______________________________

____

____

	Score	Color
Predominant:		
Secondary:		
Third:		
Fourth:		

Color Group Discussion

How many students are in each group?

Orange	Gold	Blue	Green

What is the predominant personality color in the class?______________________________

What is Typical?

Describe your group:____

____

____

____

Group Descriptions

Color:______________________

Description: ____

____

____

____

2

Appendix D
Sample Lab Worksheet Handout

load_lab( )
9

Lab 2A: All About Distributions

In the Beginning…

1. What kind of work have we done in the labs so far?

2. What will we be doing in this lab?

How to Talk About Data

3. What do we usually want to know when we make plots?

    a.

    b.

    c.

    d.

Let's Begin!

4. Write down the names of the 4 variables that contain the point-totals, or scores, for each personality color.

5. Write down the names of the variables that tell us an observation's birth gender and whether they participated in playing sports.

6. How many variables are in the data set?

7. How many observations are in the data set?

Estimating Centers

8. Create a dotplot of the scores for your predominant color and paste it here:

9. Based on your dotplot, which values came up the most frequently?

10. Based on your dotplot, about how many people in your class had a score similar to yours?

11. Based on your dotplot, what, would you say, was a typical score for a person in your class for your predominant color? How does your own score for this color compare?

Means and Medians

12. What is your predominant color? What is the mean score for this color?

13. What is the code that you can use to find the median of your predominant color?

1

14. What is the median of your predominant color?

15. Are the mean and median roughly the same? If not, use the dot9lot you made in the last slide to describe why.

Comparing birth_genders

16. What code can you use to create a dotplot of your predominant color faceted by gender?

17. Which measure of center is more appropriate for your predominant color? Why?

18. What code can you use to calculate the value that describes the center of each birth gender?

19. What is the center for your predominant color for males?

20. What is the center for your predominant color for females?

21. Do males and females differ in their typical scores for your predominant color? Answer this statistical question using your dotplot.

22. Assign the mean values to the name gender_means. Paste the result of gender_means here:

23. Run the code diff(gender_means). Paste the result here:

24. How much more/less did one birth gender score over the other for your predominant color?

Estimating Spread

25. Look at the spread of the dotplot you made for your predominant color then fill in the blank:

Data points in my plot will usually fall within ____ units of the center.

26. Which birth gender, if either, seem to have values that are more spread out from the center?

27. Calculate the MAD of your predominant color and paste it here:

28. Based on the MAD, which birth gender has more variability for your predominant color's scores?

On Your Own

29. Do boys and girls in your class differ in their color scores? Perform an analysis that produces numerical summaries and graphs. Then, write a few sentence interpretations that addresses this statistical question and considers the shape, center and spread of the distributions of the graphs your create.

References

Anderson, M. (2015). *Technology Device Ownership: 2015*. Pew Research Center. Retrieved from http://www.pewinternet.org/2015/10/29/technology-device-ownership-2015/

Association of College and Research Libraries. (2015). *Framework for information literacy for higher education.* Chicago, IL: American Library Association. Retrieved from http://www.ala.org/aclr/standards/ilframework

Au, W. (2011). Teaching under the new Taylorism: high-stakes testing and the standardization of the 21st century curriculum. *Journal of Curriculum Studies*, *43*(1), 25–45. https://doi.org/10.1080/00220272.2010.521261

Bang, M., Warren, B., Rosebery, A. S., & Medin, D. (2012). Desettling expectations in science education. *Human Development*, *55*(5–6), 302–318. https://doi.org/10.1159/000345322

Barr, D.A., Gonzalez, M.E., and Wanat, S.F. (2008). The leaky pipeline: factors associated with early decline in interest in premedical studies among underrepresented minority undergraduate students. *Academic Medicine 83*(5), 503-511.

Bhargava, R., Kadouaki, R., Bhargava, E., Castro, G., & D'Ignazio, C. (2016). Data Murals: Using the Arts to Build Data Literacy. *The Journal of Community Informatics*, *12*(3), 197–216.

