"I LOVE MATH ONLY IF IT'S CODING": A CASE STUDY OF STUDENT EXPERIENCES IN THE INTRODUCTION TO DATA SCIENCE COURSE

ERICA HEINZMAN
University of California, San Diego
eheinzman@ucsd.edu

ABSTRACT

Many important voices—including The National Council for Teachers of Mathematics (NCTM), the Dana Center's Launch Years initiative, and others—advocate for expanding the traditional course offerings in high school mathematics and statistics to include courses such as the Introduction to Data Science (IDS). To date, the research on the IDS course has primarily focused on pedagogy, professional learning for teachers, and the curriculum. This mixed-methods case study expands our understanding by analyzing the perspective of IDS students at a California public high school. Self-determination theory provides a useful frame for interpreting how these students experience the IDS course. The theory focuses on conditions for students to engage in meaningful learning: competence (self-efficacy), autonomy (agency), and relatedness (a sense of belonging). The findings from this case study suggest the IDS students feel confident, empowered, and part of a vibrant community, unlike previous mathematics and statistics courses they may have completed; and use specific language to describe their joy in problem-solving and the accessibility of the course. These findings have implications for the development and refinement of any high school data science course, including IDS.

Keywords: *Statistics education research; Data science education; Non-traditional high school mathematics; High school mathematics pathways*

1. INTRODUCTION

There is a national movement to create multiple pathways for high school mathematics and statistics that reflect the interests of students, address inequities and access in course-taking, and include new and dynamic content (California Math Council [CMC], 2020; Charles A. Dana Center [Dana Center], 2020; Daro & Asturias, 2019; Fitzpatrick & Sovde, 2019; Moussa et al., 2020; National Council of Teachers of Mathematics [NCTM], 2018). Data science has generated much interest from policymakers, researchers, and professional organizations as a rigorous alternative to the traditional algebra to calculus pathway (California Department of Education [CDE], 2022; Dana Center, 2020; Daro & Asturias, 2019). The pathway movement builds on decades of research into the content and pedagogy of K-12 mathematics and seeks to address the pressing issues of equity and the ever-changing needs of the labor market and our world (Burdman, 2015; Business–Higher Education Forum and PwC [BHEF], 2017; Committee on STEM, 2018; Dana Center, 2020; Daro & Asturias, 2019; NCTM, 2018). Although the impetus for creating these pathways is to make learning mathematics and statistics more meaningful and relevant to students, little is known in the research literature about student experiences in non-traditional pathways such as data science.

2. BACKGROUND

Prominent mathematics organizations have issued a call to action, which recommends giving agency to high school students to select a mathematics pathway that aligns with their personal and professional interests (Dana Center, 2020; Daro & Austorias, 2019; NCTM, 2018). The traditional high school pathway from algebra to calculus may hold little relevance for many students. Therefore, scholars and industry leaders posit modeling, statistics, and computational thinking may be more useful

Statistics Education Research Journal, 21(2), Article 5. https://doi.org/10.52041/serj.v21i2.43

for a wide variety of careers and may pique students' interests. Notably, data science combines all three elements (Arnold et al., 2020; BHEF, 2017; Burdman, 2015; Carnevale & Desrochers, 2003; Dana Center, 2020; Gould, 2010; NCTM, 2018; Wilkerson & Pollard, 2020). Multiple high school pathways offer an opportunity to expand course offerings beyond the narrow focus on algebra and geometry. States have begun to enact these recommendations. For example, under Oregon's *2+1 course model,* high school students select from one of three pathways ending in calculus, data science, or quantitative mathematics (CDE, 2020; CMC, 2020; Dana Center, 2020; Moussa et al., 2020; Oregon Department of Education [ODE], 2020).

Advocates contend multiple high school pathways can address persistent issues of equity and access, including the need for more diverse graduates in science, technology, engineering, and mathematics (STEM) fields (Burdman, 2015, 2018; Dana Center, 2020; Daro & Asturias, 2019; NCTM, 2018). High school mathematics, including statistics, can serve a crucial role in attracting young adults to STEM-related careers and fostering a life-long appreciation of mathematics (Burdman, 2018; Carnevale & Desrochers, 2003; Committee on STEM, 2018; NCTM, 2018; President's Council of Advisors on Science and Technology [PCAST], 2012), yet pernicious inequities prevent and marginalize too many students, especially Black, Indigenous, and Students of Color from enrolling in higher-level mathematics and statistics courses (Burdman, 2015, 2018; NCTM, 2018; U.S. Department of Education Office of Civil Rights, 2018). Offering non-traditional mathematics pathways, such as data science, have the potential to increase access to rigorous courses and may engage high school students who previously never developed an affinity for mathematics.

Data science's emerging popularity as a high school pathway coincides with the growing recognition of how data pervades nearly all aspects of daily life (Arnold et al., 2020; Dana Center, 2020; Daro & Asturias, 2019). The *Guidelines for Assessment and Instruction in Statistics Education* report, referred to as GAISE II, advocates for Pre-K to high school seniors to develop proficiency in interpreting and critiquing data (Arnold et al., 2020). The Business-Higher Education Forum (2017) posits the number of jobs requiring knowledge of data science is expected to grow in all industries as data becomes more integral to our lives. Nearly a decade ago, the *Introduction to Data Science* (IDS) course was first launched in the Los Angeles Unified School District. Now IDS is being offered in 45 school districts across four states (Dana Center, 2020; Daro & Asturias, 2019; Gould et al., 2016; UCLA, 2022a). This course offers a unique opportunity to examine student experiences in a non-traditional pathway.

The IDS course is designed as an advanced high school mathematics course to be taken after students have completed at least two years of high school mathematics (Gould et al., 2016; UCLA, 2022b). To teach the yearlong IDS course, teachers participate in extensive professional learning, which prepares them to use an inquiry-based approach and to support students programming in R. The IDS course consists of descriptive and inferential statistics, data visualizations and collection methods, and using models to make predictions (UCLA, 2022b). Gould and collaborators (2016) describe the IDS curriculum as, "consist[ing] of classroom activities and discussions, computer lab exercises in which students work collaboratively to learn to use the statistical programming language R via *RStudio* (RStudio team 2015), and participatory sensing campaigns" (p. 3). In participatory sensing campaigns, students collect and analyze data directly related to their lives. For example, students record their stress levels over a few days in the Stress-Chill campaign and then apply their learning from the unit to analyze the class data set (Gould et al., 2016; UCLA, 2022b.).