Boaler, J. & Greeno, J. (2000). Identity, agency, and knowing in mathematical worlds, In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 45-82). Stamford, CT: Ablex.

Bollier, D. (2010). The promise and peril of Big Data. The Aspen Institute. Retrieved from http://www.aspeninstitute.org/sites/default/files/content/docs/pubs/The_Promise_and_Peril_of_Big_Data.pdf boyd, D., & Crawford, K. (2012). Critical Questions for Big Data. *Information, Communication & Society*, *15*(5), 662–679. https://doi.org/10.1080/1369118X.2012.678878

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Brickhouse, N. (1994). Bringing in the outsiders: Reshaping the sciences of the future. *Journal of Curriculum Studies*, *26*(4), 401–416. https://doi.org/10.1080/0022027940260404

Cain Miller, C. (2013, April 11). Data Science: The Numbers of Our Lives. *The New York Times*. Retrieved from http://www.nytimes.com/2013/04/14/education/edlife/universities-offer-courses-in-a-hot-new-field-data-science.html

Calabrese Barton, A. (1998). Reframing "Science for All" Through the Politics of Poverty. *Educational Policy*, *12*(5), 525–541.

Calabrese Barton, A., & Yang, K. (2000). The culture of power and science education: Learning from Miguel. *Journal of Research in Science Teaching*, *37*(8), 871–889. https://doi.org/10.1002/1098-2736(200010)37

Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating data literacy into information literacy programs: Core competencies and contents. *Libri, 63*(2) 123-134.

Carlone, H. B. (2004). The Cultural Production of Science in Reform-based Physics: Girls' Access, Participation, and Resistance. *Journal of Research in Science Teaching*, *41*(4), 392–414. https://doi.org/10.1002/tea.20006

Carlson, J., Fosmire, M., Miller, C. C., & Sapp Nelson, M. (2011). Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *Libraries and the Academy*, *11*(2), 629–657. https://doi.org/10.1353/pla.2011.0022

Chandler, A. (2015). The Myth That Links Poor Families to Fast Food. *The Atlantic*. Retrieved from https://www.theatlantic.com/national/archive/2015/09/cdc-fast-food-poor-families-american/406447/

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications*, *19*(2), 171–209. https://doi.org/10.1007/s11036-013-0489-0

Chen, Y., Chen, H., Gorkhali, A., Lu, Y., Ma, Y., & Li, L. (2016). Big data analytics and big data science: a survey. *Journal of Management Analytics, 3*(1), 1–42. https://doi.org/10.1080/23270012.2016.1141332

Chong, D. and Shi, H. (2015). Big data analytics: a literature review. *Journal of Management Analytics, 2*(3), 175-201.

Clarke, R. (2016). Big data, big risks. *Information Systems Journal*, *26*(1), 77–90. https://doi.org/10.1111/isj.12088

Cobb, P., & Hodge, L. L. (2010). Culture, identity, and equity in the mathematics classroom. In E. Yackel, K. Gravemeijer, A. Sfard(Eds.), *A journey in mathematics education research* (pp. 179–195). Dordrecht, The Netherlands: Springer.

Cobb, P., & Hodge, L. L. (2002). A Relational Perspective on Issues of Cultural Diversity and Equity as They Play Out in the Mathematics Classroom. *Mathematical Thinking and Learning*, *4*(2–3), 249–284. https://doi.org/10.1207/S15327833MTL04023

Cobb, P., Confrey, J., DiSessa, A., Lehrer, R., & Schauble, L. (2003). Design Experiments in Educational Research. *Educational Researcher*, *32*(1), 9–13. https://doi.org/10.3102/0013189X032001009

Cobb, P., Gresalfi, M., & Hodge, L. L. (2009). An interpretive scheme for analyzing the identities that students develop in mathematics classrooms. *Journal for Research in Mathematics Education*, *40*(1), 40–68. https://doi.org/10.4324/9780203879276

Cobb, P., Wood, T., Yackel, E., & McNeal, B. (1992). Characteristics of Classroom Mathematics Traditions: An Interactional Analysis. *American Educational Research Journal*, *29*(3), 573–604. https://doi.org/10.3102/00028312029003573

Couldry, N. (2014). Inaugural: A necessary disenchantment: Myth, agency and injustice in a digital world. *The Sociological Review*, *62*(4), 880–897. https://doi.org/10.1111/1467-954X.12158

Covarrubias, A. (2011). Quantitative intersectionality: A critical race analysis of the Chicana/o educational pipeline. *Journal of Latinos and Education*, *10*(2), 86-105.