Research supplies the *why* for multiple high school pathways, yet there is little to no research on *how* students experience non-traditional pathways (Burdman, 2018; Dana Center, 2020; Daro & Asturias, 2019; Gould, 2010; Gould et al., 2016; NCTM, 2018). The little published research on the IDS course focuses on instructional practices and content (e.g., Gould et al., 2016; Gould, Bargagliotti, & Johnson, 2017), not on the student experience in the course. Without research into student experiences in non-traditional courses such as IDS, it is unknown if the intentions are being achieved. This research is all the more vital as multiple states move toward implementing data science pathways and more high school data science courses are developed by UCLA Psychology Department's Teaching and Learning Lab (CourseKata) and YouCubed at Stanford University to name a few (CDE, 2022; CourseKata, n.d.; Dana Center, 2020; Moussa et al., 2020; ODE, 2020; YouCubed, 2020).

3. METHODOLOGY

This research utilized a mixed methods case study design. Mixed methods build on the individual strengths of quantitative and qualitative methods. Quantitative methods capture the perceptions of a larger set of participants, while the qualitative methods flesh out the *how* and *why* of those perceptions from a carefully selected subset of participants. Together, they form a more thorough understanding of the phenomena (Creswell, 2012; Teddlie & Tashakkori, 2009). A case study offers a boundary to the research—in this case, one IDS classroom (Creswell, 2012; Stake, 2005).

3.1. THEORETICAL FRAMEWORK

Self-determination theory provides a useful frame for examining the experiences of students enrolled in the IDS course. The theory explores the inner workings of motivation (Deci & Ryan, 2008; Deci et al., 1991; Guay et al., 2008; Ryan & Deci, 2000a, 2000b). Self-determination theory conceives of motivation along a continuum from amotivation (disengagement) to intrinsic motivation (self-directed) with extrinsic motivation situated in the center. Positive motivation, whether intrinsic or extrinsic, is achieved when the individual finds meaning in the work and the three psychological needs of competence, autonomy, and relatedness are met (Deci & Ryan, 2008; Ratelle et al., 2007; Ryan & Deci, 2000a, 2000b). Competence is characterized by feelings of confidence and efficacy to engage in challenging tasks. Autonomy describes a sense of agency, a person acting freely regardless of any external pressures. Finally, a sense of belonging and being part of the community define relatedness. All three psychological needs touch upon vital aspects of a data science classroom. Ideally, students experience confidence, agency, and a sense of community when engaging in the course (Deci & Ryan, 2008; Ryan & Deci, 2000a, 2000b).

Studies show the three psychological needs of competence, autonomy, and relatedness have a profound impact on education. Students who feel a stronger sense of competence, autonomy, and relatedness demonstrate more persistence, have higher achievement, prefer intellectual challenges, and find more enjoyment in academic pursuits (Deci et al., 1991; Guay et al., 2008; Ratelle et al., 2007; Ryan & Deci, 2000a, 2000b). Studies on mathematics education outside of the self-determination field also identify these three psychological needs as essential, citing how the lack of competence, autonomy, and/or relatedness marginalizes and drives students from the discipline (Boersma & Savina, 2019; Bressoud, 2015; Good et al., 2012; Ladson-Billings, 1997).

Prominent frameworks for teaching K–12 mathematics allude to the three psychological needs of competence, autonomy, and relatedness from self-determination theory (Ryan & Deci, 2000a, 2000b). For example, Aguirre and colleagues' (2013) equity-focused teaching framework posits the need for teachers to facilitate competence, autonomy, and relatedness in students by valuing students' strengths, encouraging multiple approaches to problem-solving, and building an inclusive community. Schoenfeld's (2014) *Teaching for Robust Understanding* framework implies the three needs in the fourth dimension titled, "agency, authority, and identity" (p. 407). Agency entails autonomy. Authority infers competence and identity suggests the students viewing themselves as members of a community.

3.2. CASE STUDY SELECTION, CONTEXT, AND PARTICIPANTS

The one IDS classroom for the case study was selected in consultation with the Director of the Introduction to Data Science Project at UCLA Center X. The program director recommended Ms. Arcega who had taught the IDS course for multiple years and served as a mentor and trainer for new teachers of the course (S. Machado, personal communication, October 16, 2019). Pseudonyms were assigned to the school, teacher, and all students in reporting the findings of this study.

Ms. Arcega taught at Sierra Madre High School, a high school of 2,000 students in a populous region in Southern California. Among the high schools in the district, Sierra Madre had the highest percentage of socioeconomically disadvantaged students (80.1 percent) and a lower percentage of English learners (20 percent) than the district overall (CDE, 2017). The school was located in an older residential neighborhood. Some of the homes had fallen into disrepair and others had been razed to make way for larger homes and two-story apartment complexes. A short distance from the school was an historic commercial area. The ground-level businesses reflected diverse communities, with many restaurants serving regional cuisines from Asia and Latin America alongside familiar national chains. These communities were reflected in the demography of Sierra Madre High; 94 percent of the student population is Asian (59.5 percent) and Hispanic (33.9 percent) (CDE, 2017).

Ms. Arcega had taught a full line of high school statistics in all its variations—Statistics, AP Statistics, and IDS—and had taught the IDS course for the past four years. Ms. Arcega's IDS class was small and predominantly male. Nineteen IDS students out of 23 consented to participate in the research study, self-reporting demographic and academic data (see Table 1). Similar to the overall school population, 84 percent of the surveyed IDS students self-identified as either Latinx or Asian. However, unlike Sierra Madre High School as a whole, there was a higher percentage of Latinx (47 percent) than Asian (37 percent) students. Seventy-nine percent of participating students qualified for the federal lunch program closely reflective of the 80 percent of the school population identified as socioeconomically disadvantaged (CDE, 2017). The eight IDS focus group students (see Table 1) reflected the gender imbalance and federal lunch program status of the IDS class.

Table 1. Demographics of IDS class and student focus group participants

	IDS Class (n = 19)	IDS Focus Group (n = 8)
	Gender	
Male	14 (74%)	6 (75%)
Female	5 (26%)	2 (25%)
	Ethnicity/Race	
Latinx	9 (47%)	5 (63%)
White	1 (5%)	1 (12%)
Asian	7 (37%)	2 (25%)
Multiracial	2 (11%)	0 (0%)
	Federal Lunch Status	
Qualify for Free or Reduced Lunch	15 (79%)	6 (75%)
	Grade Level	
Juniors	6 (32%)	4 (50%)
Seniors	13 (68%)	4 (50%)
	Cumulative Unweighted GPA Range	
Below 2.0	3 (16%)	1 (12%)
2.0 to 2.5	6 (32%)	2 (25%)
2.5 to 3.0	5 (26%)	3 (38%)
3.0 to 3.5	3 (16%)	1 (12%)
3.5 to 4.0	2 (10%)	1 (12%)

Note: Students selected a range that best matched their cumulative unweighted GPA.

Survey results indicated IDS students generally earned a higher fall semester grade than their cumulative unweighted high school GPA. Fifty-eight percent of the IDS students surveyed earned at least a B in the IDS course for their fall semester grade, while only 26 percent reported a cumulative unweighted GPA of at least 3.0. Most of the surveyed IDS students (79 percent) enrolled in the course after finishing either Integrated Mathematics II or III. The other 21 percent completed statistics before enrolling in the IDS course. The eight IDS students who participated in the focus groups represented all three pathways into the course and closely mirrored the cumulative unweighted GPA ranges of the surveyed class (see Table 1).

The learning environment in Ms. Arcega's IDS class was marked by camaraderie, collaboration, and curiosity. The single-person desks are grouped in pods of four throughout the room, which fostered student interactions, whether it is exchanging light-hearted banter, debugging lines of code, or debating conjectures. Ms. Arcega often generated curiosity in the students using the context of lessons. For example, Ms. Arcega had student groups engage in a lively discussion about causes of stress in their lives and write a list of questions they had (e.g., what time of day am I most stressed), before introducing the Stress-Chill campaign.

3.3. DATA COLLECTION

This research study had two phases. Data collected in the first phase informed the data collected in the second phase (Creswell, 2012). Phase 1 was classroom observations followed by a student survey. The second phase built on the areas of interest from Phase 1 and featured semi-structured student focus groups and an interview with the IDS teacher. The teacher interview was sandwiched between the student focus groups. This way, the IDS teacher shed light on emergent themes from the analysis, yet the final student focus group had the last word so to speak, expounding on themes that emerged from previous interviews with their peers and teacher.

Students completed more than half the IDS course when data collection began in January 2020, which was sufficient time for them to form a perspective on their experiences in the class. Data collection concluded before the school closures for the pandemic. The classroom observations, a survey, and interviews generated a qualitative data set including field notes, interview transcripts, and analytical memos. A preliminary analysis of data from the first phase informed the purposeful sampling of students for the focus groups and some of the questions asked in the interviews.

Observations. During the first phase, I purposefully selected three separate occasions for observations to gain insight into the variety of experiences students have in the IDS course from classroom activities to programming in R (Gould et al., 2016; Emerson et al., 2011; Maxwell, 2012). During my first observation, IDS students used R to find the probability of “which song will play next” (Lab 2c). On another occasion, Ms. Arcega sparked curiosity and built anticipation by having students reflect on stress in their lives before introducing the Stress-Chill participatory sensing campaign. During my third observation, students explored and debated how they might determine if different factors affected Titanic passengers’ chance of survival as an introduction to Lab 2F (see *IDS Curriculum* [UCLA, 2022b] for an overview of the lessons).

Survey. After three observations in the IDS class, the teacher invited students, who submitted the necessary permissions, to complete an electronic survey via Qualtrics (see Appendix to view the Likert-type scale items from the survey). The survey had common demographic questions and adapted items from three different instruments designed to address the three psychological needs of self-determination theory: competence, autonomy, and relevance (Ryan & Deci, 2000a, 2000b). The items addressing the psychological needs of competence and autonomy are adapted from two scales on mathematics attitude, Shortened Form of the Fennema-Sherman Mathematics Attitude Scales (Mulhern & Rae, 1998) and Attitudes Toward Mathematics Inventory (Tapia & Marsh, 2004). The instruments were created for Irish school children (Mulhern & Rae, 1998) and secondary students in the United States (Tapia & Marsh, 2004). The relatedness items were adapted from Good et al. (2012) survey on a sense of belonging in mathematics. I updated the formal language of the established instruments for current high school students and narrowed the focus to their experiences in the IDS course. Students responded to these items using a 5-point Likert-type scale ranging from *strongly disagree* to *strongly agree*. None of the instruments had items addressing specific aspects of autonomy and relatedness, such as self-direction on starting problems (autonomy) or forming connections with their peers in the course (relatedness); therefore, I included additional items to capture these subtleties. Finally, the survey had one open-ended item, which prompted students to describe their perceptions of mathematics in the IDS course and was analyzed with the other qualitative data.

Semi-structured focus group and interview. The semi-structured approach to the focus groups and interview provides structure and flexibility to the interviewer. There was a predetermined set of questions to gather information on student experiences, paying particular attention to the three components of self-determination theory. The semi-structured nature provided flexibility to ask additional follow-up questions based on unanticipated responses of interest during the interviews and emergent themes from prior data collected (Creswell, 2012; Maxwell, 2012; Siedman, 2006). I recorded the student focus groups and the teacher interview for transcription. I conducted two semi-structured focus groups, with eight students. I situated the IDS teacher interview in between the two student focus groups, keeping the perspectives of the students at the forefront.

The structure of a focus group has benefits and limitations. Student participants deepened their responses by building on what their peers shared. Being with friendly peers may have alleviated some anxiety students felt about participating. However, students may have felt reluctant to share in front of their peers or may have limited time to share (Creswell, 2012). To partially mitigate these drawbacks, the focus groups were limited to a handful of students and careful consideration went into the composition of each focus group. Every effort was made to capture a diversity of perspectives among the students who submitted consent forms. I relied on my preliminary analysis Phase I data, teacher recommendations, and demographics to purposefully sample the pool of consenting students (Creswell, 2012; Maxwell, 2012).

3.4. DATA ANALYSIS

Preliminary data analysis occurred iteratively during the two phases to inform subsequent data collection. After the conclusion of the second phase, I analyzed the corpus of data in its entirety.