D'Ignazio, C., & Bhargava, R. (2016). DataBasic: Design Principles, Tools and Activities for Data Literacy Learners. *Journal of Community Informatics*, *12*(3), 83–107.

Delgado Bernal, D. (1999). Chicana/o Education From the Civil Rights Era to the Present. In J. F. Moreno (Ed.), *The Elusive Quest for Equality: 150 Years of Chicano/a Education* (pp. 77-108). Cambridge: Harvard Educational Review.

Delpit, L. (1988). The silenced dialogue: power and pedagogy in educating other people's children. *Harvard Educational Review, 58*(3), 280-298.

Ebach, M. C., Michael, M. S., Shaw, W. S., Goff, J., Murphy, D. J., & Matthews, S. (2016). Big data and the historical sciences: A critique. *Geoforum*, *71*, 1–4. https://doi.org/10.1016/j.geoforum.2016.02.020

Ekbia, H., Mattioli, M., Kouper, I., Arave, G., Ghazinejad, A., Bowman, T., … Sugimoto, C. R. (2015). Big Data, Bigger Dilemmas: A Critical Review. *Journal of the Association for Information Science and Technology*, *66*(8), 1523–1545. https://doi.org/10.1002/asi

Elmborg, J. (2006). Critical information literacy: Implications for instructional practice. *The Journal of Academic Librarianship*, *32*(2), 192–199. https://doi.org/10.1016/j.acalib.2005.12.004

Enyedy, N., & Mukhopadhyay, S. (2007). *They Don't Show Nothing I Didn't Know: Emergent Tensions Between Culturally Relevant Pedagogy and Mathematics Pedagogy*. *Journal of the Learning Sciences* (Vol. 16). https://doi.org/10.1080/10508400701193671

Freire, P. (1970/1993). *Pedagogy of the Oppressed*. New York and London: The Continuum International Publishing Group Inc.

Gonzalez, H.B. and Kuenzi, J.J. (2012). *Science, Technology, Engineering, and Mathematics (STEM) Education: A Primer*. Congressional Research Service. Retrieved from https://fas.org/sgp/crs/misc/R42642.pdf

Gunter, G. A. (2007). Building student data literacy: An essential critical-thinking skill for the 21st Century. *Multimedia & Internet@Schools*, *14*(3), 24–28.

Gutiérrez, K. D., & Jurow, A. S. (2016). Social design experiments: Toward equity by design. *Journal of the Learning Sciences*, *25*(4), 565–598. https://doi.org/10.1080/10508406.2016.1204548

Halford, S., Pope, C., & Weal, M. (2013). Digital Futures? Sociological Challenges and Opportunities in the Emergent Semantic Web. *Sociology*, *47*(1), 173–189. https://doi.org/10.1177/0038038512453798

Harris, C. I. (1995). Whiteness as property. In K. Crenshaw, N. Gotanda, G. Peller, & K. Thomas (Eds.), *Critical race theory* (pp. 279-291). New York: New Press.

Hogenboom, K., Holler Phillips, C., & Hensley, M. (2011). Show Me the Data! Partnering with Instructors to Teach Data Literacy. In *In Declaration of Independence: The Proceedings of the ACRL 2011 Conference* (pp. 410–417). Philadephia, PA. Retrieved from https://www.ala.org/ala/mgrps/divs/acrl/conferences/confsandpreconfs/national/2011/papers/show_me_the_data.pdf

Hull, G. A., & Greeno, J. G. (2006). Identity and agency in nonschool and school worlds. In *The Informal Education Reader* (Vol. 249, pp. 77–97).