Qualitative data analysis. The two phases of data collection resulted in a trove of field notes, memos, and interview transcripts for analysis. During the observations of the IDS class, I attended to descriptions and dialogue of student experiences in field notes and memos for later coding (Emerson et al., 2011). I also wrote memos after each focus group and teacher interview focusing on intriguing themes from the sessions (Maxwell, 2012).

Thematic coding of the qualitative data was both open and focused on etic and emic themes (Emerson et al., 2011; Saldaña, 2013). I utilized the coding strategies of predetermined, descriptive, versus, emotion, and in vivo (Miles et al., 2014; Saldaña, 2013). I derived the predetermined codes of competence, autonomy, and relatedness from the self-determination framework (Miles et al., 2014; Ryan & Deci, 2000a, 2000b). Emic codes emerged beyond the self-determination framework that enriched my interpretation and analysis of how the students conceptualized data science in this course. Descriptive codes focus on topics, such as the students' perceptions of creativity and freedom in the IDS course. The versus coding technique was especially helpful in highlighting the difference between students' experiences in previous mathematics and statistics classes and the IDS course. I gained further insight into this dichotomy using emotion coding and focusing on the students' descriptions of their mathematics trajectories in high school. Finally, I noticed the common terms of *fun* and *easy* held deeper meanings for the students and explored these nuances using in vivo coding (Miles et al., 2014; Saldaña, 2013).

The coding of the data was an iterative process done using MAXQDA software. I recorded emergent ideas and impressions in memos, which informed future iterations of coding and inquiry (Emerson et al., 2011; Miles et al., 2014; Saldaña, 2013). To reduce the potential for bias, I engaged in peer debriefing with colleagues of de-identified data to uncover any emergent themes or interpretations that I may have missed (Lincoln & Guba, 1985).

Quantitative data analysis. The electronic survey collected demographic information and student self-ratings on items regarding competence, autonomy, and relatedness in the IDS course (see Appendix). The students responded to the items about their mathematics experiences using a 5-point Likert-type scale from *strongly agree* (1) to *strongly disagree* (5). I analyzed the ordinal and categorical data from the survey using JASP statistical software, developed at the University of Amsterdam. The Likert-type scale items on the survey were categorized into the three psychological needs from self-determination theory: competence, autonomy, and relatedness (Ryan & Deci, 2000a, 2000b). Through principal component analysis (PCA), I confirmed which Likert-type scale items clustered around their related element of self-determination theory. I generated a composite score for competence, autonomy, and relatedness from the items in each category with a component coefficient threshold of at least 0.5, and disregarded the items which fell below this threshold. I selected the component coefficient threshold of 0.5 to increase reliability given the small sample size of 19 students (DeVellis, 2017; de Winter, Dodou, & Wieringa, 2009; Price et al., 2015). I investigated the descriptive statistics of competence, autonomy, and relatedness composite scores by analyzing the measures of center and spread. To identify the relationships that existed between demographic factors (e.g., gender, race/ethnicity, first-semester grades, etc.) and the three psychological needs of self-determination theory, I performed the appropriate t-test, one-way ANOVA test, Mann-Whitney test, or Kruskal-Wallis test (Price et al., 2015; Ryan & Deci, 2000a, 2000b).

4. RESULTS AND DISCUSSION

This case study analyzed how IDS students experienced the three psychological factors of self-determination theory: competence, autonomy, and relatedness (Ryan & Deci, 2000a, 2000b). While these findings are not generalizable, they provide insights into how the participating IDS students perceived data science and coding. The IDS students derived a sense of competence or self-efficacy from the purposefulness of coding and use the descriptor *easy* to highlight the accessibility of the course. They gained a sense of autonomy or agency from the freedom they had to display and interpret data, frequently describing data science as *fun.* Finally, the IDS students perceived relatedness or a sense of belonging from the uniqueness of the course and frequently noted how it compared with their prior experiences in mathematics and statistics, namely detachment and failure. However, these comparisons rest solely on students' recollections of their previous courses, since additional information (e.g., pedagogical approach) is beyond the scope of this study.

4.1. COMPETENCE IN THE IDS COURSE

Useful and purposeful. Previously, IDS students found high school mathematics and statistics to be onerous and tedious, especially memorization of formulas. In contrast, they viewed the IDS course, in particular coding, as useful and necessary for the challenging data science labs. Connor, a Latinx senior, encapsulated these sentiments when he declared, "I love math only if it's coding math now." Grace and Catherine found memorizing the codes useful, unlike their prior experiences of memorizing formulas in mathematics classes. Grace, an Asian senior, complained, "I would say math … is quite hard for all the formulas you had to remember like from elementary till now," musing, some of the stuff you don't really need in the future." But knowing codes in RStudio, Grace found useful, "For IDS everything is on the computer and you just to take down the code …. You just got to know what it means and it's done." Similarly, for Catherine, a Latinx junior, the purpose of memorizing formulas was elusive, "I feel it [previous mathematics] was too much things that I feel I don't need in life. Like a whole bunch of formulas, you had to memorize." Yet Catherine found meaning in memorizing codes, "And with IDS ... you just memorize codes, and the codes aren't that hard to memorize because it's all the same codes throughout the whole year .... You learn new codes on the way, but I still use codes that I learned from the first day of school." Connor even found the precision required for coding useful, "If you miss a comma … you can't have a proper graph. If you miss a parenthesis, you can't have [a] proper graph. If you misspell something, it won't work. You have to be right 100% of the time for it to work. That's what I like." As indicated, any omitted or incorrect character will make a code inoperable.

***The meaning of* easy.** The IDS students used the word *easy* 20 times and associated it with the accessibility of the content. For example, Grace viewed her previous mathematics courses, including statistics, as nearly impenetrable, :For regular math, it's a harder way to learn so it's more difficult for me, and I can't really understand, it's hard to understand. While IDS it's really easy to understand." Grace attributed her success to the nature of data science, "It was much easier, all I had to do was remember all the codes and what they meant. I didn't have to remember any formulas or anything. I didn't have to put all the numbers down in a calculator. I didn't have to write anything down, I just had to type it down and the computer does it for me, which was way easier than all those other stuff." Dismissing hallmarks of her previous courses—memorizing formulas, entering data in the calculator, and writing notes—Grace declared data science *easier*. She found purpose and meaning that were lacking in her previous courses.

Will and Caleb used *easy* to allude to their newfound confidence in the IDS course, which stemmed from an understanding of coding. Admitting his previous disengagement from mathematics, Will, an Asian senior, mused, "So my past math experiences were kind of bad. I hated math [including statistics]. You know, I was the type of person who would slack off on homework, fail my test. Yet, he described the IDS class as satisfying, "I … started using coding and that was a lot easier for me to understand …. Taking like an hour on like one question and then finding an answer to that. It's quite nice."

Similarly, Caleb, a Latinx junior, previously failed a high school mathematics course but has gained a sense of confidence coding in data science, "I failed it [Integrated Mathematics I] …. I'm not very good …. But somehow, like for coding, it's been explained to me in a way where it's easy for me to understand and for me to do." Caleb described how he problem-solved with coding, "There's like a categorical problem, or like a numerical problem, … you can choose different plots to help you solve [them] …. We have a wide variety of things like histogram, bar graphs, XY plots." The IDS students employed the word *easy* to express their newfound feelings of competence in the course. In Will's case, he persevered in problem-solving and Caleb relished the freedom he had in creating visual displays of data.

4.2. AUTONOMY IN THE IDS COURSE

Empowering and engaging. In the IDS course, students expressed feeling empowered and connected to the curriculum, fostering a sense of autonomy. Nolan, a Latinx junior, valued the freedom he had in making representations of the data, "I don't have to follow one way to get this answer, I like how I can just change it up a bit. I like how I can use a histogram or maybe a bar graph or like tally or a XY plot [for] my answer." Similarly, Catherine found empowerment in executing the codes to create visualizations of the data, "One interesting thing that I like with the coding, I like how you could change the colors of the graphs … like every piece of code can do a big thing with the graph." Although Connor previously failed statistics, his sense of curiosity empowered him to learn more advanced concepts, "I ask[ed] Miss, our teacher, about like hard graphs if there were any. And she gave me a paper, and it was about GG plots. So I started messing around, and I got all the possible plots, somehow."

Underlying the interest in data science is the intriguing context of the IDS data sets. Students can be both a *consumer* and *producer* of data, bringing unique relevancy to the IDS course (Gould, 2010, p. 310). The IDS students explored pre-existing data sets (e.g., Titanic passenger logs) and created their own (e.g., Stress-Chill campaign). The IDS students expressed genuine interest in the context-driven, challenging labs and had the freedom to explore these data sets using coding. Grace fondly recalled both the context and the crux of the lab investigations, "I thought some of the labs that we did were really cool. We did one with songs and another with horror movies," and continued to speculate on why males may perish more frequently in horror movies. Joseph, a Latinx junior, perked up as he described the participatory sensing lab about his and his classmates' stress levels, "Stress and chill …. It was one of the labs that we had to go home and at a specific time, we would record what we think our stress level was. Then we would work with the data that we all put in." He was fascinated by a data set generated from his peers' stress levels and what he could learn using the skills he gained in the IDS course. Connor valued the opportunity he had to explore a data set of personal interest, "We uploaded our own data onto R Studio. We were asking questions that we want to know. Like, for me, I think of weight training, powerlifting. I was wondering like if age had anything to do with how much weight they pick up and specific workouts." Connor chose a database on weightlifting to explore his hobby using the techniques he learned in the IDS course. The given and curated contexts intrigued the students and let them engage meaningfully in data science.

***The meaning of* fun.** The IDS students used the word *fun* to capture their agency and engagement, or autonomy, in data science. The word *fun* appeared near hallmarks of the IDS course such as computers, coding, and graphical displays. Liam, a white senior, noted his initial uncertainty about the IDS course, "When I first came to IDS, I was a little bit nervous ... It was ..., kind of fun because like, it's not like the typical math classes. It's ... math on the computers. So that was a kind of interesting to me." Caleb compared coding to solving puzzles using the adjective *fun*, "I like puzzles. I like putting things together and coming up with something. So like, coding has been really fun for me."
Catherine associated *fun* with coding and doing data science, "So, I took IDS and I don't regret it at all. It's been a cool experience and it's nothing compared to IM3 [Integrated Math III] .... [IDS] It's all coding and it's really interesting and fun."

4.3. RELATEDNESS IN THE IDS COURSE

A vibrant community. The IDS students perceived themselves as being part of a vibrant community of learners. This sense of community or relatedness stemmed from pride in being some of the first to enroll in the course and an appreciation for the collective wisdom of their peers. The students perceived the IDS course as novel for two reasons. It was only offered at Sierra Madre High School for a handful of years and felt like a sharp departure from their previous mathematics courses with an emphasis in programming. The IDS course was developed less than a decade ago at the University of California, Los Angeles in partnership with Los Angeles Unified School District (Gould et al., 2016; Personal Communication with Ms. Arcega, December 16, 2019; UCLA, 2022).

The students felt they were participating in a groundbreaking experience by taking the IDS course, which translated into a feeling of community and relatedness in the IDS classroom. Nolan sensed his enrollment in the course was historic, “This class [IDS] is just starting and we’re going to be part of something very big. And I’m really grateful for the opportunity of being something that’s going to be going big later on.” Connor appreciated the camaraderie from being part of a new course, “So we’re learning as we go. It’s the best part for me.”

The paradigm of the teacher as the holder of all knowledge (Harel & Rabin, 2010) is unsustainable in data science since the students engage in “investigations in which learners pose questions, obtain data, and communicate findings” (Wilkerson & Polman, 2020, p. 5). The IDS students frequently depicted a supportive community in the IDS classroom. Within this community, the IDS students viewed their peers as capable doers of mathematics. Nolan valued learning from his peers, even forming new friendships, “I love the aspect of working together in groups of two or more. And sometimes I don’t understand it, I’ll be like, ‘Hey Will [another IDS student], will you help me out?’ And so I really like that because you get to convers[e] with someone new, make new friends.” Catherine acknowledged how the novelty of the IDS curriculum levels the playing field, “It’s good because everyone in that class could help you because we’re all at the same level. We’re all beginners into IDS, so we’re at the same level and we could really help each other. And I feel like, if I’m ever stuck, I have multiple people to go to.” Ms. Arcega concurred, “They [IDS students] ... know who needs more help … and they help each other out.” The sense of community or relatedness extended beyond collaboration to genuine concern for each other. Grace noted, “For us, we’re all kind to each other, we help one another and we give ideas to each other.” Connor observed, “So yeah, we do get to collaborate a lot in this class, and it does give us a chance to kind of worry about each other and help.”