Hursch, D. (2005). The growth of high-stakes testing in the USA: accountability, markets and the decline in educational equality. *British Educational Research Journal*, *31*(5), 605–622. https://doi.org/10.1080/01411920500240767

IBM. (n.d.). What is big data analytics? Retrieved April 21, 2017, from http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics

Ignatow, G., & Robinson, L. (2017). Pierre Bourdieu: theorizing the digital. *Information, Communication & Society*, 1–17. https://doi.org/10.1080/1369118X.2017.1301519

Jacob, B. A. (2005). Accountability, incentives and behavior: the impact of high-stakes testing in the Chicago Public Schools. *Journal of Public Economics*, *89*(2005), 761–796. https://doi.org/10.1016/j.jpubeco.2004.08.004

Jones, B. & Gazzar, B. (2014, January 23). LAUSD iPad status looks good, school board vows ‘we are committed to this. We will move forward.’ *Huffington Post*. Retrieved from http://www.huffingtonpost.com/2013/11/06/lausd-ipad-status-good_n_4226209.html

Glaser, B., & Strauss, A. (1967). Grounded theory: The discovery of grounded theory. *Sociology The Journal Of The British Sociological Association*, *12*, 27-49.

Jones, B., & Gazzar, B. (2014). LAUSD iPad Status Looks Good, School Board Vows“ We Are Committed To This. We Will Move Forward.” *Huffington Post*. Retrieved from http://www.huffingtonpost.com/2013/11/06/lausd-ipad-status-good_n_4226209.html

Kahle, J.B. & Meece, J.L. (1994). Girls and science education: A developmental model. In D. Gabel (Ed.), *Handbook of Research in Science Teaching and Learning* (p. 1559-1610). Washington, D.C.: National Science Teacher Association.

Kaufman, T. E., Graham, C. R., Picciano, A. G., Popham, J. A., & Wiley, D. (2014). Data-driven decision making in the K-12 classroom. In J.M. Spector, M.D. Merril, J. Elen, M.J. Bishop, (Eds.), *Handbook of Research on Educational Communications and Technology* (p. 337-346). Springer New York.

Kearns, L.-L. (2011). High-stakes standardized testing and marginalized youth: an examination of the impact on those who fail. *Canadian Journal of Education*, *34*(2), 112–130.

Kellner, D. (1998). Multiple Literacies and Critical Pedagogy in a Multicultural Society. *Educational Theory*, *48*(1), 103–122. https://doi.org/10.1111/j.1741-5446.1998.00103.x

Kellner, D. (2003). Toward a Critical Theory of Education. *Democracy & Nature*, *9*(1), 51–64.

Kellner, D., & Share, J. (2007). Critical Media Literacy , Democracy , and the Reconstruction. In D. Macedo & S. R. Steinberg (Eds.), *Media Literacy: A Reader* (pp. 3–23). New York: Peter Lang Publishing. Retrieved from https://pages.gseis.ucla.edu/faculty/kellner/essays/2007_Kellner-Share-Steinberg and Macedo_ch1.pdf

Khalifa, M. A., Jennings, M. E., Briscoe, F., Oleszweski, A. M., & Abdi, N. (2014). Racism? Administrative and Community Perspectives in Data-Driven Decision Making: Systemic Perspectives Versus Technical-Rational Perspectives. *Urban Education*, *49*(2), 147–181. https://doi.org/10.1177/0042085913475635

Kohn, A. (2000). *The case against standardized testing : raising the scores, ruining the schools*. Portsmouth, NH: Heinemann. https://doi.org/10.2307/821673

Konold, C. & Pollatsek, A. (2002). Data analysis as the search for signals in noisy processes. *Journal for Research in Mathematics Education*, *33*(4), 259-289.

Krogstad, J.M. (2016). *5 facts about Latinos and education*. Pew Research Center. Retrieved from http://www.pewresearch.org/fact-tank/2016/07/28/5-facts-about-latinos-and-education/#

Lenhart, A. (2012). *Teens, smartphones & texting*. Pew Research Center. Retrieved from http://www.pewinternet.org/2012/03/19/teens-smartphones-texting/

Lenhart, A., Duggan, M., Perrin, A., Stepler, R., Rainie, L., & Parker, K. (2015). *Teens, Social Media & Technology Overview 2015*. Pew Research Center.