4.4. MEASURES OF COMPETENCE, AUTONOMY, AND RELATEDNESS

IDS student survey data sheds light on the degree to which students perceived competence, autonomy, and relatedness. On the survey, 6 competence, 10 autonomy, and 10 relatedness items passed the principal component analysis coefficient threshold of 0.5, resulting in a composite score for each factor (see Table 2). The quantitative data reflected a positive sense of competence, autonomy, and relatedness expressed by the students in the focus groups. Both the means and medians for the three factors are close to 2.0 or *agree*. The standard deviations and interquartile ranges (IQR) are 1.0 or less, indicating many of the student responses fell between *strongly agree* (1) and *neutral* (3).

Table 2. Statistics of Competence, Autonomy, and Relatedness Composite Scores (n = 19)

	Mean	Median	Std. Deviation	IQR	Range
Competence Composite Score	2.079	2.000	0.802	1.000	3.000
Autonomy Composite Score	2.053	1.900	0.654	0.800	2.600
Relatedness Composite Score	2.189	2.200	0.706	0.800	2.300

Note: The Likert-type scale ranged from *strongly agree* (1) to *strongly disagree* (5).

Visually, the histograms for the competence, autonomy, and relatedness composite scores show a similar trend (Figure 1). The majority of the data lie on the agreement side of *neutral* (3.0), with the tallest bin being *strongly agree* (1.0) for all three factors. Competence has the most variation with one score in the 4.0 (*disagree)* bin.

There is, however, less power in the small sample (19 students), making it hard to detect statistically significant differences. Thus, not surprisingly, no statistical differences were detected between academic and demographic factors of the IDS students (e.g., first-semester grade in IDS, ethnicity, grade level, gender, etc.) and their composite ratings of competence, autonomy, and relatedness.

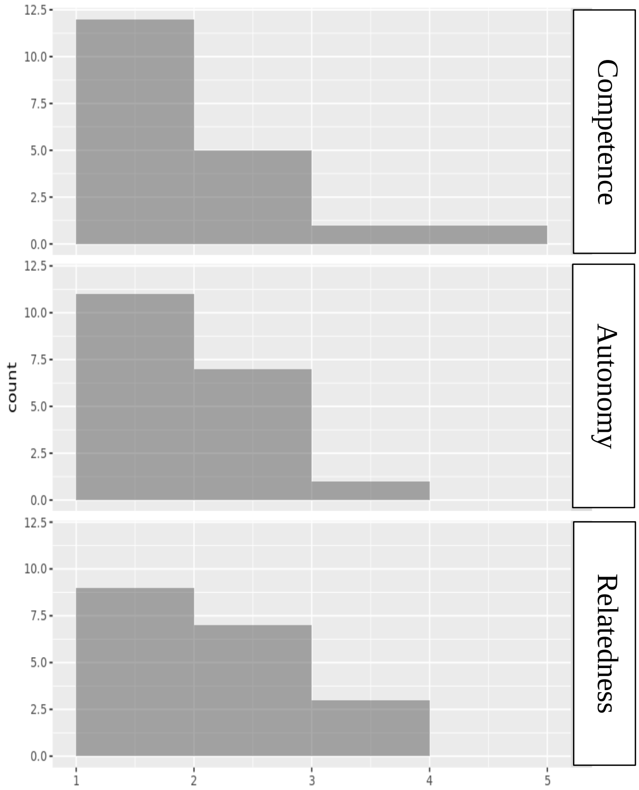

Figure 1. Histograms of the IDS composite scores for competence, autonomy, and relatedness (strongly agree (1) to strongly disagree (5)).

Individual Likert-type scale items provided a further perspective of how students perceive competence, autonomy, and relatedness in the IDS course. For example, nearly 90 percent of IDS students responded they *agree* or *strongly agree* with the competence item, “I’m confident about learning the content in this course.” Their response to this item echoed the newfound efficacy of IDS students described in the focus groups. Many admitted to prior disengagement and failure in high school mathematics and statistics, yet perceived data science as accessible and relevant. For the autonomy item, “I believe there is more than one way to solve a problem,” 89 percent of IDS students *agreed* or *strongly agreed* with the statement, which mirrored the freedom the students described in the IDS course. They felt unconstrained in their approaches to representing and interpreting data in the course. Finally, 14 of the 19 IDS students responded they *agreed* or *strongly agreed* with the relatedness item, “When I am working in this math class, I appreciate working with my peers on problems,” reflecting the strong sense of community the IDS students expressed in the focus group.

5. LIMITATIONS AND IMPLICATIONS

The goal of a case study is to closely examine a particular phenomenon—student experiences in one IDS class, making it impossible to generalize the findings beyond this single case (Creswell, 2012; Stake, 2005). Thus, the above findings are constrained by context, time, and participants. I only studied one section of IDS at Sierra Madre High School. Therefore, I am unable to extrapolate the findings to all IDS courses or generalize to other high school data science courses. I collected the data before the pandemic school closures, preventing any findings related to student experiences during remote learning. Ms. Arcega, the IDS teacher, graciously consented to participate along with 19 IDS students. Ms. Arcega was a highly qualified teacher and not necessarily indicative of all IDS teachers. And the small student cohort limits the power of the statistical tests making it harder to detect significant differences. Another limitation was relying on IDS students' comparisons of their prior experiences in mathematics and statistics courses to understand their current experiences in the IDS course since no additional details of their prior courses are known (e.g., qualifications of their teachers).

Although a limitation of case studies is generalizability, the data collected and subsequent analysis illuminate directions for future inquiry (Creswell, 2012; Stake, 2005). The findings of this case study suggest IDS students found meaning and empowerment in data science and expressed a new sense of confidence, agency, and belonging. Further studies may shed light on how pedagogy and different elements of the course, such as coding and engaging data sets, contribute to the students' positive perceptions of data science. Also, of interest is how the duration, malleability, and evolution of these beliefs after students finish the IDS course. For example, a longitudinal study might explore if IDS students who previously eschewed the STEM career pathway, now join it. This positive view of data science cultivated by the IDS students is especially significant given all but one of the students interviewed identified as Black, Indigenous, or a Student of Color, and all described previous marginalizing and disenfranchising experiences in high school mathematics and statistics. If policymakers, industry, and educators are committed to growing and diversifying STEM professions, data science may be uniquely situated to attract more diverse students to STEM, and at the very least allow students to find personal value in the discipline, regardless of their future careers (Arnold et al., 2020; Burdman, 2018; Committee on STEM, 2018; Dana Center, 2020; NCTM, 2018; PCAST, 2012; Su & Jackson, 2020).

Even though this study focused on student experiences, it was apparent the teacher's content knowledge and pedagogy influenced the experiences the IDS students had (Gould et al., 2016; UCLA, 2022). In the rush to implement data science in more high schools, it is vital to study the role professional learning plays in preparing teachers to teach this non-traditional course. For multiple pathways in mathematics to thrive, and data science in particular to flourish, the public must perceive the intellectual rigor and merit of these courses (Burdman, 2018; Fitzpatrick & Sovde, 2019; NCTM, 2018).

ACKNOWLEDGEMENTS

I wish to thank Ms. Arcega and her IDS students for participating in the research study. This research study was part of a larger dissertation study on non-traditional high school mathematics courses.

REFERENCES

Aguirre, J., Mayfield-Ingram, K., & Martin, D. (2013). *The impact of identity in K–8 mathematics: Rethinking equity-based practices*. The National Council of Teachers of Mathematics.

Arnold, K., Bargagliotti, A., Franklin, C., Gould, R., Johnson, S., Perez, L., & Spangler, D. (2020). *Pre-K–12 guidelines for assessment and instruction in statistics education II (GAISE II) report.* https://www.amstat.org/asa/education/Guidelines-for-Assessment-and-Instruction-in-Statistics-Education-Reports.aspx

Business–Higher Education Forum and PwC. (2017, April). I*nvesting in America's data science and analytics talent: The case for action.* PriceWaterhouseCoopers. https://www.bhef.com/sites/default/files/bhef_2017_investing_in_dsa.pdf

Boersma, S., & Savina, F. (2019). Re-envisioning the pathway to calculus: Supporting all students. In R. Hartzler & R. Blair (Eds.), *Emerging issues in mathematics pathways* (pp. 13–22). University of Texas at Austin & Charles A. Dana Center.

Bressoud, D. (2015). Insights from the MAA national study of college calculus. *Mathematics Teacher*, 109(3), 178–185.

Burdman, P. (2015, April). *Degrees of freedom: Diversifying math requirements for college readiness and graduation.* PACE & Learning Works. https://www.edpolicyinca.org/publications/degrees-freedom-diversifying-math-requirements-college-readiness-and-graduation

Burdman, P. (2018). *The Mathematics of opportunity: Rethinking the role of math in educational equity.* https://justequations.org/resource/the-mathematics-of-opportunity-report/

California Department of Education. (2017). *California school dashboard.* [website] https://www.caschooldashboard.org/

California Department of Education. (2022). *Mathematics framework.* [website] https://www.cde.ca.gov/ci/ma/cf/

California Math Council. (2020, October). The transition from high school to postsecondary mathematics designing modernized and socially just pathways through collaboration. https://www.cmc-math.org/assets/CMC%20Quantitative%20Reasoning%20Statement.pdf

Carnevale, A. P., & Desrochers, D. M. (2003). The democratization of mathematics. In B. L. Madison, L. A. Steen (Eds.), *Quantitative literacy: Why numeracy matters for schools and colleges* (pp. 21–31). National Council on Education and the Disciplines.

Charles A. Dana Center. (2020). *Launch years: A new vision for the transition from high school to postsecondary mathematics*. University of Texas at Austin. https://utdanacenter.org/launchyears

CourseKata. (n.d.). *CourseKata statistics and data science.* https://coursekata.org/

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Pearson.

Committee on STEM Education of the National Science & Technology Council. (2018). *Charting a course for success: America's strategy for STEM education* (Report). [ERIC Document Reproduction Service No. ED590474]

Daro, P. & Asturias, H. (2019, October). *Branching out: Designing high school mathematics pathways with equity in mind.* https://justequations.org/resource/branching-out-designing-high-school-math-pathways-for-equity/

Deci, E. L., & Ryan, R. M. (2008). Self-determination theory: A macrotheory of human motivation, development, and health. *Canadian Psychology [Psychologie canadienne]*, *49*(3), 182–185.

Deci, E. L., Vallerand, R. J., Pelletier, L. G., & Ryan, R. M. (1991). Motivation and education: The self-determination perspective. *Educational Psychologist*, *26*(3-4), 325–346.

DeVellis, R. F. (2017). *Scale development: Theory and applications* (4th ed.). SAGE Publications.

de Winter, J. C. F, Dodou, D., & Wieringa, PA (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, *44*(2), 147–181. https://doi.org/10.1080/00273170902794206

Fitzpatrick, L. P., & Sovde, D. (2019). The case for mathematics pathways from the launch years in high school through postsecondary education. In R. Hartzler & R. Blair (Eds.), *Emerging issues in mathematics pathways* (pp. 97–104). University of Texas at Austin & Charles A Dana Center.

Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology, 102*(4), 700–717. https://doi.org/10.1037/a0026659

Gould, R. (2010). Statistics and the modern student. *International Statistical Review*, *78*(2), 297–315.

Gould, R., Bargagliotti, A., & Johnson, T. (2017). An analysis of secondary teachers' reasoning with participatory sensing data. *Statistics Education Research Journal*, *16*(2), 305–334. https://doi.org/10.52041/serj.v16i2.194

Gould, R., Machado, S., Ong, C., Johnson, T., Molyneux, J., Nolen, S., ... & Zanontian, L. (2016). Teaching data science to secondary students: The mobilize introduction to data science curriculum. In J. Engel (Ed.), *Promoting understanding of statistics about society. Proceedings of the Roundtable Conference of the International Association of Statistics Education* (IASE), July 2016, Berlin, Germany. ISI/IASE. https://iase-web.org/documents/papers/rt2016/Gould.pdf?1482484533

Guay, F., Ratelle, C. F., & Chanal, J. (2008). Optimal learning in optimal contexts: The role of self-determination in education. *Canadian Psychology [Psychologie canadienne]*, *49*(3), 233–252.