Love, N. (2004). Taking data to new depths. *National Staff Development Council*, *25*(4), 22–26. Retrieved from http://courseweb.hopkinsschools.org/pluginfile.php/63192/mod_resource/content/0/CFA/Data_to_new_depths.pdf

Lubienski, S. T. (2002). Research, reform, and equity in US mathematics education. *Mathematical thinking and learning*, *4*(2-3), 103-125.

Makar, K. & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal*, *8*(1), 82-85.

March, J. G. (2006). Rationality, foolishness, and adaptive intelligence. *Strategic Management Journal*, *27*(3), 201–214. https://doi.org/10.1002/smj.515

Margolis, J. & Suarez-Orozco, M. M. (2014, January 21). Handing Out iPads to LAUSD Students Isn't Enough. *Los Angeles Times*. Retrieved from http://www.latimes.com/opinion/commentary/la-oe-margolis-schools-computers-20140119,0,2183468.story#axzz30nxEFO86

Martin, D. (2000). *Mathematics success and failure among African-American youth: The roles of sociohistorical context, community forces, school influence, and individual agency.* Mahwah, New Jersey: Lawrence Erlbaum.

Martin, A., McAlear, F. and Scott, A. (2015). Path not found: Disparities in access to computer science courses in California high schools. Level Playing Field Institute.

Martin, A., McAlear, F., & Scott, A. (2015). *Path not found: Disparities in Access to computer science courses in California high schools*.

McAbee, S. T., Landis, R. S., & Burke, M. I. (2017). Inductive reasoning: The promise of big data. *Human Resource Management Review*, *27*, 277–290. https://doi.org/10.1016/j.hrmr.2016.08.005

McAfee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*, *90*(October), 1–9. https://doi.org/10.1007/s12599-013-0249-5

McAuley, D., Rahemtulla, H., Goulding, J., and Souch, C. (2014). How Open Data, data literacy and Linked Data will revolutionise higher education. Retrieved from http://pearsonblueskies.com/2011/how-open-data-data-literacy-and-linked-data-will-revolutionise-higher-education/

McCartney, P. (2015). Big Data Science. *The American Journal of Maternal/Child Nursing*, *40*(2), 130.

Mervis, J. (2010). Better intro courses seen as key to reducing attrition of STEM majors. *Science*, *330*(6002), 306-306.

Morris, E. W., & Perry, B. L. (2016). The punishment gap: School suspension and racial disparities in achievement. *Social Problems*, *63*(1), 68–86. https://doi.org/10.1093/socpro/spv026

Murrell Jr, P. C. (1999). Chartering the village: The making of an African-centered charter school. *Urban Education*, *33*(5), 565-583.

Museus, S. D., Palmer, R. T., Davis, R. J., & Maramba, D. C. (2011). *Racial and Ethnic Minority Students' Success in STEM Education. ASHE Higher Education Report*. https://doi.org/10.1002/aehe.3606

National Governors Association Center for Best Practices, Council of Chief State School Officers. (2010). Common Core State Standards (Statistics and Probability). National Governors Association Center for Best Practices, Council of Chief State School Officers, Washington D.C.

U.S. Department of Education. (2001). *The No Child Left Behind Act of 2001*.

Oakes, J. (1985) *Keeping track: How students structure inequality*. New Haven, NJ: Yale University Press.

Partnership for 21st Century Skills. (2015). P21 framework definitions. Retrieved from www.p21.org/storage/documents/docs/P21_Framework_Definitions_New_Logo_2015.pdf

Pérez Huber, L., Vélez, V. N., & Solórzano, D. G. (2014). *The growing educational equity gap for California's Latina/o students*.

Perrin, A., & Duggan, M. (2015). *Americans' Internet access: 2000-2015*. Pew Research Center.

Philip, T. M. (2017). Learning with mobile technologies: Considering the challenges, commitments, and quandaries. *Communications of the ACM*, *60*(3), 34–36. https://doi.org/10.1145/2976735

Philip, T. M. & Garcia, A. (2013a). The Importance of Still Teaching the iGeneration: New Technologies and the Centrality of Pedagogy. *Harvard Educational Review (83)*2, p. 300-319.

Philip T. M. & Garcia, A. (2013b, October 10). iFiasco in LA's Schools: Why Technology Alone is Never the Answer. *Dmlcentral: Digital Media + Learning: The Power of Participation*. Retrieved from http://dmlcentral.net/blog/thomas-m-philip-and-antero-garcia/ifiasco-las-schools-why-technology-alone-never-answer

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. https://doi.org/10.1080/07370008.2016.1210418

Philip, T. M., Rocha, J., & Olivares-Pasillas, M. C. (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), 1–21.

Philip, T. M., Schuler-Brown, S., & Way, W. (2013). A framework for learning about big data with mobile technologies for democratic participation: Possibilities, limitations, and unanticipated obstacles. *Technology, Knowledge and Learning*, *18*(3), 103–120. https://doi.org/10.1007/s10758-013-9202-4

Philip, T., & Olivares-Pasillas, M. C. (2016). Learning Technologies and Educational Equity: Charting Alternatives to the Troubling Pattern of Big Promises with Dismal Results. *Teachers College Record*. Retrieved from http://www.tcrecord.org/PrintContent.asp?ContentID=21616

Qin, J., & D'ignazio, J. (2010). The Central Role of Metadata in a Science Data Literacy Course. *Journal of Library Metadata*, *10*(2–3), 188–204. https://doi.org/10.1080/19386389.2010.506379

Raible, J., & Irizarry, J. G. (2010). Redirecting the teacher's gaze: Teacher education, youth surveillance and the school-to-prison pipeline. *Teaching and teacher education*, *26*(5), 1196-1203.

Rocque, M., & Paternoster, R. (2011). Understanding the antecedents of the "school-to-jail" link: The relationship between race and school discipline. *The Journal of Criminal Law and Criminology*, *101*(2), 633–665. https://doi.org/10.1017/CBO9781107415324.004

Rosebery, A. S., Warren, B., & Tucker-Raymond, E. (2016). Developing interpretive power in science teaching. *Journal of Research in Science Teaching*, *53*(10), 1571-1600.

Rubin, A. and Hammerman, J.K. (2006). Understanding data through new software representations. In G.F. Burrill (Ed.), *Thinking and reasoning with data and chance: Sixty-eighth NCTM yearbook* (p. 241-256). Reston, VA: National Council of Teachers of Mathematics.

Rumsey, D.J. (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education 10*(3), 6-13.

Seiler, G. (2013). New metaphors about culture: Implications for research in science teacher preparation. *Journal of Research in Science Teaching*, *50*(1), 104-121.

Sellar, S. (2015). Data infrastructure: a review of expanding accountability systems and large-scale assessments in education. *Discourse: Studies in the Cultural Politics of Education*, *36*(5), 765–777. https://doi.org/10.1080/01596306.2014.931117

Selwyn, N. (2015). Data entry: towards the critical study of digital data and education. *Learning, Media and Technology*, *40*(1), 64–82. https://doi.org/10.1080/17439884.2014.921628

Selwyn, N. (2016). Minding our language: Why education and technology is full of bullshit... and what might be done about it. *Learning, Media and Technology*, *41*(3), 437–443. https://doi.org/10.1080/17439884.2015.1012523

Selwyn, N., Henderson, M., & Chao, S. H. (2015). Exploring the role of digital data in contemporary schools and schooling - "200,000 lines in an Excel spreadsheet." *British Educational Research Journal*, *41*(5), 767–781. https://doi.org/10.1002/berj.3186

Shapiro, J.J. and Hughes, S.K. (1996). Information literacy as a liberal art: Enlightenment proposals for a new curriculum. *Educom Review 31*(2), 31-35.

Standaert, R. (1993). Technical Rationality in Education Management : A Survey Covering England, France and Germany. *European Journal of Education*, *28*(2), 159–175.

Stanley, W. B., & Brickhouse, N. W. (2000). Multiculturalism, Universalism, and Science Education. *Science Education*, *78*(4), 387–398.

Steen, L. A. (1999). Numeracy: The New Literacy for a Data-Drenched Society. *Educational Leadership*, *57*(2), 8–13.

Stephenson, E., & Schifter Caravello, P. (2007). Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project. *Reference Services Review*, *35*(4), 525–540. https://doi.org/10.1108/00907320710838354

Stepler, R. and Brown, A. (2016). *Statistical Portrait of Hispanics in the United States*. Pew Research Center: Hispanic Trends.

Strasburger, V.C. (2015). The new technology revolution: collaborative efforts between pediatricians, schools, and millennials for media education. In J. Frechette & R. Williams (Eds.), *Media Education for a Digital Generation* (p. 61-80). New York: Routledge.

Harris, C. I. (1995). Whiteness as property. In K. Crenshaw, N. Gotanda, G. Peller, & K. Thomas (Eds.), *Critical race theory* (pp. 279-291). New York: New Press.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* Newbury Park, CA: Sage Publications.Szymanski, M. (2015, December 11). LAUSD getting computers to all students at 103 schools. *LA School Report*. Retrieved from http://laschoolreport.com/lausd-gets-computers-for-every-student-at-103-pilot-schools/

Szymanski, M. (2015, December 11). LAUSD getting computers to all students at 103 schools. *LA School Report*. Retrieved from http://laschoolreport.com/lausd-gets-computers-for-every-student-at-103-pilot-schools/

Tygel, A. F., & Kirsch, R. (2015). Contributions of Paulo Freire for a critical data literacy. In *Web Science 2015 Workshop on Data Literacy*. Retrieved from www.dataliteracy.eita.org.br/wp-content/uploads/2015/02/Contributions-of-Paulo-Freire-for-a-critical-data-literacy.pdf van Dijck, J. (2014). Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveillance & Society*, *12*(2), 197–208.

Vikraman, S., Fryar, C. D., & Ogden, C. L. (2015). *Caloric Intake From Fast Food Among Children and Adolescents in the United States, 2011 – 2012*. Retrieved from https://www.theatlantic.com/national/archive/2015/09/cdc-fast-food-poor-families-american/406447/

Warren, B. and Rosebery, A.S. (1996). "This question is just too, too easy!" Students' perspectives on accountability in science. In L. Schauble & R. Glaser (Eds.), *Innovations in learning: New environments for education* (p. 97-125). Mahwah, NJ: Erlbaum.

Warren, B., & Rosebery, A. S. (2011). Navigating Interculturality: African American Male Students and the Science Classroom. *Journal of African American Males in Education*, *2*(1), 98–115.

White House Initiative on Educational Excellence for Hispanics. Retrieved from: http://www.ed.gov/edblogs/hispanic-initiative/

Yackel, E., & Cobb, P. (1996). Sociomathematical Norms, Argumentation, and Autonomy in Mathematics. *Journal for Research in Mathematics Education*, *27*(4), 458–477.

Yarbrough, B. (2014, April 12). How Common Core Education Standards are Changing the Way LAUSD Schools Test Children. *LA Daily News*. Retrieved from http://www.dailynews.com/social-affairs/20140412/how-common-core-education-standards-are-changing-the-way-lausd-schools-test-children

Zheng, N., Zhang, J., & Wang, C. (2014). Special issue on big data research in China. *Knowledge and Information Systems*, *41*(2), 247–249. https://doi.org/10.1007/s10115-014-0792-5

Zurkowski, P. G. (2013, December). Information Literacy Is Dead… Long Live Information Literacy. In *Worldwide Commonalities and Challenges in Information Literacy Research and Practice: European Conference, ECIL 2013, Istanbul, Turkey, October 22-25, 2013. Revised Selected Papers* (Vol. 397, p. 1). Springer.