Harel, G., & Rabin, J. M. (2010). Teaching practices associated with the authoritative proof scheme. *Journal for Research in Mathematics Education*, *41*(1), 14–19. https://www.jstor.org/stable/40539362

Ladson-Billings, G. (1997). It doesn't add up: African American students mathematics achievement. *Journal for Research in Mathematics Education,28*(6), 697–708. https://doi.org/10.2307/749638

Lincoln, Y., & Guba, E. (1985). *Naturalistic inquiry* (1st edition). SAGE Publications.

Maxwell, J. A. (2012). *Qualitative research design: An interactive approach* (3rd ed.). SAGE Publications.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative data analysis*. SAGE Publications.

Moussa, A., Barnett, E. A., Brathwaite, J. R., Fay, M. P., & Kopko, E. M. (2020). *A changing paradigm in high school mathematics.* https://ccrc.tc.columbia.edu/media/k2/attachments/changing-paradigm-high-school-mathematics.pdf

Mulhern, F., & Rae, G. (1998). Development of a shortened form of the Fennema-Sherman Mathematics Attitudes Scales. *Educational and Psychological Measurement*, *58*(2), 295–306.

National Council of Teachers of Mathematics. (2018). *Catalyzing change in high school mathematics: Initiating critical conversations*. Author.

Oregon Department of Education. (2020, August). *High school core math guidance: Version 4.* https://www.oregon.gov/ode/educatorresources/standards/mathematics/Documents/High%20School%20Core%20Mathematics%20Guidance.pdf

President's Council of Advisors on Science and Technology. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_2-25-12.pdf

Price, P. C., Jhangiani, R., & Chiang, I. C. A. (2015). *Research methods in psychology*. BC Campus.

Ratelle, C. F., Guay, F., Vallerand, R. J., Larose, S., & Senécal, C. (2007). Autonomous, controlled, and amotivated types of academic motivation: A person-oriented analysis. *Journal of Educational Psychology*, *99*(4), 734–746. https://doi.org/10.1037/0022-0663.99.4.734

Ryan, R. M., & Deci, E. L. (2000a). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54–67. https://doi.org/10.1006/ceps.1999.1020

Ryan, R. M., & Deci, E. L. (2000b). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*(1), 68–78. https://doi.org/10.1037/0003-066X.55.1.68

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). SAGE Publications.

Schoenfeld, A. H. (2014). What makes for powerful classrooms, and how can we support teachers in creating them? A story of research and practice, productively intertwined. *Educational Researcher*, *43*(8), 404–412. https://doi.org/10.3102/0013189X14554450

Seidman, I. (2006). *Interviewing as qualitative research: A guide for researchers in education and the social sciences* (4th edition). Teachers College Press.

Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The SAGE handbook of qualitative research* (2nd ed., pp. 435–454). SAGE Publications.

Su, F. E., & Jackson, C. (2020). *Mathematics for human flourishing.* Yale University Press.

Tapia, M., & Marsh, G. E. (2004). An instrument to measure mathematics attitudes. *Academic Exchange Quarterly*, *8*(2), 16-21.

Teddlie, C., & Tashakkori, A. (2009). *Foundations of mixed methods research: Integrating quantitative and qualitative approaches in the social and behavioral sciences*. SAGE Publications.

UCLA. (2022a). *Introduction to data science: Our data: Our lives*. https://www.mobilizingcs.org/

UCLA. (2022b). *Introduction to Data Science Curriculum.* https://www.ucladatascienceed.org/introduction-to-data-science-curriculum

U.S. Department of Education Office for Civil Rights. (2018, April). 2015–16 Civil rights data collection STEM course taking. https://www2.ed.gov/about/offices/list/ocr/docs/stem-course-taking.pdf

Wilkerson, M. H., & Polman, J. L. (2020). Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences*, *29*(1), 1–10. https://doi.org/10.1080/10508406.2019.1705664

YouCubed. (2020, December 03). *21st century teaching and learning: Data Science*. Stanford University. https://www.youcubed.org/21st-century-teaching-and-learning/

ERICA HEINZMAN
University of California, San Diego
eheinzman@ucsd.edu

APPENDIX

The following are the Likert-type scale items from the original survey. I have placed a double asterisk next to the items which exceeded the component coefficient threshold of 0.5 (PCA) and comprise the composite score for competence, autonomy, and relatedness.

Competence Statements

In this math class:

1. I feel secure about attempting problems **
2. I am confident about learning the content in this course **
3. I think I can handle more challenging problems
4. I feel lost, though I do okay in my other classes
5. I look forward to taking more math classes
6. The content makes me feel nervous
7. I have a sinking feeling when I try problems
8. I have self-confidence
9. My teacher has encouraged me to study more math **
10. I am comfortable expressing my own ideas on how to look for solutions to a difficult problem. **
11. The content is very interesting.**
12. I am comfortable answering questions in math class.**

Autonomy Statements

1. I believe I will need math in my future work **
2. I am in this course because it is useful **
3. I see math as a subject I will rarely use in daily life
4. I believe this math class is worthwhile **
5. This class is enjoyable for me **
6. When I am confused by a problem in this math class, I stick with it
7. I'd rather have someone give me the answer to a difficult math problem.
8. Knowing math will help me earn a living.**
9. When I start working on a problem in this course, I find it hard to stop.**
10. This class is dull and boring.
11. I like to solve new problems in this class **
12. I would like to avoid using math in college.
13. This class would be very helpful no matter what I decide to study.**
14. The challenge of this class appeals to me. **
15. My teacher values only one method and answer
16. In this class, I can choose my way to solve a problem
17. I believe there is more than one way to solve a problem **

Relatedness Statements

When I am in this math class:

1. I feel that I belong
2. I feel like I am part of this community
3. I feel like an outsider **
4. I feel accepted **
5. I feel valued **
6. I feel excluded **
7. I feel comfortable **
8. I feel inadequate
9. I wish I could fade into the background and not be noticed **
10. I try to say as little as possible
11. I enjoy being an active participant **
12. Even when I struggle, I trust my teacher to have faith in my potential **
13. I appreciate working with my peers on problems **
14. I feel supported by peers**

In this class, I would describe math [write what comes to mind]: