Psychological Factors as Mediators of Demographic Differences in Performance on the QuaRCS Quantitative Literacy Assessment

Soon-Young Shimizu

Advisors:
Professor Kate Follette,
Professor Nicholas Horton,
Professor Patricia O'Hara
Dr. Sarah Bunnell

May 15, 2020

Submitted to the
Department of Special Programs of Amherst College
in partial fulfillment of the requirements for
the degree of Bachelor of Arts with honors

Abstract

Quantitative literacy or numeracy is the ability to utilize, apply, and reason with mathematical, statistical, and numerical skills. The importance of developing quantitative literacy is well-recognized, yet, there are significant disparities in quantitative literacy across demographic groups. To explore why these disparities exist, we utilized data from the Quantitative Reasoning for College Science (QuaRCS) assessment. It measures students' quantitative literacy performance and gathers their responses to non-quantitative questions that probe their affects and attitudes. We focused on the demographics of gender, race/ethnicity, and socioeconomic status, comparing the performance and responses on the assessment between underrepresented and non-underrepresented groups. To prepare the data for analysis, we performed Factor Analysis, a technique used to extract underlying variables. From this procedure, several composite variables constructed from questions on the assessment emerged: Numerical Self-Efficacy, Numerical Anxiety, Numerical Relevancy, School Math Affect, Daily Life Math Affect, and Course Motivation. Using linear regression analysis, we find that all composite variables are predictive of score. Using mediation analysis, we find that certain composite variables are able to partially mediate the relationship between an underrepresented group and their lower QuaRCS score. Among our composite variables, Numerical Self-Efficacy, Numerical Anxiety, and Daily Math Affect are the strongest mediators to explain the achievement gap. Which of our composite variable(s) acted as mediators varied between underrepresented demographic groups, suggesting that factors that explain differences are unique to each. However, no single variable can explain more than 50% of the gap, and it's likely that compound effects of multiple mediating variables or additional variables not measured by the QuaRCS instrument are needed to fully explain the gaps. There are many factors that are at play in these differences, and additional variables beyond the scope of this thesis may underlie or contribute to differences in score.

Acknowledgments

The decision to write an education thesis and declare a corresponding interdisciplinary major was an impromptu one. I am immensely grateful for those who supported my decision and guided me throughout the process of developing and completing this thesis.

First, I would like to thank Professor Michelle Tine at Dartmouth College, who took me off the waitlist for her Intro to Education course and subsequently reinstilled my interest in education. I would not have otherwise decided to pursue this education thesis (and hopefully later graduate work in education). I must also thank Professor Eric Van Erp who let me (not an enrolled student at the time) audit his introductory statistics course at Dartmouth–a prerequisite I needed to declare and complete an interdisciplinary major so late in my college career.

As I scrambled to get my interdisciplinary proposal together, I am appreciative of my committee members, Professor Nick Horton, Professor Pat O'Hara, and Dr. Sarah Bunnell, who graciously agreed to serve so quickly and willingly. I am additionally appreciative of Professor Horton for his statistical guidance and Dr. Bunnell for her extensive comments and suggestions!

None of my work would have been possible without the efforts, assistance, and support from previous and current members of the QuaRCS group. They have built the QuaRCS infrastructure on which this thesis stands, attended to my troubleshooting issues, and listened to my weekly thesis updates. Cat Sarosi, in particular, has facilitated my onboarding process to the project, and I am especially grateful for her expertise.

I am indebted to the Stats Fellows and Professor Brittany Bailey for assisting me with all my ad hoc statistical needs. Special thanks to Dawit Wachelo for reading and reviewing my entire thesis with much enthusiasm. Thanks to AK, AY, and SH for listening to my horrendously inarticulate and lengthy practice presentations. Additional thanks to SH, I guess, for reading half of my thesis.

I extend my gratitude to friends, professors, and acquaintances, who may not have contributed in any direct way to this thesis, but instead have contributed to my own personal growth during my time at Amherst. As a physics major, much of my time was spent around fellow physics folks at Amherst, and I have been supported by them especially. I would not have been able to persist or survive as a physics major, or as a student at Amherst, without this community. While this thesis has nothing to do with physics in the slightest, I have learned (and suffered :)) much from my experiences in this discipline, which surely contributed in some way. I am appreciative of all of it.

And in this quarantine period of COVID-19, under these unusual circumstances during which I have completed this thesis, the friends of Morrow 3rd and 4th floor deserve much appreciation and gratitude. Quarantine on campus would have been dull and dispiriting without the collection of these eclectic individuals. They have created a joyously unique space in this odd time, which I will remember dearly.

Of course, I must give my greatest thanks to my principal advisor, Professor Kate Follette. I randomly came along with a request mid-way through Junior year about potentially working on her education research. She immediately pressed forward with it (even though she already takes on too much!), and what resulted was this thesis. She has supported me through the entire process, bearing with me while I learned everything for this thesis, as I was working on the thesis. I could not have completed a project of this nature without her guidance, patience, and time. Thank you.

Lastly, I need to give a small thanks to *Team Sports - Live* by Cory Wong & Metropole Orkest[1], a song I listened to on repeat for *hours* as I finished up this thesis. Writing is a dreadfully slow and effortful process for me, and it has helped retain my energy in the last stretches of writing. Please go take a listen.

[1] I emphasize, the LIVE version on the Live from Amsterdam album, not the studio version

Contents

Chapter 1

Introduction

We might intuitively understand why mathematics is a fundamental component of school curricula, nationally required in all classrooms from kindergarten to high school. Perhaps the most common intuition is the utility of mathematics[1]. Basic skills such as counting and performing simple algebraic operations to skills involving more complex mathematics such as trigonometry and calculus are deemed useful. However, when we–students, parents, educators, and administrators–discuss the instrumental value of learning mathematics, perhaps it is not mathematics itself that we are ultimately referring to but rather the context in which we apply mathematical and numerical skills. In day-to-day circumstances, it is not the ability of 'solving for x' in an algebraic equation that we find useful, but rather the use of algebraic thinking to solve problems, for example, determining how many 5-seater cars are needed to transport 18 people. It is not the ability to 'prove the triangle is an isosceles' that we are concerned with but the logical reasoning needed to make a sound

[1]Perhaps to the dismay of mathematicians...see [1]

conclusion. So, what we aim to derive from a good mathematics education is not the mathematical and numerical skill in itself, but rather the application and understanding of such skills in broader contexts. In other words, what we are seeking to achieve in our education is *quantitative literacy* [2].

1.1 What is Quantitative Literacy?

The concept of literacy is known with greater familiarity as the ability to read, write, understand, and contextualize text. However, literacy can be understood more generally and elegantly as "not merely as a measurable amount of technical skill, but also a judgement about the nature and quality of an interaction between a person with that skill and a particular social or environmental situation." This calls for the need to periodically update our understanding or 'judgement' of literacy due to the ever-changing circumstances in society [2].

In particular, with the development of technology, the 21st century demands a new set of abilities for the data-driven world in order to be successful in the workplace and to be informed participants in a democratic society. For instance, the collection, creation, and dissemination of information has exponentially increased since the accessibility of the internet. How we digest the plethora of information into knowledge is perhaps a greater issue now then it was in the last century.

Several branches of literacy emerged in the modern world in addition to the traditional sense of literacy: financial literacy, health literacy, and science literacy. Soon, we will see that these literacies are coupled with a more fundamental type of literacy: *quantitative literacy*.

The skills and traits of quantitative literacy, or more historically known as numeracy, have been desirable characteristics in individuals since the commercialization of industries in seventeenth-century England or nineteenth-century America. However, the term numeracy did not emerge until the mid twentieth century with the basic definition of "an ability with or knowledge of numbers"[3]. Such a vague definition of numeracy is not descriptive of the value in or the aspects of the literacy of numbers.

An updated comprehensive definition of what quantitative literacy or equivalently, numeracy or quantitative reasoning, draws from similar yet distinct terms such as: *mathematical literacy*, which requires the understanding of the role of mathematics in greater contexts and the ability to make mathematically based judgements; and *mathematical reasoning*, which involves critical thinking about what mathematical tools are appropriate and whether or not results make sense [2]. Table 1.1 tabulates the key elements of quantitative literacy from *Mathematics and Democracy*.

1.2 Why is Quantitative Literacy Important?

Ultimately, quantitative literacy is the ability to utilize, apply, and reason with mathematical, statistical, and numerical skills, which makes it a fundamentally interdisciplinary concept. Because of its applicable nature, we see that deficiencies in quantitative literacy have consequences across wide-ranging domains. As the author of *Mathematics and Democracy*, Lynn Arthur Steen,

Table 1.1: Elements of Quantitative Literacy. Paraphrased from Steen 2001

Elements	**Description**
Confidence with Math	Being comfortable with using and applying quantitative ideas and methods.
Cultural Appreciation	Understanding the history and role of mathematics in wide-ranging applications, and recognizing its relevance in day-to-day contexts.
Interpreting Data	Being able to interpret tables, figures, charts, and draw out pertinent information and make conclusions.
Logical Thinking	Being able to analyze evidence, have sound reasoning and conclusions, understand the basis of arguments, and identify fallacies.
Making decisions	Utilizing mathematics in day-to-day contexts to solve problems and make informed decisions.
Mathematics in Context	Utilizing mathematics in particular contexts such that the context provides meaning to the mathematical results and concepts.
Number sense	Having the familiarity and intuition of the meaning of numbers such as scale, proportional reasoning, and estimation.
Practical Skills	Utilizing elementary mathematics in common circumstances.
Prerequisite Knowledge	Having algebraic, geometric, and statistical knowledge to utilize them as tools in greater fields of study and education.
Symbol Sense	Familiarity and comfort in seeing and using algebraic symbols; understanding the syntax and grammar of mathematical symbols.

puts it, "an innumerate citizen today is as vulnerable as the illiterate peasant of Gutenberg's time"[2]. It acts as an underpinning for other kinds of literacies. Science literacy, for example, is founded on similar principles of logical and analytical reasoning. It requires a basic understanding of the sciences, its processes, and methods, as well as the ability to make evidence-based arguments. The latter requires numerical and quantitative skills to work with and interpret data such that quantitative literacy is a corequisite to science literacy [4]. It is a similar case with financial and health literacy. Financial literacy encompasses the ability to make sound financial decisions by using and understanding basic financial concepts like interest rate and inflation [5]; Health literacy requires the ability to understand basic health information or instructions in various forms such as nutrition labels, risk assessments, or dosing, which inform personal decisions for good health [6]. Studies have shown that differences in financial literacy were best explained by participants' level of numeracy even after adjusting for income, gender, age, education and cognitive and emotional factors such as cognitive reflection, self-efficacy, math anxiety, and financial anxiety [5]. Studies in health literacy show similar results such that proper health literacy corresponds to higher numeracy levels. [6]. Both rely on the ability to work with percentages, probabilities and basic arithmetic calculations to understand and apply interest and inflation rates or interpreting chances of success or failure. Therefore, quantitative literacy is necessary for literacy in a variety of other domains.

Numeracy is also an essential skill in proper decision-making as evidenced by a set of four studies conducted with participants from an introductory psychology course. Participants who were more numerate were able to assess probabilities of risk or loss and choose the better outcome, compared to those who were less numerate. They were also less prone to the framing effect, a cognitive bias of choosing identical options differently depending on how they are put in context [7]. Consequently, numeracy ties into financial and health decisions, for example, on loans and investments, or on a treatment's risks and benefits, which are often founded upon numerical assessments [5][6]. Thus, innumerate individuals are more likely to make sub-optimal decisions which affects personal successes in life. But such decisions can have greater impacts as well. Choosing to participate in the stock and housing market, to take out loans, or to make investments, influence the allocation of resources which in turn impacts the health and stability of the economic system [8][9].

Thus, it is no surprise that numeracy is highly sought by employers. Especially now, there is a high demand for STEM-related occupations, for which quantitative literacy is essential [4]. But in a growing technologically-based society, even non-STEM-related occupations are in need of quantitative skills [10]. Therefore, low numeracy skills can be a barrier to working higher paying jobs and achieving financial stability, acting as an impediment to social mobility [10][11]. Essentially, numeracy is a form of human capital which subsequently increases and sustains participation in the labor market and fulfills its demands which in turn, increases economic activity and health. In particular, numeracy acted as a better predictor of employment status, income level, and labor force participation than literacy when controlling for years of schooling [12].

Even in political contexts do we see the importance of quantitative literacy. *Mathematics and Democracy* argues that a well-functioning democracy is contingent on properly educated public citizens of which being quantitatively literate is a necessary component. As U.S. citizens, we must be able to grapple with data-driven claims to make our own judgements and contribute to public reasoning and influence greater decisions that impact the entire democracy as a whole. "Numerical thinking [is] essential to the 'discourse of public life." Quantitative literacy is a 'liberating' kind of literacy. A minimal level of literacy, or 'inert literacy', is necessary to perform basic tasks, follow routine procedure, and function as a participant in society. But a 'liberating literacy' is an additional step towards using skills for critical thinking and reflection, making sound conclusions, and acting on well-founded decisions. This is what democracy promotes and necessitates. We can see that quantitative literacy fits into the equation of a liberating literacy due to the extensive use of numbers and quantitative reasoning in the modern world that are used to justify claims and ground political decisions. However, we see too frequently that the numbers that are thrown into political and social contexts as evidence are misused, misinterpreted, misunderstood or without context–contributing to the issue of numerical misinformation [13][14].

What simply follows are dubious or faulty claims, and unwarranted conclusions from our innumerate public, or from our trusted representatives and elected officials which may result in unfounded, ineffective, or even harmful action[2].

[2]Consider the federal response to the COVID-19 Pandemic in the United States, for example...

1.3 The United States and Quantitative Literacy

Quantitative literacy is clearly something of importance. Unfortunately, the education system in the United States is failing to develop adequate numeracy skills.

In 2016, the Organisation for Economic Co-operation and Development (OECD) conducted an international study of literacy. Questions[3] of various difficulties were given to adults 16-65 years of age assessing "the ability to access, use, interpret and communicate mathematical information and ideas in order to engage in and manage the mathematical demands of a range of situations in adult life." Out of 34 countries, the U.S. ranks 28th in numeracy proficiency (Figure 1.2)[10]. When considering the U.S.' expenditures on education, these values fall far below expectations. The US spends an average of $ 31,000 per full-time student in post secondary education, the highest of those surveyed in the OECD study, and 93% greater than the OECD average. Similarly, the U.S. ranks second in highest expenditures in primary and secondary education. On average, $12,800 is spent per full-time student, 35% greater than the OECD average [15]. While the literacy rate is higher than the average of OECD countries, the quantitative literacy rate is surprisingly below average.

The study establishes several levels of numeracy proficiency which corresponds to a range of numeracy scores. A description of each level is provided

[3] Refer to Appendix A.1 for sample questions from the OECD

in Figure 1.1.

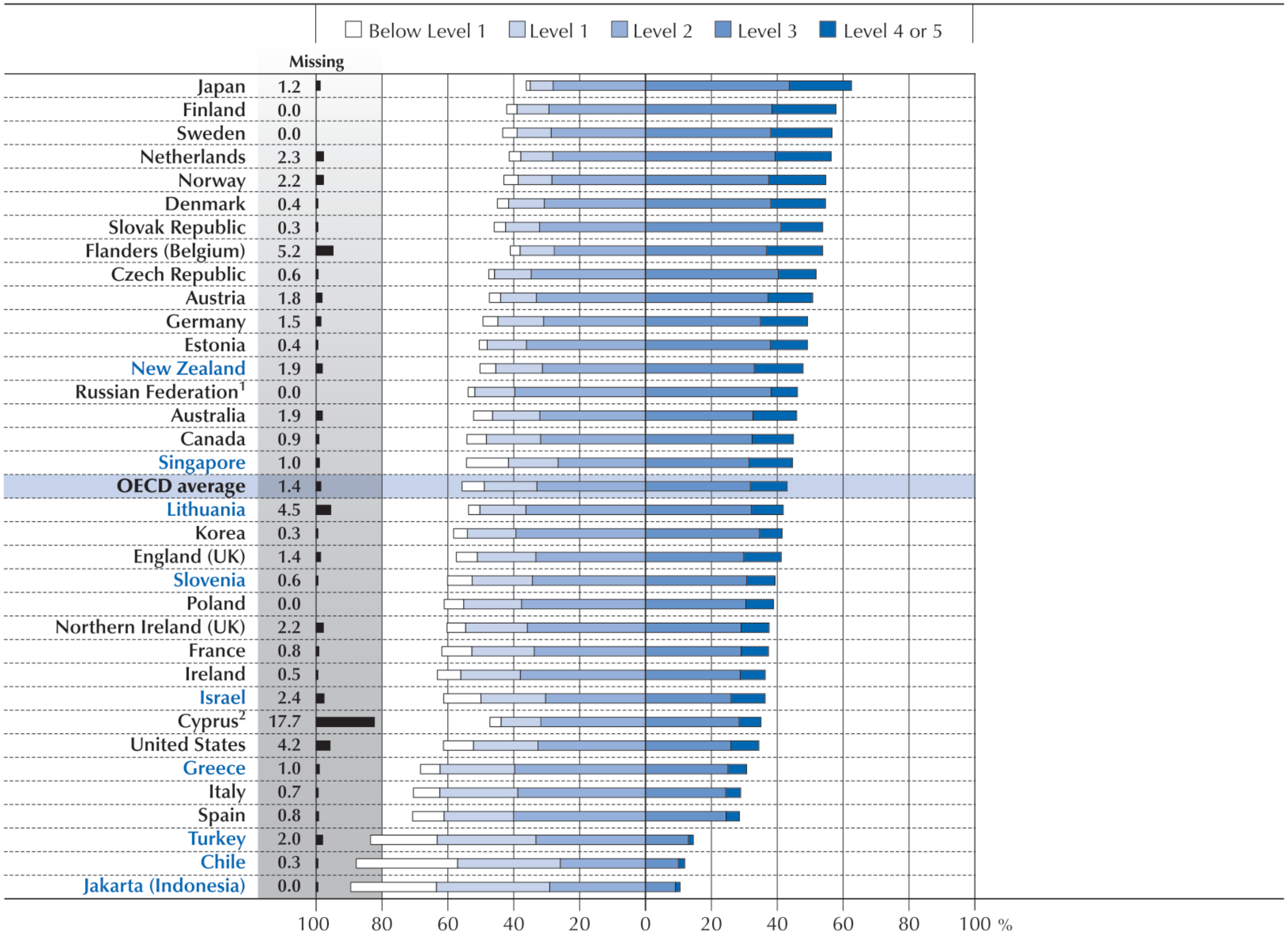
Level	Score range	Percentage of adults scoring at each level (average)	The types of tasks completed successfully at each level of proficiency
Below Level 1	Below 176 points	6.7%	Tasks at this level require the respondents to carry out simple processes, such as counting, sorting, performing basic arithmetic operations with whole numbers or money, or recognising common spatial representations in concrete, familiar contexts where the mathematics content is explicit with little or no text or distractors.
1	176 to less than 226 points	16.0%	Tasks at this level require the respondent to carry out basic mathematical processes in common, concrete contexts where the mathematical content is explicit, with little text and minimal distractors. Tasks usually require one-step or simple processes involving counting, sorting, performing basic arithmetic operations, understanding simple percentages, such as 50%, and locating and identifying elements of simple or common graphical or spatial representations.
2	226 to less than 276 points	33.0%	Tasks at this level require the respondent to identify and act on mathematical information and ideas embedded in a range of common contexts where the mathematics content is fairly explicit or visual with relatively few distractors. Tasks tend to require the application of two or more steps or processes involving calculation with whole numbers and common decimals, percentages and fractions; simple measurement and spatial representation; estimation; and interpretation of relatively simple data and statistics in texts, tables and graphs.
3	276 to less than 326 points	31.8%	Tasks at this level require the respondent to understand mathematical information that may be less explicit, embedded in contexts that are not always familiar and represented in more complex ways. Tasks require several steps and may involve the choice of problem-solving strategies and relevant processes. Tasks tend to require the application of number sense and spatial sense; recognising and working with mathematical relationships, patterns and proportions expressed in verbal or numerical form; and interpretation and basic analysis of data and statistics in texts, tables and graphs.
4	326 to less than 376 points	10.2%	Tasks at this level require the respondent to understand a broad range of mathematical information that may be complex, abstract or embedded in unfamiliar contexts. These tasks involve undertaking multiple steps and choosing relevant problem-solving strategies and processes. Tasks tend to require analysis and more complex reasoning about quantities and data; statistics and chance; spatial relationships; and change, proportions and formulas. Tasks at this level may also require understanding arguments or communicating well-reasoned explanations for answers or choices.
5	Equal or higher than 376 points	1.0%	Tasks at this level require the respondent to understand complex representations and abstract and formal mathematical and statistical ideas, possibly embedded in complex texts. Respondents may have to integrate multiple types of mathematical information where considerable translation or interpretation is required; draw inferences; develop or work with mathematical arguments or models; and justify, evaluate and critically reflect upon solutions or choices.

Note: The proportion of adults scoring at different levels of proficiency adds up to 100% when the 1.4% of numeracy-related non-respondents across countries/economies are taken into account. Adults in the missing category were not able to provide enough background information to impute proficiency scores because of language difficulties, or learning or mental disabilities (see section on literacy-related non-response above).

Figure 1.1: OECD Levels of Numeracy Proficiency

Compared to the international average of 53%, 61% of U.S. adults fall below level 3 proficiency. Some examples of tasks below level three include identifying trends in graphs or reading off a thermometer (Appendix A.1).

Percentage of 16-65 year-olds scoring at each proficiency level in numeracy



Note: Adults in the missing category were not able to provide enough background information to impute proficiency scores because of language difficulties, or learning or mental disabilities (referred to as literacy-related non-response).

1. See note at the end of this chapter.

2. See note 1 under Figure 2.1.

Countries and economies are ranked in descending order of the combined percentage of adults scoring at Level 3 and at Level 4 or 5.

Source: Survey of Adult Skills (PIAAC) (2012, 2015), Table A2.4.

StatLink http://dx.doi.org/10.1787/888933365863

Figure 1.2: Numeracy Proficiency in Adults in OECD countries

Such tasks are relevant in day-to-day contexts from watching and reading the news to checking the speedometer as you drive. About 200 million Americans do not meet this level of quantitative proficiency. This raises the question of how we can expect Americans to be properly informed and make good decisions both in their personal lives and their roles as citizens and voters.

Additionally, the U.S.' variability in score is troubling. Measured as the difference in the first and third quartiles, the U.S. has a difference of 76 points– the fourth largest of all surveyed countries (Figure 1.3). This demonstrates the wide range of numeracy proficiency levels and suggests large disparities in ability among the U.S. population.

This begins to illustrate another key concern with the United States' education system. It fails to educate populations equally such that the levels of innumeracy are more severe for certain demographic groups than for others. Results from the 2013 National Assessment of Adult Literacy (NAAL) provides such evidence (Figure 1.4). Here, 4 levels of quantitative literacy are defined from proficient, intermediate, basic, and below basic[4]. 'Basic literacy' is defined as the skills needed to perform simple every-day literacy activities such as calculating the total cost of a lunch or locating and using information from a graph[5] [16]. Of American adults 16 years and older, 55% fall at or below basic literacy, but the differences in quantitative literacy levels for underrepresented minorities are striking. A clear and concerning disparity between these two populations exists, which likely contributes to the high variability in the US OECD numeracy scores. A much higher proportion of underrepresented

[4]Description of all levels is found in Appendix A.2

[5]Refer to Appendix A.3 for a sample question from NAAL

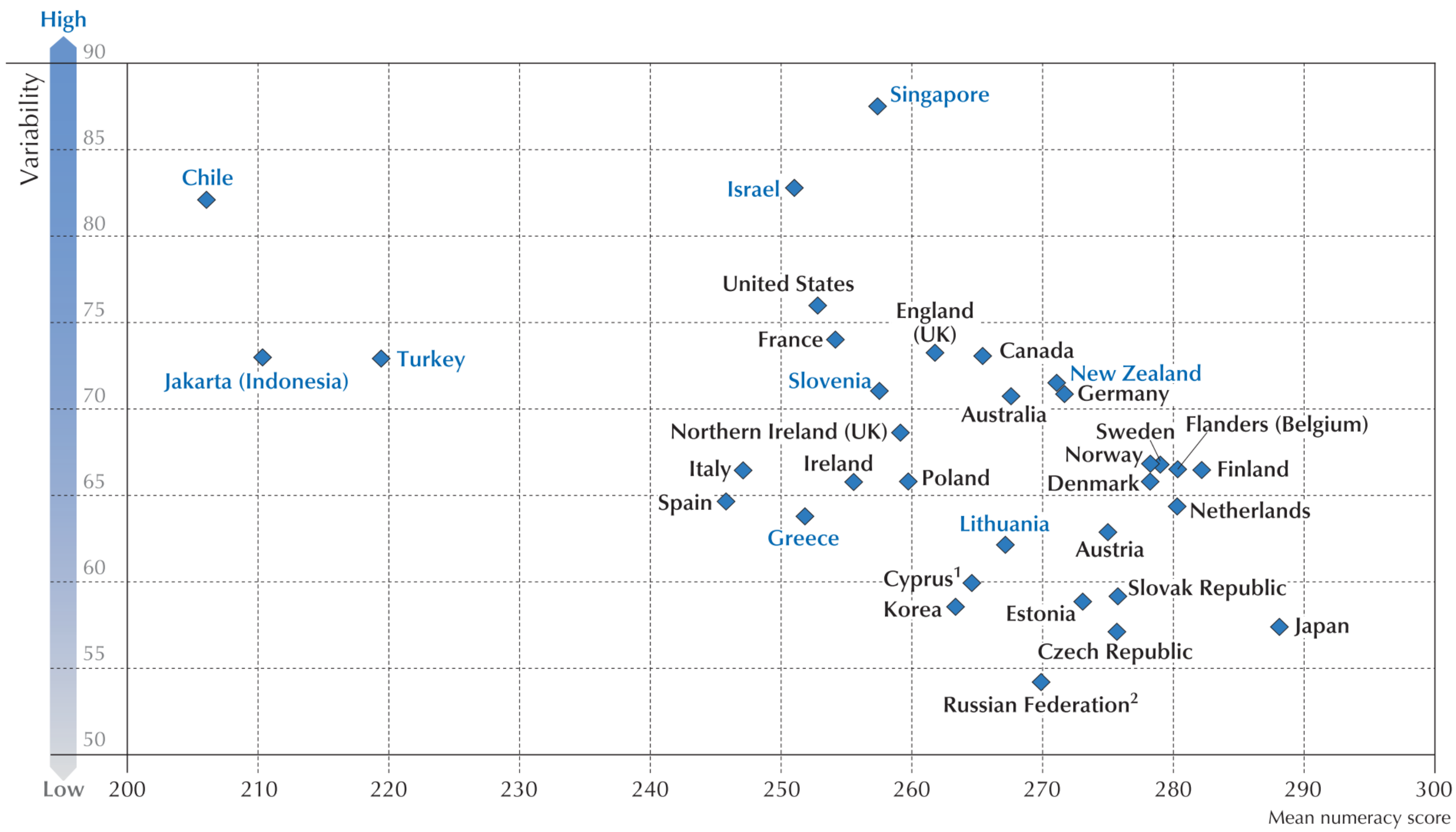
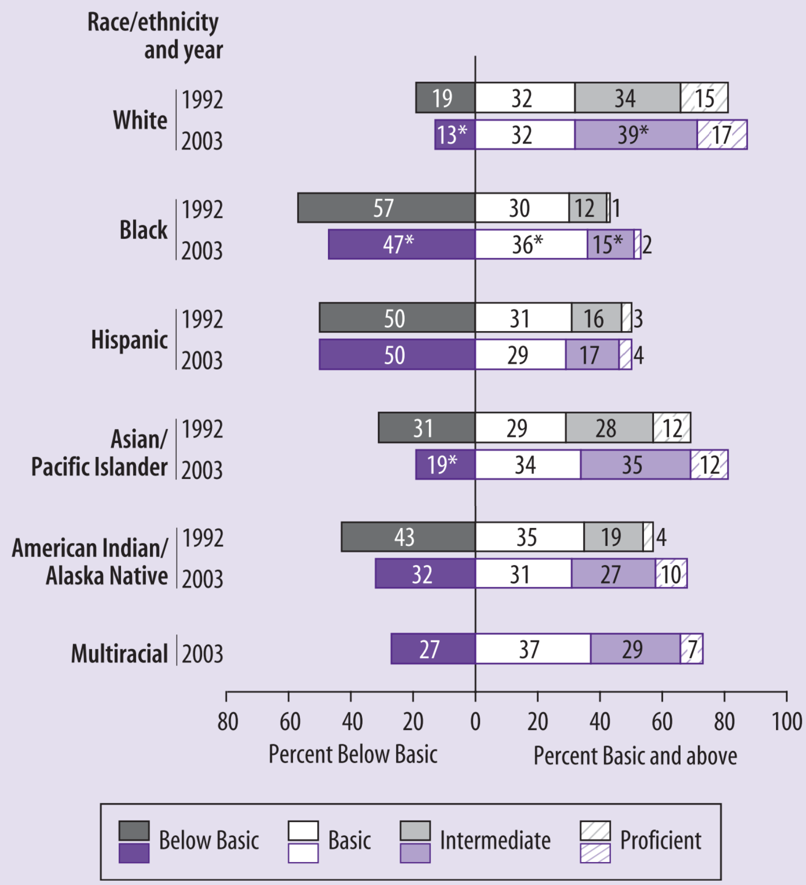


Figure 1.3: Variability vs. Mean Numeracy Score by country



*Significantly different from 1992.

NOTE: Detail may not sum to totals because of rounding. Adults are defined as people 16 years of age and older living in households or prisons. Adults who could not be interviewed because of language spoken or cognitive or mental disabilities (3 percent in 2003 and 4 percent in 1992) are excluded from this figure. In 1992, respondents were allowed to identify only one race but could identify "other" as their race. In 2003, respondents were allowed to identify multiple races but could not choose "other" as their race. All adults of Hispanic origin are classified as Hispanic, regardless of race. The Asian/Pacific Islander category includes Native Hawaiians. Results for adults who identified "other" as their race in 1992 are not included in this figure. The percentage of the population that identified "other" as their race in 1992 rounds to 0.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 1992 National Adult Literacy Survey and 2003 National Assessment of Adult Literacy.

Figure 1.4: Percentage of Adults in each quantitative literacy level by race/ethnicity: 1993 and 2003.

minorities fall below basic levels of literacy, and a lower proportion fall above basic levels compared to the white population. For example, 47% and 50% of Black and Hispanic adults fall in the 'below basic' level of literacy while only 13% of White adults do; and only 2% and 4% of Black and Hispanic adults reach the level of 'proficiency' compared to 17% of the white adult population.

We understand the importance of quantitative literacy, and yet the state of our innumerate population is abysmal. Why is the United States education system failing to develop our numeracy skills? While this certainly is an important question to investigate, there is another piece that is also of concern: why is there a disparity in numeracy skills in the United States? In other words, why are the levels of innumeracy more severe for certain demographic groups than for others?

In this thesis, we focus on the latter component, for these salient differences in literacy illuminate the long standing issue of social inequality within the United States. The severe consequences of innumeracy likely arise most frequently within these demographic groups impeding their participation in democratic, economic, and social contexts and maintaining a cycle of social inequality. To understand this disparity, we investigate what factors predict quantitative literacy, and whether these factors explain differences between demographic groups.

In chapter 2, we provide an overview of the historical trends demonstrating these differences known as the achievement gap. Then, we discuss various factors that contribute to the gap, and identify which factors we choose to investigate more specifically. In chapter 3, we begin to discuss the methods of our investigation, outlining how our data is collected and processed. In chapter 4, we continue our data processing discussion around a procedure called Factor Analysis. Once equipped with cleaned and refined data, we finally analyze our data in chapter 5 using various statistical methods and discuss our results.

Chapter 2

Background

While the numeracy results of the National Assessment of Adult Literacy are alarming, the disparities in performance by race/ethnicity are not new. Known as the 'achievement gap' particularly in the context of academic achievement, differences in academic performance have persisted for decades, not only among racial/ethnic groups but among various demographic groups as well, such as gender and socioeconomic status. In this chapter, we give a brief overview of the historical trend in the achievement gap, and the factors that contribute to these differences. Such factors inform potential reasons for differences in numeracy specifically, and drive the hypotheses in our current investigation.

2.1 The Achievement Gap

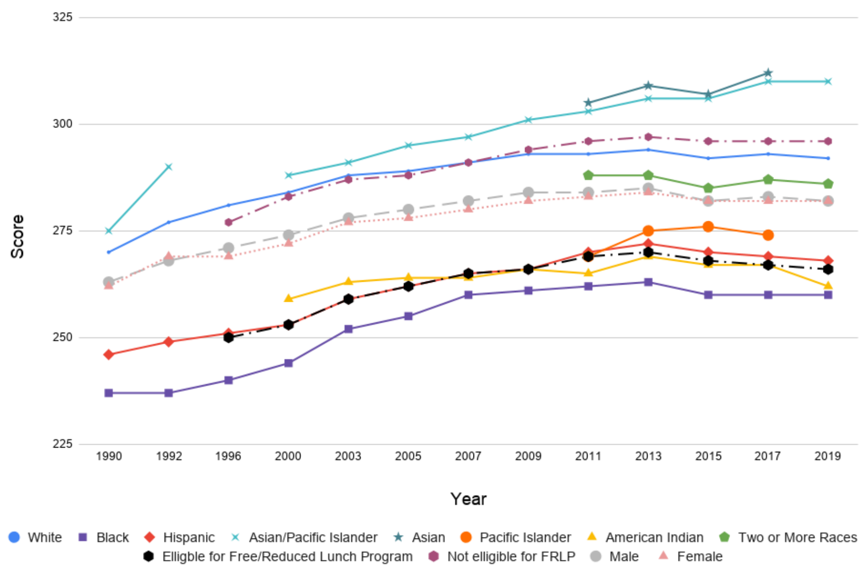
The achievement gap is defined as the difference in academic performance between different demographic groups such as gender, race/ethnicity, disability status, and socioeconomic status. Differences are reflected not only in test scores but also in grades, graduation rates, and college attendance rates [17]. The historical trend in the achievement gap in race and gender is best documented in the United States by the National Assessment for Educational Progress (NAEP).

The NAEP has conducted national exams since 1970 to measure academic achievement of 4th, 8th, and 12th grade students in various domains. From its inception, we have seen consistent patterns of performance between demographic groups in reading, mathematics, and science scores for decades. Unfortunately, little progress has been made in diminishing persistent racial/ethnic disparities. While data on numeracy itself has not been collected as early as the other subjects, mathematics and numeracy go hand in hand. Thus a look into the historical trends in the mathematics score gap can act as a rudimentary proxy in estimating trends in numeracy gaps. Figure 2.1 illustrates the trend in NAEP mathematics scores of eighth graders of several demographic groups from 1990 to 2019[1]. Figure 2.1a shows a general positive trend in score for all demographics. Figure 2.1b shows the differences in score between demographic groups; the average score for White students act as the baseline comparison for all race/ethnicity groups; and the eligibility status for Free/Reduced Price Lunch (FRPL) acts as a proxy for socioeconomic status of the student.

Since its administration, there has been a positive trend in math scores, although its growth has stagnated for most groups in the last decade. Moreover,

[1] While we are interested in older students (i.e. 12th graders) as well, data for this demographic is not fully available. However, it is known that the gap only widens with age.[18][19]

A.



B.

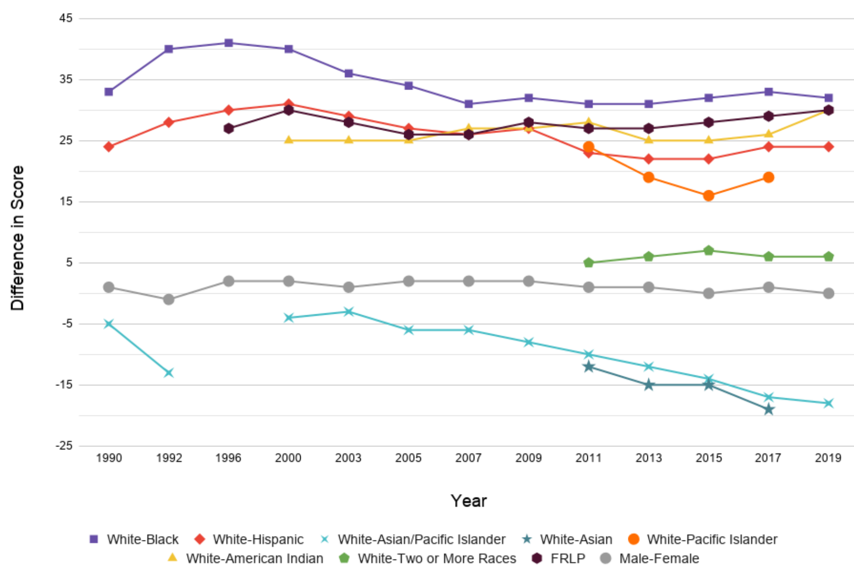


Figure 2.1: NAEP Mathematics Score 1990 - 2019. A: trends in score by demographic group. B: differences in score between demographic groups

there continues to be a clear distinction between groups. From 1992 to 2003, the Black-White student gap narrowed with a progressive change every few years until 2005 where the gap has seemingly reached a relatively steady state since with no significant difference between 2005 and 2019 scores. From 2000 to 2015, the White-Hispanic student gap has continued to narrow with the exception of 2009; however, the gap has begun to widen again with significant differences between 2015 and 2019. Unfortunately, no progress has been made for American Indian students since 2000 when data was first collected from this group, and the gap only continues to grow. The gap between students who are not eligible for FRPL and students who are eligible resemble differences between White and minority students, as most students who are eligible are students of minority groups [15]. Students who identified with two or more races have the smallest gap between White students, and remained constant since 2011.

Unlike previous minority groups, Asian students have consistently performed significantly higher than White students and other demographic groups since 2003. While earlier data consolidated Asian and Pacific Islander groups, more recent data provides a breakdown of both groups. It demonstrates that the Asian student group is the driver of the trend in earlier data. In fact, Pacific islander students perform more similarly to Hispanic students than to Asian students.

The performance of each race/ethnicity group relative to White students correlate similarly with the 2003 NAAL numeracy results in adults. While we cannot be completely certain of numeracy trends, we can suspect that it will follow the NAEP mathematics score trend in a similar manner. With the exception of Asian students who have made significant progress since 2003 surpassing White students in math performance, we can predict that a large gap still exists between other minority groups.

Fortunately, the gender gap has remained fairly small since 1990. While the differences are statistically significant, the two groups have only differed in score by 1 or 2 points. It is important to note, however, that this trend does not reflect the gender gaps that may be present within demographic groups.

With this overview of various demographic trends in achievement, it is clear that the achievement gap is a longstanding and complex issue. There have been attempts to find explanations of the gap with a variety of theories, demonstrating the multifaceted nature of the issue [20]. In the next section, we give an overview of the literature on the factors suspected to drive the achievement gap.

2.2 Drivers of the Achievement Gap

Historically, biological and genetic factors acted as the explanation for differences in cognitive ability, an idea that was further reinforced with the publication of *The Bell Curve* in 1994. Authors Richard Herrnstein and Charles Murray attributed economic success with higher levels of IQ, which they claimed were biological, heritable traits that differed between races. However, this claim has largely been discredited and dismissed. The achievement gap is an indication of differences in cognitive performance to some extent, however, largely attributed to historical, socioeconomic, environmental, and psychological factors [17][21][22][20].

2.2.1 Educational Debt

The political, economic, and social underpinnings of the achievement gap is well encompassed by the term "educational debt." Coined by Gloria Ladson-Billings, it is a term often used in lieu of 'the achievement gap' to recognize its complexities and contextualize the problem. It frames academic disparities as a consequence of historical, economic, sociopolitical, and moral 'debt' accumulated by society as a result of decisions and policies that were made, and largely owed to minority, underprivileged, and disadvantaged groups [23]. For example, historical debt refers to the educational inequalities that initially manifested as a consequence of racial segregation and discriminatory policies. Sociopolitical debt refers to the lack of representation and exclusion of peoples in the civic process.

Most blatantly, economic debt is reflected in the disparities in school funding and spending. High-poverty neighborhoods, which tend to consist mainly of minorities, receive substantially less funding than affluent neighborhoods. In several urban areas, for example, expenditures per student is between \$3,000 and \$11,000 less for schools with high Hispanic and African American student populations compared to schools with lower minority populations [24][25]. Since schools are funded in part by local taxes, disparities in funding are also tied to the 'income' gap—differences in the average income between demographics. In 2016, the median income for Black and Hispanic households were

$31,082 and $30,800, respectively; for a White household, the median income was $47,948 [26]. Household income determines which communities, and thus quality school districts, are accessible to families.

The obvious consequence of the economic debt is an ‘opportunity gap’—-differences in the access/provision of quality schooling and teachers, and educational and basic resources between groups which contribute to the achievement gap [17]. We consistently see that schools with higher populations of African American and Latinx students are taught by inexperienced teachers, receive substantially less funding, offer fewer advanced courses, and have lower teacher salaries [25]. A study of high-performing minority students across the United States found that the gap between Black students and White or Asian students increased over the course of high school even with similar starting scores; of which socioeconomic status (SES) played a large role in explaining these differences. As mathematical content becomes more difficult i.e. in high school, school and teacher quality are critical in attaining higher levels of achievement. Moreover, students of higher-income households can afford additional resources outside of school as well such as tutors and additional summer courses [27].

While SES status is a highly significant predictor of academic achievement, the achievement gap cannot be attributed entirely to financial deficiencies or the opportunity gap. A study looked at academic development across groups by SES and race/ethnicity using a time-varying effects model (TVEM) over longitudinal data from the Peabody Individual Achievement Test (PIAT) in math (Figure 2.2). There are differences in test scores between groups in

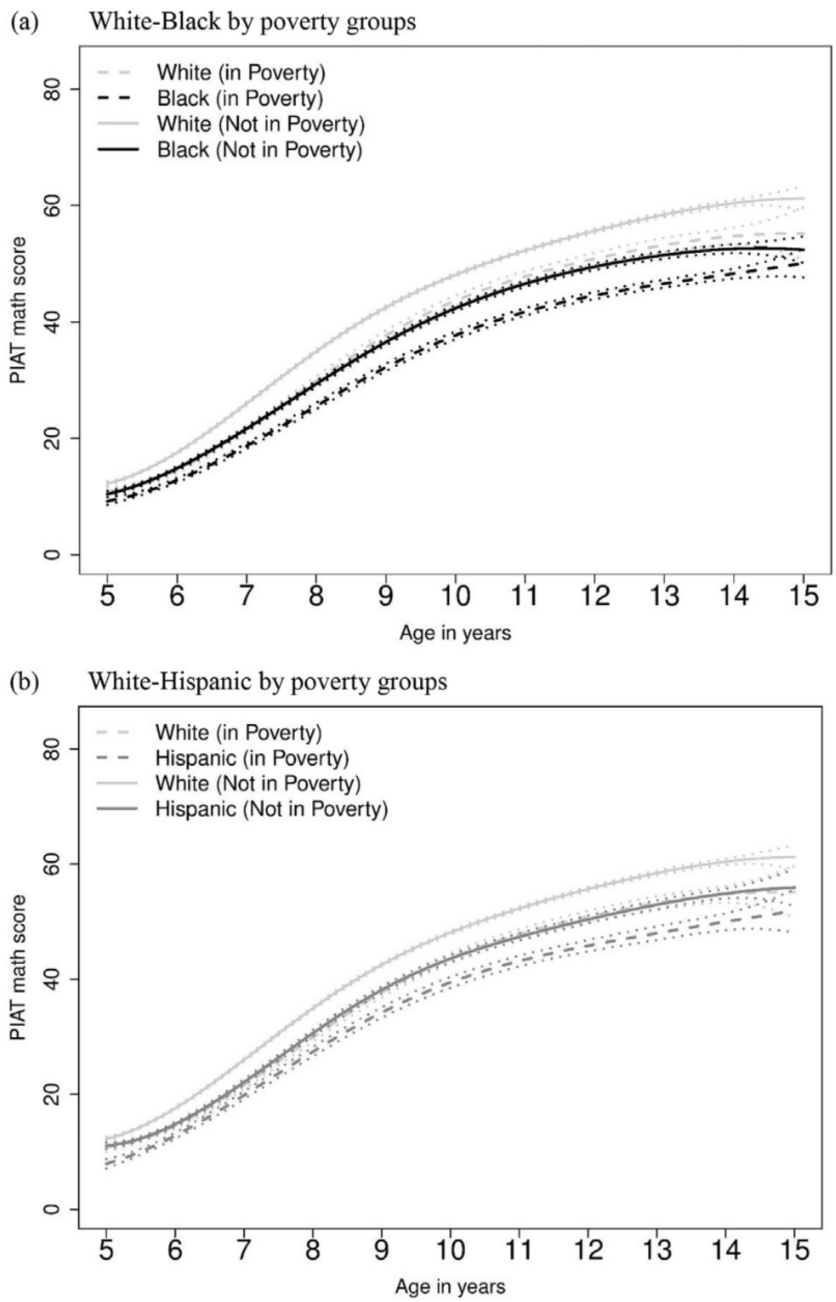


Figure 2.2: TVEM Model of PIAT Math Score over time in by a) White and Black students, and b) White and Hispanic students

poverty and not in poverty for each demographic group which only widen with time. More surprisingly, however, Black students *not* in poverty had similar trajectory of PIAT math score as White students in poverty. Similar relationships were found with the Hispanic group [18]. The range of income in the 'not in poverty' group may differ greatly between White, Black, and Hispanic groups which may partially explain why there is a gap between these groups. However, the similar scores for White in poverty and Black or Hispanic not in poverty further demonstrate that income or economic debt to explain the achievement gap cannot fully account for differences. By contextualizing the achievement gap in terms of other historical factors, educational debt provides additional insight into this issue.

2.2.2 Psychological Factors

The concept of educational debt provides a substantial background to the achievement gap. However, additional research in a separate domain regarding the influence of a student's attitude and affect on academic performance begins to explain a separate element of the story.

Affect describes a reactive and fundamental state of mind that is "ephemeral," for example an emotion or mood; it is an evaluative state of an object. Attitude describes a tendency to evaluate an object in a particular way, positive or negative, that is influenced by past experiences or beliefs, and influences future judgement and behavior [28][29][30]. Factors that influence students' affect and attitude have been shown to correlate with academic performance and competence. We will discuss a few constructs well-versed in the literature that influence student affect and attitude towards mathematics, which in turn influences mathematical performance/competency: self-efficacy and math anxiety.

Self-efficacy is an individual's belief in their ability to produce an important effect; in other words, how capable or effective one believes they are in solving/doing a problem/task. In the context of numeracy, this includes an individual's confidence in their ability to properly solve a math word problem. It is developed and constantly reshaped through interactions, exchanges, and experiences with one's environment. The extensive literature on self efficacy demonstrates the robustness of self-efficacy as a predictor of academic achievement; and similarly, we see that mathematical self efficacy is a significant predictor of academic performance in math [31] [27][32].

Notably, women tend to have less confidence in math than men which partially explains the gender gap in mathematics [33]. There are differences in self-efficacy among races as well as indicated in a study of high achieving White, Black and Hispanic high school students. However, its effects varied by race/ethnicity. For White and Hispanic students, a higher math self-efficacy predicted higher math achievement, as expected. For African American students, on the other hand, high self-efficacy predicted lower math achievement; and for Asian students who had the lowest self-efficacy performed the best of the groups [27]. As aforementioned, African American students are less likely to be placed in advanced courses and enroll in worse schools, contrary to Asian American students. Thus, they suspect that their high math self-efficacy is in part due to the 'Big Fish, Little Pond' effect—-they may be high achievers locally, but on a larger scale they are not as high achieving, relatively speaking. Although, this hypothesis does not explain how math self-efficacy affects Hispanic students' performance and requires further investigation. Ultimately, self-efficacy may be a good measure in predicting academic achievement for some demographic groups, however, for others it may prove to be more complicated.

Math anxiety describes the feeling of anxiousness, apprehension and tension driven by encounters with mathematics in ordinary or academic contexts. The generation of such negative affects impedes math performances, the cause often attributed to the limitation of working memory. High math anxiety can lead to math avoidance behaviors, negative attitudes towards math, and low mathematical confidence. This in turn has implications of math illiteracy and lower math achievement, self-efficacy, and competence [34][35]. In particular, math anxiety and math self-efficacy are strongly and negatively correlated with each other, but nonetheless are distinct constructs [36]. The causal direction between them, however, is unclear.

Math anxiety has been known to be much more prevalent in women than men, which in turn acts as another factor that drives gender differences in mathematics [34][37]. On the other hand, there were no statistically significant differences between racial/ethnic groups in U.S. adults in an exploratory study. They did find that students in particular have higher math anxiety than adults, so differences between racial/ethnic groups may precipitate within just the student demographic and is still worth investigating [38].

However, there are differences in math anxiety and self-efficacy across certain cultures outside the U.S. For example, students in Korea and Japan tend to have higher math anxiety and self-efficacy on average while still performing above average on math performance. This contradicts the results of previous studies. Similarly, the U.S. has high math self-concept and self-efficacy, while still performing below average on math performance [36]. It is important to note that these are comparisons between countries. While the students in the U.S may have an inflated sense of self-efficacy relative to other countries, it does not mean that it cannot accurately predict higher achieving students, from lower achieving students within the U.S. Nonetheless, we should note that there are differences in the predictive power of math anxiety and math self-efficacy between cultures, and potentially between specific demographic groups.

Provided with previous investigations on the achievement gap, and factors that drive differences in academic ability, we can proceed with the current investigation on what drives differences in quantitative literacy. While previous literature tends to focus on mathematics achievement or more general cases of academic achievement, we can suspect similar factors to explain differences in quantitative literacy, specifically, among various demographic groups.

2.3 Current Investigation

2.3.1 QuaRCS

To explore our question of interest, we will be working with the Quantitative Reasoning for College Science (QuaRCS) data set. The QuaRCS study uses its own validated quantitative literacy assessment to measure the quantitative skills of college students in general education science courses. In addition to students' performance on the assessment, it also uniquely collects both demographic variables and psychological variables [39][40]. Demographic information includes students' gender, race/ethnicity, and disability status; psychological variables measured were numerical self-efficacy, numerical relevancy, and academic maturity. Therefore, the QuaRCS dataset provides us with a novel opportunity to investigate the interrelationships between demographic, psychological variables, and numeracy (Figure 2.3). In the following sections, we provide more detail on these variables and previous results from the QuaRCS study.

QuaRCS Variables

While self-efficacy was touched on previously, numerical relevancy and academic maturity are newer constructs. Numerical relevancy refers to a student's perception of how frequently they encounter numerical situations. While not well-investigated specifically in the literature, we suspect that one's perception of its relevance is another factor influences mathematics achievement. When students recognize the relevance of material being taught, they have greater motivation to learn [41][42]. We suspect then that a greater sense of numerical relevance would positively impact their academic achievement. Academic maturity is a construct that emerged post-analysis[2], measuring student's level of academic development. While it is a completely new construct, we will soon

[2] more specifically, from factor analysis. See Section 4 for the methods with which we constructed the psychological variables

see that it was shown to be a significant predictor of score (though to a lesser degree than numerical self-efficacy or numerical relevancy). Finally, there is the level of effort the student put on the assessment. This variable can be considered either as an independent or dependent variable. In other words, use it as a variable to predict score, or use it as a method of control.

Previous QuaRCS Results

Results from the first set of administrations[3] of the QuaRCS assessment already demonstrate the under-performance of certain demographic groups, consistent with previous literature (Figure 2.4). There is a statistically significant gap in QuaRCS score between male and female students, and Caucasian and African American or Hispanic/Latinx students. However, in a multiple regression model predicting score, the predicting power of gender and race is minimal when these psychological variables are included in the model. Table 2.1 lists all the variables that were included in the final model listed in stages–where each stage contributes additional explanatory power (i.e. R^2) to the overall regression model. Numerical self-efficacy, numerical relevancy, and academic maturity themselves explain up to 32% of the variance[4] in score. When accounting for additional significant variables such as effort, major, and age, the developed regression model explain for almost 50% of the variance in score. Additional demographic variables added less that 0.1% to the model.

[3] Full Version 1.0, Pre 2017; see table 3.2

[4] to put it crudely, variance is a measure of the spread of data. So when we say that 32% of variance is explained, we mean that our model can explain the shape of the spread of data in part, or explains how the data 'behaves'

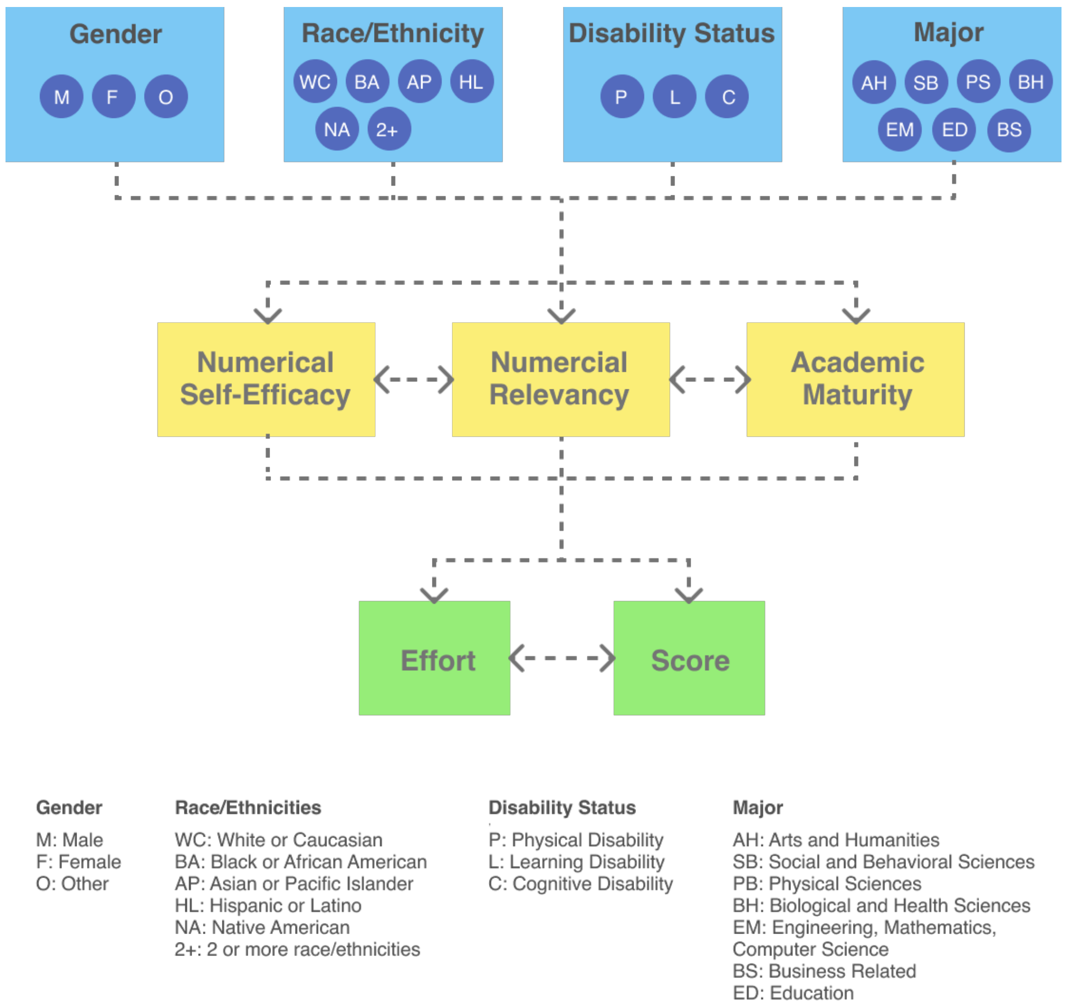


Figure 2.3: Measures of Demographic and psychological Variables and their relationships with effort and score on the QuaRCS Assessment

Therefore, we have the initial evidence that the under-performance of certain demographic groups may be explained by these psychological variables.

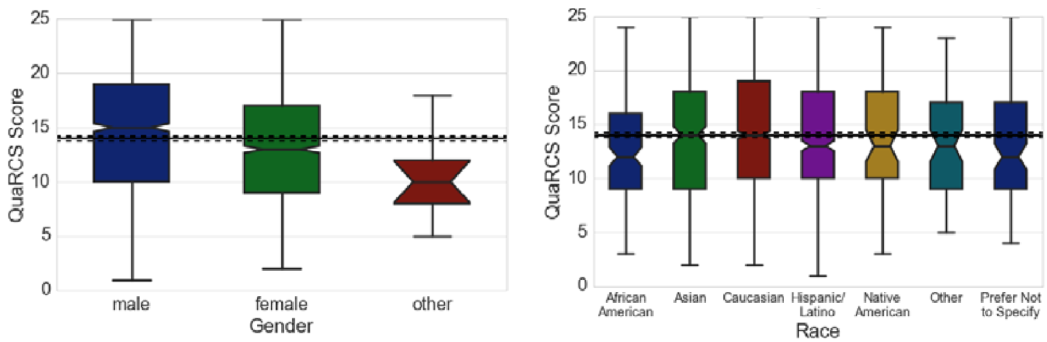


Figure 2.4: QuaRCS Score vs Gender and Race

Factor/variable	Standardized beta coefficient	Significance	Pearson's correlation coefficient (partial)	Model R^2
		Model 1: *Composite Affective Variables from Dimension Reduction*		
Numerical Self-Efficacy	0.42	<0.001	0.43	
Numerical Relevancy	0.15	<0.001	0.16	0.324
Academic Maturity	0.07	<0.001	0.09	
		Model 2: *Model 1+ Affective variables not appearing in or removed from factor solution*		
Effort	0.39	<0.001	0.44	0.483
Calculator Usage	-0.07	<0.001	-0.10	
		Model 3: *Model 2+ Academic Variables*		
Chose my major because: Other	0.05	<0.001	0.07	
Numerical skills are important to my daily life	0.09	<0.001	0.09	0.489
Humanities major	0.05	<0.001	0.07	
STEM major	0.06	<0.001	0.09	
		Model 4: *Model 3+ Basic Demographics*		
Learning disability	-0.03	0.02	-0.05	0.493
Age	0.03	0.02	0.05	

Notes: Summary of the regression model that produces Figure 7. Variables were entered in four tiers, and the additional descriptive power added by the variables in each tier can be seen in the change in the R^2 value as that tier is added to the model, where R^2 represents the proportion of variation in QuaRCS score explained by the model.

Table 2.1: Final Regression Model from Follette 2017

2.3.2 Hypotheses and Groups of Interest

In our investigation, we look more specifically into the relationships between the psychological variables measured and to what extent they explain differences in score among certain demographic groups. We will also be using a newer iteration of the QuaRCS assessment data, a 'lite' version[5] with additional affective questions such as math anxiety and a new demographic question on the highest education attained by a parent, which will provide a different territory of exploration.

From the literature and previous results, we suspect that there will be significant differences in psychological variables within demographic groups which impact their performance on the QuaRCS assessment. In particular, we hypothesize that lower math self-efficacy and higher math anxiety explain, in part, the lower levels of quantitative literacy in minority groups such as Black, Hispanic, and Native American students. Higher math self-efficacy and lower math anxiety will explain the performance of the higher achieving White demographic group. However, for Asian and multiracial students who tend to perform better on mathematics exams than other minority groups, will have similar levels of math self-efficacy and math anxiety as White students. Because our population of study are college students, the varying effects of self-efficacy on achievement discussed in Section 2.2 may not pertain to this group.

We suspect that there will be small differences in psychological variables between male and female students as the gender gap is fairly small; however, we do expect women to have higher math anxiety and lower self efficacy.

The new demographic variable, the highest level of education attained by a parent (Parent Education), can serve as a proxy to socioeconomic status

[5] see Table 3.2

(SES), as higher educational attainment are generally associated with higher levels of income[43]. Thus, we predict that students with parents with less educational attainment, or lower socioeconomic status, will have lower self-efficacy and higher math anxiety.

The QuaRCS Lite also includes other variables that probe student's math affects that the literature does not explore specifically. We will explore first-hand how they may contribute to differences in score, and how they differ among demographic groups.

Chapter 3

QuaRCS Methods and Data Processing

3.1 Study Population

The QuaRCS study aims to measure numeracy skills in college students, particularly students enrolled in introductory general-education (gen-ed) science courses. These courses are tailored to the general student body, open to all regardless of their previous background in the sciences. As a result, such courses may contain a greater proportion of students who are not exposed to the more quantitative-heavy courses, and perhaps may be in the most need of developing quantitative skills. Therefore, these gen-ed courses are seen as crucial venues for the development of such skills. Additionally, general education science courses may be the final quantitative course taken at college for a large proportion of students [39]. Thus, the QuaRCS study is focused on investigating this particular population, and understanding to what extent intro gen-ed courses are successfully (or unsuccessfully) improving quantitative literacy.

The QuaRCS study, however, does not limit its administration to only general education science courses. It collects data from students enrolled in introductory science and math courses for majors as well. We have removed students from such courses from our data set, as general education and non-general education populations have shown to be statistically different. In other words, the validity and reliability of the QuaRCS instrument was only demonstrated with a non-science major population [39].

3.2 Assessment Development and Revision

The assessment was developed using a bottom-up approach, surveying science and math educators about which numerical skills they deemed the most important, then developing questions that targeted these skill-areas. Fifteen numeracy experts were asked to categorize the developed questions by what numerical skill(s) it assessed. The question would be identified as measuring a particular numerical skill if seven or more educators agreed on its categorization. This approach establishes content validity; in other words, it ensures that what the assessment measures covers the entire breadth of what we seek to measure. Figure 3.1 gives a descriptive list of all numerical skills tested on the QuaRCS assessment, the number of questions that fall in each category, and the difficulty range of the questions. Difficulty is measured by the p_D-value, an indicator of how many students answered the question correctly. The lower the p_D-value, the more difficult the question.

Skill	Definition	Abbrev.	N_{Quest}	p_D-value range
Graph Reading	Read, interpret or extrapolate graphical data.	GR	5	0.22-0.76
Table Reading	Read and interpret information presented in tabular form.	TR	3	0.35-0.80
Arithmetic	Add, subtract, multiply or divide two or more numbers.	AR	21	0.24–0.80
Proportional Reasoning	Compare two or more numbers, rates, ratios, fractions.	PR	13	0.24–0.75
Estimation	Approximate an answer or choose the closest value to a precise calculation.	ES	4	0.24–0.76
Percentages	Compute or compare percentages	PC	5	0.28–0.73
Statistics and Probability	Statistics= interpretation of data, including distributions and descriptive statistics (mean, median, mode, etc). Probability = compute odds or risk or determine the most likely outcome.	SP	6	0.22–0.59
Area and Volume	Compute or compare areas or volumes	AV	5	0.48–0.68
Error	Evaluate uncertainty in graphs or numbers	ER	4	0.22–0.36
Unit Conversions and Dimensional Analysis	Unit Conversions = Use the relationship between two or more units to transform one number into another. Dimensional Analysis = Draw inferences about the relationship between two or more quantities based on the units attached to them.	UD	6	0.30–0.75

Table 3.1: Numerical skills tested on the QuaRCS assessment

The QuaRCS Full Version of the assessment, contains 25 quantitative questions each followed by a question asking about the student's confidence in their answer. The section is followed by 25 non-quantitative questions. These include demographic questions about gender, race/ethnicity, major, and disability status, and questions that ask how frequently quantitative problems are encountered, the student's confidence in their quantitative skills, and their reasons for choosing their particular gen-ed course and their major. The final question on the assessment asks how much effort the student put on the exam. A full list of questions administered on the assessment can be found in Table B.1 in the Appendix.

However, due to the concern of test fatigue, a shorter version of the assess-

Table 3.2: Versions of the QuaRCS assessment

Version	Semesters Administered	# of Questions
Full 1.0	Fall 2015 - Spring 2017	25
Full 2.0	Fall 2019 - Present	25
Lite 1.0	Fall 2016 - Summer 2017	11
Lite 2.0	Fall 2017 - Spring 2019	15
Lite 2.1	Fall 2019 - Present	15

ment, QuaRCS Lite, was created. QuaRCS Lt1.0 was piloted with 11 quantitative questions. Additional affective questions were included, and asked about levels of anxiety in various quantitative situations and levels of difficult, stress, and confusion according to context, namely using math in daily life or school. However, unlike the Full version, the number of questions (11) proved to be too few for sufficient validation of the test, resulting in a low Cronbach's α, a statistical measure of validity. In other words, the assessment could not accurately measure each quantitative skill as intended. As a result, a second light version, QuaRCS Lt 2.0, with 15 quantitative questions was created and administered since 2017. This version is validated, with a Cronbach's α of 0.81. Table 3.2 gives an overview of the versions of the QuaRCS assessment. Tables B.4 and B.5 in the appendix gives a list of questions that have been asked or are currently asked in each version of the assessment.

In preparation for the work in this thesis, we made additional changes to the demographic questions (QuaRCS Lt.2.1). The options offered in a question asking about a student's race or ethnicity were expanded, as previous options were limited. For instance, students who chose 'other' as their race on the Full assessment often specified their race as 'Middle Eastern' or 'Black'. Thus, we updated the question to include an ‘Arab or Middle Eastern’ option and revised ‘African American’ to ‘Black or African American’. We also provide more specific options for the race option ‘Asian’ as it covers a variety of distinct ethnic groups, both minority and majority populations. When ‘Asian’ is selected, ‘East Asian’, ‘South Asian’, ‘Southeast Asian’, and ‘Native Hawaiian or Pacific Islander’ are now offered as options.

A question on the highest education attained by a student’s parents was also added. An individual’s level of educational attainment is correlated with their socioeconomic status (SES); thus, we will use the highest education variable as a proxy for SES [44]. As discussed in Section 2.2, SES is an important factor in academic achievement and likely a valuable piece of information for understanding a student’s quantitative achievement. While a direct question on SES could be asked, we suspect that a student will be able to answer a question on their parent’s highest education more accurately than their income level. The question also allows us to simultaneously identify first-generation students in addition to estimating their SES.

Because there have been various changes to the assessment, some versions are not completely compatible with others. For instance, QuaRCS Full 1.0 cannot be combined with QuaRCS Lt 1.0 or 2.0 because of the different number of quantitative and non-quantitative questions. QuaRCS Lt 1.0 cannot be used for its assessment score since the instrument failed to meet standard validity measures. Therefore, there will be 4 ‘master’ data sets that we will perform analysis on; the contents and uses of each are tabulated in Table 3.3.

Table 3.3: QuaRCS Data Sets. X denotes full use of a data set; x denotes use of half

Data Set	Full	Lt. 1.0	Lt. 2.0	Lt. 2.1	Uses
S0	X				SPSS vs R comparison Correlation matrix comparison
S1†		x	x	x	Exploratory Factor Analysis
S2†		x	x	x	Confirmatory Factor Analysis
S3			X	X	Regression Models
Sample Size	2736	1031	1180	1240	

† S1 and S2 were created by combining all Lt. 1.0, Lt. 2.0, Lt. 2.1 assessments, randomizing the order, then splitting it in half.

3.3 Data Collection, Scoring, and Cleaning

The assessment was administered through Qualtrics, a survey collection and management software. Instructors were given access to the survey and asked to administer the assessment at the beginning and end of the semester.

From Qualtrics, the raw data from each individual student and course is retrieved for 'pre' or 'post' semester assessments then run through the scoring and cleaning Python script. The scoring script calculates the total score on the quantitative section in percent and sub-scores for each numerical skill measured on the assessment. A total and average confidence score is also calculated, from the student's responses regarding their confidence in each answer. The cleaning script drops irrelevant columns (e.g. IP address, operating system) and incomplete assessments. It also tidies the dataset by filling in empty cells with NaN (Not a Number) values, renaming, and consolidating variables. Next, the cleaned data set is anonymized by assigning each student an ID number, which is then matched with their instructor's anonymized ID[1]. 'Pre' and 'Post' assessments are run through the Python scripts individually, then matched by student ID and merged into one complete semester data set. The first row in Figure 3.1 illustrates the flow of this process. After scoring multiple semesters of data, we can merge them to create a master data set(s) to perform further analysis on. These steps will be discussed in the following chapters.

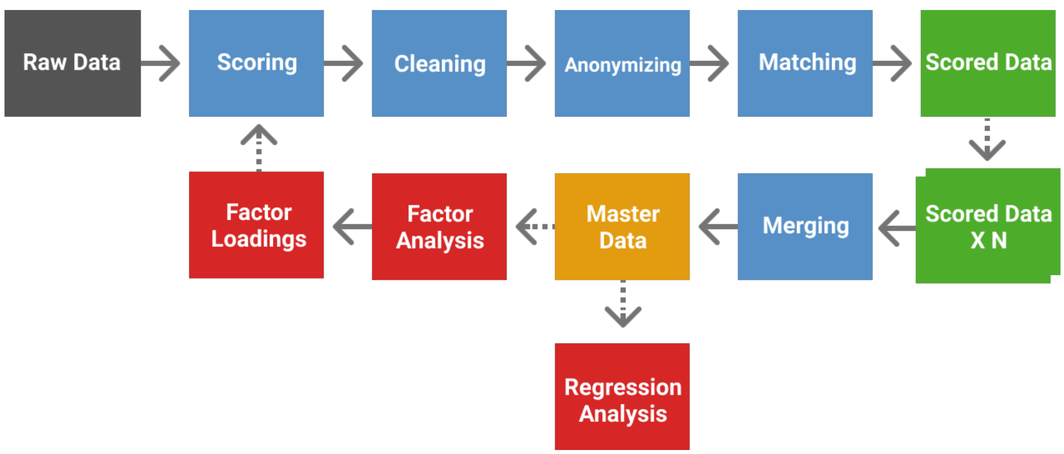


Figure 3.1: Data Processing Flow Chart

3.4 Data Wrangling

Data wrangling is a process that transforms or maps data into a form that is more useful or easily interpreted without changing its weight or value. For example, standardizing scores, or renaming variables are common wrangling

[1]Anonymization is needed per IRB requirements and as data from QuaRCS assessments are available online through public database

steps. We carried out several wrangling steps on our master data sets. The wrangling code can be found in Appendix D.

The first was to rename variable or column names so that they were (1) shortened and/or (2) more descriptive. Next, we relabeled the numeric indicators of demographic variables into an appropriate string. For example, for the GENDER column, values of 1, 2, 3 were labeled 'male', 'female', and 'other', respectively. Unlike the gender or highest education variables, disability, major and race variables are distributed across many columns–one for each option–to accommodate the ability of students to select multiple responses. Thus, a choice 'white/caucasian' would be indicated by a value of 1 in the column, RACEV2_WHITE. For each of these columns, we labeled values of 1 with their respective descriptor, e.g. in the previous case, 'white'. Because the race variables are separated by column, we collapse all the columns into one aggregated RACE column. Students who chose multiple races were identified as 'multiracial.' Similar steps were performed with disability status and academic major variables.

However, there is a complication with collapsing the race variable between different assessment versions because of the recent modification of the race question. We labeled race/ethnicity values based on the second version of the question, which does not completely correspond to the first iteration. In other words, there are some students that may not have been correctly categorized. For example, students who identified as Middle Eastern on assessments prior to QuaRCS Lt. 2.1 have previously chosen 'other', and would not be correctly categorized as 'Arab or Middle Eastern.' Therefore, the 'Other' category includes a mix of students that may technically fall under a more appropriate category. The 'Asian' category also contains students who identify as 'Pacific Islander' or 'Southeast Asian' in Lt 2.0 versions of the assessment, which would otherwise be placed in the 'Asian Minority' category.

To make the initial process of exploring the data easier, we categorized demographic groups into an underrepresented (UR) group and non-underrepresented (Non-UR) group. The distinction was made post analysis. Generally, demographic groups with higher numeracy scores, tend to be a non-underrepresented group in the sciences, and demographic groups with lower numeracy scores tend to be an underrepresented group in the sciences. The parent education variable was also split into students with parents with post-secondary education and without post-secondary education[2]. Note that we will later interpret these groups as high socioeconomic and low socioeconomic groups. Table 3.4 indicates which demographic groups are being considered as an underrepresented or non-underrepresented group.

3.5 Summary Statistics

To get a general understanding of our sample, here we provide a general summary of data set S3[3].

Figure 3.2 gives the distributions of score and effort on the assessment. Medians are indicated by the dashed vertical line. A majority of students

[2]For the sake of consistency (rather than precise terminology), 'Secondary or Less' is labeled as URG Education in the code, and 'Post-Secondary' is labeled as Non-URG Education

[3]Our analysis using demographic variables will utilize S3, rather than S0, S1, or S2. Thus, statistical summaries for those datasets will not be provided.

Table 3.4: UR and Non-UR Groups Identified

	UR	Non-UR
Gender	Female, Other	Male
Race	Black or African American, Hispanic or Latino, Native American, Asian Minority (South East Asian, Native Hawaiian or Pacific Islander, Other), Multiracial Minority	White, Asian, Multiracial, Arab or Middle Eastern,
Disability	Cognitive, Learning	None
Parent Education (SES)	Elementary education, secondary education, technical certificate, Associate's degree	Bachelor's, Master's, Professional Degree, Doctorate Degree

do exert high levels of effort (efforts 4,5), with a significant decrease in the number of students with lapsed effort (effort 3) compared to the previous administration of QuaRCS Full assessment[4]. This reflects the efficacy of the Lite version in collecting more accurate quantitative literacy scores of students who put effort into all questions. The demographics of our sample is described in Table 3.5. Note that certain demographics have only been recorded since the administration of QuaRCS Lt 2.1. In particular, the question asking about the students' parent's education was only administered since Fall 2019, explaining the large number of no responses (N/A).

[4]Distribution of effort on the QuaRCS Full 1.0 can be found in Appendix B.1

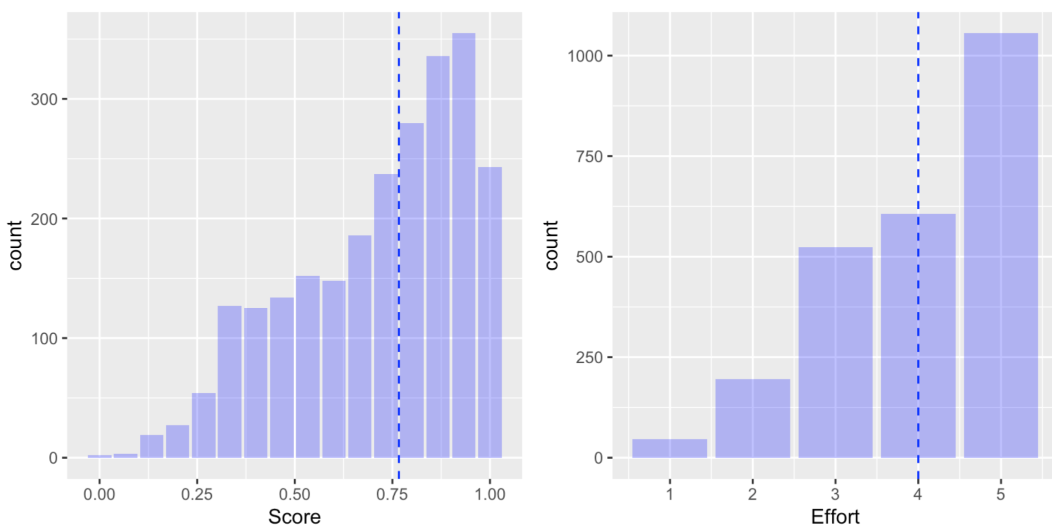


Figure 3.2: Distribution of Score and Effort. Medians are marked by the dashed line

Table 3.5: Demographic Summary

Gender	n	%
Female	1287	53.0
Male	1091	44.9
Other	37	1.5
N/A	13	0.5
Total	2428	100
Race	n	%
Arab or Middle Eastern	15	0.6
Asian (East Asian, South Asian)	410	16.9
Black or African American	186	7.7
Hispanic or Latinx	376	15.5
Multiracial	219	9.0
Native American	11	0.5
Other	44	1.8
Asian Minority (SE Asian, Pacific Islander, Hawaiian)	96	4.0
Multiracial Minority	80	3.3
White	864	35.6
N/A	127	5.2
Total	2428	100
Parent Education	n	%
No Secondary Education	115	4.7
Secondary Education	256	10.5
Associate's Degree	113	4.7
Technical Certification	41	1.7
Bachelor's Degree	326	13.4
Master's Degree	264	10.9
Professional Degree	84	3.5
Doctorate Degree	81	3.3
N/A	1148	47.3
Total	2428	100
Effort Level	n	%
5: Tried My Best	1056	43.5
4: Tried Pretty Hard	607	25.0
3: Stopped Trying Midway	524	21.6
2: Didn't Try Hard	195	8.0
1: Randomly Chose Answers	46	1.9
Total	2428	100

Chapter 4

Factor Analysis

The QuaRCS data provides more than 25 questions to consider for analysis. However, the non-quantitative questions asked on the assessment were designed to probe at underlying psychological variables such as numerical self efficacy or numerical anxiety. Thus, many of the questions are related to one another, while also prompting unique responses. Instead of using the 25 questions as separate variables for analysis, we can aggregate them to form several composite variables using a technique called factor analysis.

In this section, we provide an extensive overview of how we developed composite variables (or later referred to as *factors*) from the initial set of attitudinal questions using data set S1. The results of this section will be very useful as a dimension reduction technique for later analytical work. Corresponding code can be found in Appendix D.

4.1 Exploratory Factor Analysis

Factor analysis is a technique that is used to condense multiple variables measured in an experiment into a few unobserved, indirectly measured, variables called factors [45]. For example, if an experiment measures participants' ability to mentally visualize, rotate, or manipulate objects, these measures may be aggregated to create an overall factor that an experimenter interprets as a measure of "spatial intelligence." Typically, an experiment is designed to measure variables with some preconception of how they are related to one another. Even without this preconception, factor analysis ultimately eases the use and interpretation of data by reducing the number of variables and helps to eliminate issues of multicollinearity[1]. In our case, the 25 attitudinal questions (listed in Appendix B.5 are the variables which we seek to condense.

In more abstract terms, the factor is like a vector (think a line in 2D or 3D space) with an associated numerical value (later known as the factor score) that is a linear combination of each measured variable multiplied by a corresponding weight or loading. It is constructed such that the deviation of each observance from the vector is minimized. The loadings can be interpreted similarly to regression coefficients, which describe how much each variable contributes to the factor, and describes its relationship with the factor (positive or negative). The sum of squared loadings is called an eigenvalue, which describes the variance each factor explains. The higher the eigenvalue (> 0), the more our factors explain the variance of the data. Essentially, our factor

[1]multicollinearity is an indication of high correlations between two or more variables, such that one predicts the other

aims to describe, as best as possible, an underlying variable that cohesively explains measured variables as a group [45][46].

To explain the precise process of factor analysis is beyond the scope of this thesis; in fact, much of the process is done computationally using R or Python modules. However, there are preliminary considerations to make before generating factors and their loadings in R or Python, which we will describe in the following sections [2]

4.1.1 Correlation Matrix

Before performing factor analysis, we must first check whether there is a good incentive to proceed. A simple justification is a correlation matrix [3] with pairs of variables with high correlations which tend to aggregate onto the same factor. This indicates that there may be an underlying factor that explains multiple variables simultaneously and can be determined via factor analysis. However, variables that are extremely correlated should be removed (>0.90) as it does not add a unique contribution to the factor [47]. The correlation matrix is also used as an input into the factor analysis function in R or Python.

The method of calculating coefficients can vary depending on the value-type of each variable–whether it is numerical, ordinal, binary, or nominal. The Pearson correlation coefficient is used between numerical variables; the Spearman correlation coefficient is used between numerical or ordinal variables; the polyserial coefficient is used between numerical and ordinal variables; the

[2] In previous work, factor analysis was done via SPSS; since I am not familiar with SPSS syntax, I used R instead. Section C in the appendix compares the outcomes of SPSS and R

[3] a table of values of correlation coefficients between all variables included in factor analysis

polychoric coefficient is used between ordinal or binary variables. Since our variables of interest have a mixture of value-types–numerical, ordinal, and binary–there is reason to compute different correlation coefficients for each pair of variables. However, this would convolute subsequent analysis, so choosing one coefficient is advantageous in its simplicity. To compare these approaches, we created three separate matrices from previous data (QuaRCS Full) and generated the factor loadings (Figure 4.1). Pearson and Spearman columns use their titular coefficient matrix; Hetcor columns use a heterogeneous correlation matrix, appropriately applying Pearson, Polychoric, and Polyserial coefficients to each variable according to its type.

Table 4.1: Comparison of factor loadings using various correlation matrices

	Factor 1			Factor2			Factor3		
	P	S	H	P	S	H	P	S	H
PRE_LK1	0.886	0.896	0.961						
PRE_LK2	0.830	0.836	0.903						
PRE_ATT_3	0.590	0.588	0.619						
PRE_DIFF	0.592	0.576	0.606						
CONF_AVG	-0.530	-0.551	-0.516						
PRE_DAILYM				0.729	0.725	0.779			
PRE_DAILYG				0.655	0.662	0.697			
PRE_FREQEN				0.587	0.590	0.634			
PRE_ATT_2				-0.508	-0.491	-0.584			
PRE_LK4				-0.430	-0.403	-0.475			
PRE_ATT_1				-0.399	-0.385	-0.449			
PRE_LIKEJOB							0.599	0.596	0.619
PRE_GOOD							0.528	0.528	0.690
PRE_MONEY							0.493	0.491	0.500
PRE_LIKE							0.445	0.448	0.708
PRE_NOTSURE							-0.429	-0.431	-0.829

Pearson (P), Spearman (S), Hetcor (H)

The loadings between Pearson and Spearman differ by no more than 0.03, while Hetcor loadings differ more substantially. However, the relative magnitudes/relationships between the loadings are similar between methods. The only exceptions are the binary WHYMAJ variables, which load onto a factor we named 'academic maturity.' Previously, this factor has not shown any

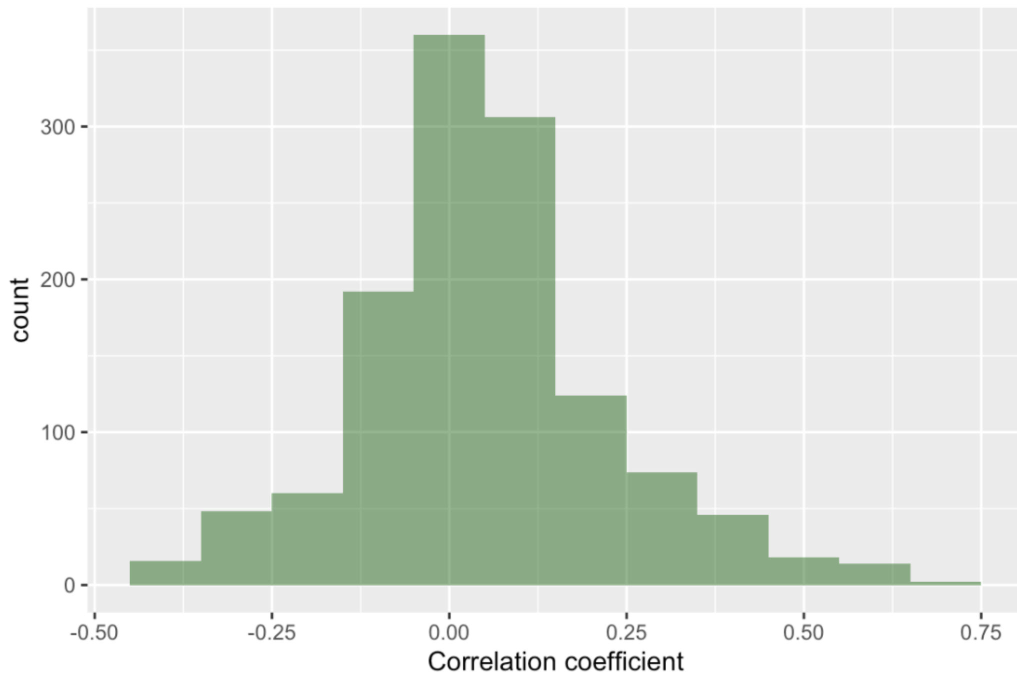


Figure 4.1: Distribution of correlations between variables of interest

significance in our models predicting QuaRCS score and will not be used in subsequent analyses. Therefore, to keep our interpretations simple, we decided to proceed with the more versatile Spearman correlation coefficient.

In a histogram of correlations in our correlation matrix, there are a number of high correlations (both negative and positive) between our variables, and thus can proceed towards factor analysis (Figure 4.1).

4.1.2 Principal Component Analysis & Parallel Analysis

Another consideration to make is the number of factors we want to extract from our data. This decision is guided by Principal Component Analysis

(PCA) or Parallel Analysis. PCA is a procedure similar to factor analysis but assumes the variables in the data have no intrinsic error in themselves that would contribute to the overall variance of the data. Components (analogous to factors) are extracted from our measured variables with an associated value called an eigenvalue. This describes how much the component explains variance in the data. In our case, we have 36 measured variables from which to extract components/factors.

When we plot the components and their corresponding eigenvalues, a plot known as a scree plot, there often is a kink or 'elbow' where the eigenvalues start to steady. There are two criteria that are typically used to decide how many factors to extract using PCA: (1) the number of components above the elbow on the Scree Plot or (2) the number of components with an eigenvalue greater than 1. However, there is no consensus on which method to use because of the varying results and interpretations of scree Plots [45][46]. For example in our scree Plot, the location of the elbow is not particularly distinct, and may be at 2 or 3 or 6 components (Figure 4.2); and ten components have an eigenvalue greater than 1. Thus, the scree plot does not provide an immediate or clear suggestion on how many factors should be extracted.

We can also turn to Parallel Analysis, another method of determining the appropriate number of factors said to be more accurate than PCA. First it involves generating a 'parallel data set', which contains random values for each variable. Then, it performs exploratory factor analysis with the parallel data, and computes the factor's eigenvalues. It repeats this process 500-1000 times, generating a new set of eigenvalues with a new parallel data set. These eigen-

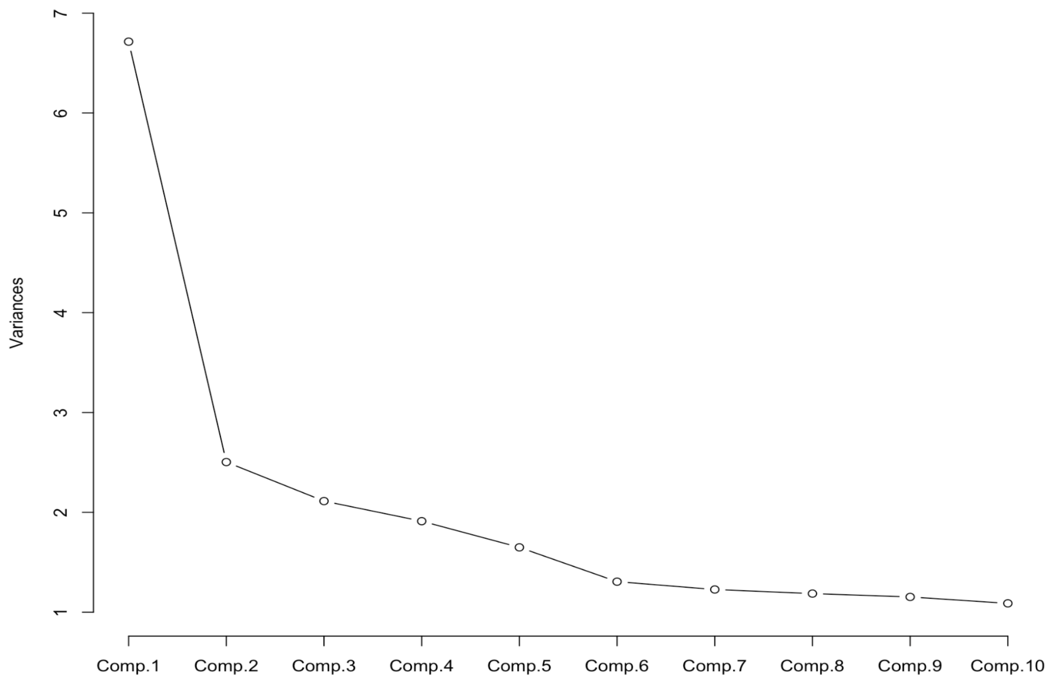


Figure 4.2: Scree Plot

values (Random Ev) are averaged then compared with eigenvalues determined from the real data set (Unadjusted Ev). If the difference (Adjusted Ev) is greater than 0, then that eigenvalue's factor is retained [46]. Essentially, a factor is retained if its eigenvalue is different from a randomly generated one. The results of parallel analysis recommend that we retain 17 factors (Figure 4.3)[48]. As the purpose of factor analysis is to condense the number of variables, this procedure did not prove to be helpful.

Remember that PCA and Parallel Analysis are guidelines. Ultimately, the decision of how many factors to retain is based on the interpret-ability and utility of our factors. The factors that are retained must be meaningful in terms of the variables as a collective. Knowing the bounds of 2 - 17 factors

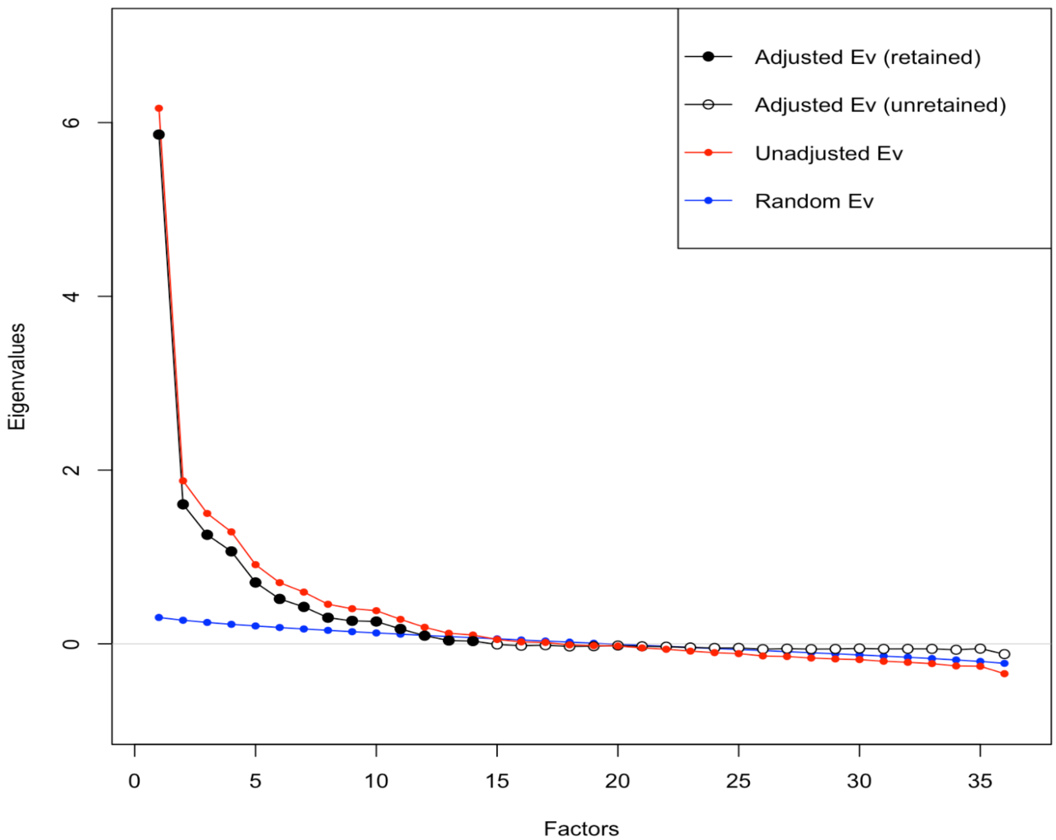


Figure 4.3: Parallel Analysis Plot

we can iterate through factor analysis and settle on a number of factors that seems reasonable given our preconceptions about how our variables should group together.

4.1.3 Rotations

Factors can be defined to be orthogonal or oblique i.e. independent or dependent of one another depending on the suspected relationships between the variables. In our case, we assume our factors to be correlated to some extent as we expect the likely factors (e.g. numerical anxiety, numerical self efficacy) to be related to one another; thus, we choose to extract oblique factors. This also determines how loadings are maximized onto one factor rather than several factors by a rotation process. This eases the interpretation of the factors by identifying variables that load strongly onto a particular factor and nothing else. There are several options for rotation. In the case of an oblique rotation, the 'oblimin' and 'promax' rotations are the most common. Because our dataset is large ($N > 500$), we use the 'promax' rotation, which is computationally faster and better suited for large datasets [45][46].

4.1.4 Factor Solution

Now that we know which correlation matrix to use, a range of how many factors to retain, and what kind of factors we desire, we can begin the process of finalizing our factor solution. Ultimately, our factor solution will contain a list of variables that contribute to a single factor.

Finding a Factor Solution

Again, this is an iterative process. We begin with our full selection of potential variables then run it through the factor analysis code[4]. Because we still have yet to determine how many factors to retain, we begin arbitrarily with 5 factors and compare this factor solution to solutions with different numbers of factors. In doing so, we find that the variables distribute onto 6 factors fairly well. A 5-factor solution aggregated variables that we believed to be distinct from each other (i.e. variables that are later associated with numerical self-efficacy

[4]See Appendix B.5 for the list of questions associated with each variable

and numerical anxiety). A higher term solution yielded factors with two or less variables; a factor with less than three variables is not practical since we want to reduce the number of variables as succinctly as possible. A 6-factor solution provided factors that were more interpretable and aligned with our initial assumptions. The first run with 6 factors is shown in Table 4.2.

Table 4.2: Factor Solution. All variables included

Variable	ML1	ML2	ML4	ML5	ML3	ML6	h2	u2	com
PRE_ATT_SC_1	-0.02	**0.76**	-0.09	0.06	-0.01	0.03	0.58	0.42	1.04
PRE_ATT_SC_2	0.11	**0.46**	**0.45**	-0.09	0.10	-0.06	0.47	0.53	2.35
PRE_ATT_SC_3	0.05	**0.34**	**0.47**	-0.18	-0.01	0.05	0.35	0.65	2.18
PRE_ATT_SC_4	-0.04	**0.78**	0.00	-0.03	0.00	0.03	0.61	0.39	1.01
PRE_ATT_SC_5	-0.10	**0.77**	-0.06	-0.02	0.04	-0.01	0.61	0.39	1.06
PRE_ATT_DL_1	**-0.43**	0.10	0.02	0.21	-0.06	0.01	0.39	0.61	1.64
PRE_ATT_DL_2	-0.06	0.10	**0.53**	0.04	0.04	-0.05	0.40	0.60	1.14
PRE_ATT_DL_3	-0.26	-0.12	**0.59**	-0.15	0.00	0.00	0.37	0.63	1.61
PRE_ATT_DL_4	**-0.41**	0.14	0.05	0.17	-0.05	-0.01	0.36	0.64	1.65
PRE_ATT_DL_5	**-0.46**	0.18	-0.04	0.19	0.01	-0.06	0.43	0.57	1.73
PRE_CF_MEAN	-0.27	-0.04	0.12	**0.34**	0.13	0.09	0.43	0.57	2.69
PRE_LK1	-0.02	-0.01	0.01	**0.83**	0.05	0.15	0.76	0.24	1.08
PRE_LK2	-0.06	-0.09	0.03	**0.79**	0.03	0.17	0.70	0.30	1.14
PRE_LK5	-0.03	0.08	-0.22	**0.63**	-0.01	0.10	0.35	0.65	1.33
PRE_DAILYM	0.06	-0.07	**0.58**	0.02	-0.12	0.16	0.32	0.68	1.31
PRE_DAILYG	0.02	0.00	**0.55**	-0.04	0.03	0.07	0.30	0.70	1.06
PRE_FREQEN	-0.01	-0.08	**0.52**	0.00	0.03	0.08	0.27	0.73	1.10
PRE_ANX.1_1	**0.56**	-0.12	0.00	-0.07	0.03	0.01	0.41	0.59	1.14
PRE_ANX.1_2	**0.72**	-0.01	0.01	0.05	-0.04	0.04	0.48	0.52	1.02
PRE_ANX.1_3	**0.63**	0.12	-0.01	0.02	0.06	0.00	0.34	0.66	1.10
PRE_ANX.1_4	**0.70**	0.01	-0.03	0.09	0.05	-0.01	0.41	0.59	1.05
PRE_WHYMAJ_1	-0.12	-0.04	0.12	-0.08	0.17	0.20	0.13	0.87	3.85
PRE_WHYMAJ_2	0.07	0.00	0.08	0.03	0.01	**0.59**	0.35	0.65	1.07
PRE_WHYMAJ_3	0.07	-0.01	0.07	0.06	-0.03	**0.57**	0.32	0.68	1.09
PRE_WHYMAJ_4	-0.04	0.03	-0.04	0.08	0.23	**0.44**	0.31	0.69	1.67
PRE_WHYMAJ_5	-0.01	-0.04	-0.18	-0.16	0.10	0.10	0.10	0.90	3.34
PRE_WHYMAJ_6	0.02	0.00	0.02	0.08	0.02	0.18	0.04	0.96	1.45
PRE_WHYMAJ_7	0.06	0.02	-0.04	-0.01	0.10	**-0.33**	0.12	0.88	1.28
PRE_WHYMAJ_8	0.00	-0.02	0.04	0.02	0.04	-0.12	0.02	0.98	1.70
PRE_WHYCS_1	0.15	0.02	0.18	0.03	**-0.46**	0.15	0.24	0.76	1.80
PRE_WHYCS_2	-0.04	0.00	-0.20	-0.09	0.00	0.10	0.07	0.93	2.04
PRE_WHYCS_3	-0.03	0.01	0.08	-0.03	**0.63**	-0.04	0.41	0.59	1.05
PRE_WHYCS_4	0.08	0.03	-0.02	0.08	**0.32**	0.01	0.11	0.89	1.29
PRE_WHYCS_5	0.08	0.01	0.03	-0.05	**0.70**	-0.01	0.46	0.54	1.04
PRE_WHYCS_6	0.09	-0.01	0.02	0.02	**0.63**	0.01	0.38	0.62	1.05
PRE_WHYCS_7	-0.01	-0.02	0.02	0.06	0.00	-0.02	0.01	0.99	1.79

The leftmost column lists our variables of interest and the rows indicate how much of that variable loads onto each suspected factor, ML (for Maximum Likelihood, a common method of factor analysis). Variables load onto all factors, but load more strongly onto some. These loadings are bolded (>0.30) in Table 4.2.

In addition to the factor loadings, there are three statistical measures given for each variable: communality (h2), uniqueness (u2), and Hoffman's complexity index (com). Communality ranges from 0 to 1 and describes how much of the variable's variance is explained by our extracted factors. Unique variance is the portion of the variance that is not explained by the factors $(1-h2)$ [45][49]. Higher communalities and lower uniqueness are ideal, and correlate to higher factor loadings. Often variables with communalities that are less than 0.2 (or u2 >0.8) are removed [50]. The complexity index describes how many factors a particular variable helps explain. To minimize multicollinearity, variables with a complexity index greater than 1 should be removed.

First, we remove variables that are not explained by our factors i.e. values of high uniqueness. These include WHY_MAJ_1, WHY_MAJ_5, WHY_MAJ_6, WHY_MAJ_8, WHY_CS_2, WHY_CS_7 ($u2 > 0.90$). The loadings change slightly after each removal, thus we must remove variables one at a time. In this process, the rest of the WHY_MAJ variables also are dropped, and we are left with Table 4.3, where all communalities are greater than 0.2.

Note variables like ATT_SC_2 , which cross-loads, or strongly loads onto two factors, ML 2 and ML4. It is a confounding variable, indicated quantitatively with an index value (com) greater than 2. An index value around 1 is optimal as it indicates that the variable is not cross loading, and rather is

Table 4.3: Factor Solution. Insignificant variables removed

Variable	ML2	ML4	ML5	ML1	ML6	ML3	h2	u2	com
PRE_ATT_SC_1	**0.76**	-0.11	0.02	0.06	0.08	-0.04	0.58	0.42	1.08
PRE_ATT_SC_2	**0.47**	**0.38**	0.04	-0.01	-0.10	0.04	0.46	0.54	2.05
PRE_ATT_SC_3	**0.34**	**0.43**	0.02	-0.12	-0.03	0.00	0.35	0.65	2.08
PRE_ATT_SC_4	**0.77**	-0.05	-0.01	-0.01	0.07	-0.01	0.61	0.39	1.02
PRE_ATT_SC_5	**0.77**	-0.11	-0.05	0.00	0.09	0.01	0.61	0.39	1.08
PRE_ATT_DL_1	0.03	0.02	0.00	0.06	**0.70**	-0.02	0.57	0.43	1.02
PRE_ATT_DL_2	0.10	**0.50**	0.01	0.06	0.10	0.00	0.39	0.61	1.18
PRE_ATT_DL_3	-0.15	**0.58**	-0.06	-0.15	0.26	0.01	0.42	0.58	1.71
PRE_ATT_DL_4	0.08	0.03	0.01	0.01	**0.69**	0.01	0.53	0.47	1.03
PRE_ATT_DL_5	0.13	-0.06	-0.06	0.06	**0.65**	0.03	0.56	0.44	1.14
PRE_LK1	-0.01	0.14	-0.03	**0.82**	-0.02	0.02	0.78	0.22	1.07
PRE_LK2	-0.10	0.17	0.00	**0.74**	0.09	0.02	0.70	0.30	1.18
PRE_LK5	0.08	-0.11	0.00	**0.57**	0.05	-0.02	0.35	0.65	1.13
PRE_DAILYM	-0.08	**0.59**	0.07	0.06	0.00	-0.08	0.30	0.70	1.12
PRE_DAILYG	0.01	**0.55**	-0.07	0.04	-0.14	0.01	0.30	0.70	1.18
PRE_FREQEN	-0.08	**0.53**	0.00	0.05	-0.01	0.03	0.27	0.73	1.07
PRE_ANX.1_1	-0.15	0.01	**0.69**	-0.11	0.13	0.04	0.53	0.47	1.23
PRE_ANX.1_2	-0.03	0.03	**0.82**	0.02	0.07	-0.03	0.60	0.40	1.02
PRE_ANX.1_3	0.12	-0.02	**0.51**	0.00	-0.12	0.01	0.32	0.68	1.22
PRE_ANX.1_4	0.01	-0.02	**0.60**	0.05	-0.09	0.01	0.40	0.60	1.06
PRE_WHYCS_3	0.00	0.06	-0.07	0.03	-0.03	**0.48**	0.27	0.73	1.09
PRE_WHYCS_5	0.00	-0.06	0.04	-0.05	0.03	**0.80**	0.61	0.39	1.03
PRE_WHYCS_6	-0.02	-0.05	0.05	0.02	0.03	**0.70**	0.47	0.53	1.03

explained (mostly) by one factor. Confounding variables would complicate our interpretation of the factors, and thus, are also removed[5]. We are left with Table 4.4.

Lastly, we remove variables with loadings below a certain threshold. The loading threshold is often set to between 0.2 to 0.7 and is dependent on a variety of factors including the chosen variables, sample size and strength of correlations. There is no true consensus as to which threshold is best, but a justification should be provided [46]. While we initially chose a threshold of 0.35, we found that 0.4 produced a more robust factor solution during

[5]The order in which we remove variables also can change how variables load onto factors, for example, whether we remove confounding variables first or after variables that do not load. As a result, the end results may differ. In our case, however, both directions have been tested and result in the same matrix.

Table 4.4: Factor Solution. Confounding variables removed

Variable	ML1	ML5	ML6	ML2	ML4	ML3	h2	u2	com
PRE_ATT_SC_1	0.03	0.02	0.05	**0.74**	-0.02	-0.02	0.58	0.42	1.02
PRE_ATT_SC_4	-0.05	-0.01	0.03	**0.77**	0.06	0.00	0.62	0.38	1.02
PRE_ATT_SC_5	-0.05	-0.05	0.06	**0.75**	0.00	0.03	0.61	0.39	1.03
PRE_ATT_DL_1	0.05	0.01	**0.77**	0.00	-0.07	-0.03	0.58	0.42	1.03
PRE_ATT_DL_2	0.08	0.00	0.15	0.09	**0.34**	0.02	0.28	0.72	1.69
PRE_ATT_DL_3	-0.12	-0.07	0.30	-0.10	**0.42**	0.02	0.32	0.68	2.23
PRE_ATT_DL_4	0.00	0.01	**0.72**	0.06	-0.02	-0.02	0.53	0.47	1.02
PRE_ATT_DL_5	0.03	-0.06	**0.69**	0.11	-0.10	0.02	0.56	0.44	1.11
PRE_CF_MEAN	**0.35**	-0.19	0.10	-0.04	0.13	0.10	0.42	0.58	2.34
PRE_LK1	**0.90**	-0.02	-0.06	-0.03	0.07	0.00	0.80	0.20	1.02
PRE_LK2	**0.79**	0.00	0.07	-0.12	0.08	-0.01	0.69	0.31	1.08
PRE_LK5	**0.60**	0.00	0.02	0.05	-0.12	-0.04	0.33	0.67	1.11
PRE_DAILYM	0.00	0.09	-0.03	-0.02	**0.73**	-0.11	0.46	0.54	1.08
PRE_DAILYG	-0.01	-0.07	-0.20	0.08	**0.67**	0.03	0.41	0.59	1.23
PRE_FREQEN	0.01	0.01	-0.04	-0.01	**0.58**	0.04	0.33	0.67	1.02
PRE_ANX.1_1	-0.12	**0.68**	0.12	-0.15	0.02	0.05	0.52	0.48	1.25
PRE_ANX.1_2	0.02	**0.84**	0.09	-0.04	0.02	-0.03	0.62	0.38	1.03
PRE_ANX.1_3	-0.01	**0.50**	-0.13	0.11	0.00	0.03	0.32	0.68	1.25
PRE_ANX.1_4	0.05	**0.59**	-0.09	0.00	-0.03	0.02	0.39	0.61	1.07
PRE_WHYCS_1	0.03	0.12	0.00	0.03	0.12	**-0.38**	0.17	0.83	1.46
PRE_WHYCS_3	0.00	-0.09	-0.07	0.02	0.07	**0.55**	0.34	0.66	1.11
PRE_WHYCS_4	0.05	0.10	0.05	0.02	-0.01	**0.32**	0.11	0.89	1.34
PRE_WHYCS_5	-0.08	0.05	-0.04	0.02	0.04	**0.76**	0.55	0.45	1.04
PRE_WHYCS_6	-0.01	0.06	-0.02	-0.01	0.03	**0.67**	0.44	0.56	1.03

our confirmatory factor analysis as we shall discuss in the following chapter. Removed variables with loadings below 0.4 are ATT_DL_2, ATT_DL_3, WHY_CS_1, WHY_CS_4, and CF_MEAN. Our final factor solution is shown in Table 4.5

Below the loadings we include values for 'SS loadings' or sums of square loadings. These are the eigenvalues associated with each factor and indicate how much variance each factor explains in the data. The following table is a correlation matrix, which indicates how each factor is correlated with the other factors. As we made our factors oblique it is not unusual to see correlations between our factors.

Table 4.5: Factor Solution. Final solution

Variable	ML2	ML5	ML1	ML6	ML3	ML4	h2	u2	com
PRE_ATT_SC_1	**0.75**	0.04	0.04	0.03	-0.03	-0.01	0.58	0.42	1.02
PRE_ATT_SC_4	**0.79**	0.01	-0.03	0.01	0.01	0.04	0.63	0.37	1.01
PRE_ATT_SC_5	**0.75**	-0.03	-0.04	0.06	0.02	0.00	0.61	0.39	1.02
PRE_ANX.1_1	-0.14	**0.69**	-0.11	0.13	0.04	0.03	0.53	0.47	1.22
PRE_ANX.1_2	-0.02	**0.82**	0.02	0.07	-0.03	0.04	0.60	0.40	1.03
PRE_ANX.1_3	0.13	**0.51**	0.01	-0.13	0.02	-0.01	0.32	0.68	1.28
PRE_ANX.1_4	0.02	**0.60**	0.07	-0.10	0.01	-0.03	0.40	0.60	1.09
PRE_LK1	-0.01	-0.02	**0.87**	-0.03	0.02	0.06	0.78	0.22	1.01
PRE_LK2	-0.11	0.00	**0.78**	0.08	0.01	0.08	0.70	0.30	1.08
PRE_LK5	0.06	0.01	**0.57**	0.02	-0.03	-0.08	0.33	0.67	1.07
PRE_ATT_DL_1	-0.02	-0.01	0.04	**0.74**	-0.02	0.00	0.57	0.43	1.01
PRE_ATT_DL_4	0.05	0.00	0.00	**0.68**	0.01	0.03	0.52	0.48	1.02
PRE_ATT_DL_5	0.09	-0.07	0.02	**0.67**	0.03	-0.04	0.57	0.43	1.07
PRE_WHYCS_3	0.01	-0.08	0.04	-0.03	**0.48**	0.01	0.26	0.74	1.08
PRE_WHYCS_5	0.00	0.05	-0.07	0.03	**0.80**	-0.01	0.62	0.38	1.02
PRE_WHYCS_6	-0.02	0.06	0.00	0.03	**0.69**	-0.02	0.47	0.53	1.02
PRE_DAILYM	-0.04	0.09	-0.04	0.07	-0.09	**0.80**	0.59	0.41	1.08
PRE_DAILYG	0.07	-0.07	-0.01	-0.10	0.03	**0.61**	0.39	0.61	1.12
PRE_FREQEN	-0.01	0.00	0.04	0.02	0.04	**0.50**	0.28	0.72	1.03
SS loadings	1.86	1.83	1.78	1.64	1.36	1.28			
ML2	1.00	-0.34	0.36	0.42	0.05	0.21			
ML5	-0.34	1.00	-0.59	-0.58	-0.20	-0.31			
ML1	0.36	-0.59	1.00	0.61	0.20	0.35			
ML6	0.42	-0.58	0.61	1.00	0.11	0.28			
ML3	0.05	-0.20	0.20	0.11	1.00	0.17			
ML4	0.21	-0.31	0.35	0.28	0.17	1.00			

Interpreting the Factor Solution

Now, with our finalized loadings, we can interpret what our factors (with their corresponding variables) describe.

Factor 2 (ML2) contain variables that measure how hard/confusing/stressful using math is when solving problems in school. We simply call this factor 'School Math Affect.' Factor 6 (ML6) contains similar measures, but in the context of daily life. We call this factor 'Daily Life Math Affect.' Notably, there is a clear distinction between how students feel about using math to solve problems in different contexts.

Factor 5 (ML5) contain variables that measure a student's level of anxiety in a variety of quantitative situations. Because these are quantitative situations and not purely mathematical situations, we generalize the more common term of math anxiety and instead, call factor 5, 'Numerical Anxiety.' Note that questions asking about stress (SC_5, DL_5) does not load onto anxiety. While stress and anxiety may produce similar symptoms or triggered by similar situations, there is a nuanced distinction between anxiety and stress. Anxiety is internally triggered with lasting effects while (acute) stress is externally caused with short-lived effects.

Factor 1 (ML1) contains variables that measure satisfaction in one's quantitative ability and level of confidence in using numbers in school and daily life. Levels of confidence correspond to a measurement of self-efficacy, and the level of satisfaction corresponds to one's sense of ability. The more confident or satisfied one is in their quantitative ability, the more effective they will find themselves in solving numerical problems. Thus, we call Factor 1 'Numerical

Self-Efficacy.'

Factor 3 (ML3) contains binary variables that measure why students chose their course, in particular, if the course 'sounded interesting', if they 'heard the class was good', and/or if they 'heard the instructor was good.' It seems to measure a level of intrinsic motivation for taking the course, rather than an extrinsic motivation which would be indicated by choosing 'for a prerequisite' as their reason. Therefore we call this factor '(intrinsic) Course Motivation'.

Factor 4 (ML4) contain variables that measure how frequently a student believes that they encounter or solve quantitative problems in daily life. The variables indicate to what extent the student recognizes the applications or relevance of quantitative literacy in day-to-day circumstances. Thus, we call factor 5, 'Numerical Relevancy.'

All our factors describe or influence one's attitude or affect. We will refer to these as our composite variables in 5.

With these factors identified, the correlation matrix of our factors in Table 4.5 is important to address. Again, as our factors are oblique, it is not unusual to see that they are correlated to each other. Notably, Numerical Anxiety, negatively correlates with all other factors rather strongly. While we cannot assess the direction of the correlation, either Numerical Anxiety can have a wide impact across multiple factors or multiple factors can all contribute to Numerical Anxiety. School/Daily Life Math Affect and Numerical Self Efficacy are also strongly correlated; a student with a high self efficacy is likely to have a higher School/Daily Life Math Affect (or vice versa). Numerical Relevancy is mildly correlated with other factors, though most strongly with self-efficacy.

Course Motivation is the least correlated with the factors.

Some factors are correlated to other factors in similar ways, however, they all have different magnitudes/degrees of correlation. There are no instances of extreme correlation or issues of multicollinearity. Therefore, each factor can uniquely explain portions of our data.

4.1.5 Factor Scores

Now that we have our final factor solution, we can generate factor scores for each student. A factor score is a numerical value that describes the factor as a combination of factor loadings weighted by student responses of the associated variable. Because our factors are oblique i.e. not independent, we must use the Regression method to compute scores [45]. Thankfully, this is the most straightforward method: the factor score is simply the linear combination of its variables weighted by its loading. For example, the factor score for School Math Affect is,

$$\text{School Math Affect} = l_1 * x_1 + l_2 * x_2 + l_3 * x_3 \tag{4.1}$$

x_n, is the Likert value of a student's math affect in school (easy/hard, straightforward/confusing, not stressful/stressful, respectively) and l_i is the loading corresponding to the question/variable.

Factor scores for each student are calculated by integrating a separate scoring script in the data flow (Figure 3.1). After the raw data is initially scored by the main scoring script, factor scores are calculated and standardized by a factor scoring script in Python. Once all the raw data has been scored to include the new factors, the data can be re-merged into a master data set.

4.2 Confirmatory Factor Analysis

By performing exploratory factor analysis, we are able to gauge what underlying variables or factors may be present in the data. However, the factor solution can change based on the data set. The results of exploratory factor analysis (EFA) in the previous section only describes the specific dataset we used, S1. Our factor solution is more of a hypothesis of how our variables are structured rather than a definitive conclusion over all and future data. Therefore, if we seek to apply our factor solution to newly acquired data, we must test if our hypothesis is true. This is done through confirmatory factor analysis (CFA), a complementary procedure to EFA, on a separate data set, S2.

Instead of determining how many factors explain our data as it is done in EFA, we predetermine the structure of our data from our EFA results and apply it to a new data set. Essentially, we are testing how well our hypothesised model fits new data in CFA.

CFA is done in R. We pass our model and data in as parameters and the function returns statistical measurements on the quality of fit. The results of these metrics are listed in Table 4.6. There are four primary measures of interest: χ^2 p-value, RMSEA, CFI/TFI, and SRMR.

The χ^2 (chi-squared) p-value is a value that indicates the probability of two groups being identical to each other[6]. In the case of CFA, we are comparing the probability that our predicted model and actual structure of the data are the same. Ideally, we seek a high χ^2 p-value, which indicates that our model perfectly matches the structure of a new data set. Our χ^2 p-value is not at all high ($\chi^2 < 0.0001$). Luckily, this is not concerning. The χ^2 test statistic–the measure that determines the p-value–is sensitive to sample size; the larger the sample, the larger the χ^2 test statistic, which gives a small χ^2 p-value. A low χ^2 p-value is common for large samples, and as a result, is not a sufficient metric of fit. Thus, it is recommended that we also consider/meet two other fit metrics [51][52] [46].

RMSEA, or root-mean square error approximation, is an alternative measure that adjusts the χ^2 test statistic for sample size. Note that RMSEA is not a measure of probability, like the χ^2 p-value, but more like an adjusted χ^2 test statistic. Thus in this case, we desire a RMSEA value close to 0 (which is associated with a large p-value). A value less than 0.06 is recommended for good fit, but 0.08 is also acceptable [51][52]. Our RMSEA value is below 0.06 (RMSEA = 0.48), indicating that our model fits well.

SRMR or standardized root-mean square residuals is the difference in the residuals of between our predicted model and the actual model. A value less than 0.08 is recommended, and is considered a good predictive model [51][52]. Our model is below this threshold (SRMR = 0.035), therefore, indicates that our predicted and actual model share similar residual patterns.

The CFI or TFI index is an incremental fit index that compares the fit of

[6]Our null hypothesis is that there is no difference between our tested model and the actual model

Table 4.6: CFA Model Metrics

Metric	Value	Recommended Value (for a good fit)
χ^2 p-value	< 0.0001	> 0.05
RMSEA	**0.48 ± 0.03**	< 0.06
SRMR	**0.035**	< 0.08
CFI	**0.951**	> 0.90
TFI	**0.938**	> 0.90

our proposed model with a baseline model. The baseline or null model is a model that assumes all variables to be uncorrelated. Essentially, the index is a ratio of the difference between the χ^2 test statistic of the null model and test model, and the χ^2 test statistic of the null model, adjusting for sample size. An index close to 1 is ideal and an index greater than 0.90 is acceptable[51][52]. Our CFI and TFI values are both greater than 0.90 (CFI = 0.951, TFI = 0.938), which again indicates a good model fit.

Our model meets all the metrics for a good model fit, with the exception of the χ^2 p-value. Thus, we can conclude that our EFA model of factors and their associated variables is a consistent structure that underlies our data. We can assume that our factors are relatively robust, and viable for future analysis. We will adopt this solution for all future QuaRCS datasets.

Chapter 5

Modeling

With both exploratory and confirmatory factor analysis completed, we utilize the factors or composite variables that emerged to proceed with our primary investigation. Can our composite variables explain the differences in performance on the QuaRCS assessment among various demographic groups? Our initial hypotheses laid out in Section 2.3.2 accounted only for the effect of Numerical Self-Efficacy, Numerical Anxiety, and Numerical Relevancy on explaining the relationship between demographics and quantitative literacy. However, there are additional variables such as ‘School Math Affect’, ‘Daily Life Math Affect’, and ‘Course Motivation’ that unexpectedly emerged from our factor analysis. As variables that are related to our overall constructs of affect and attitude, we include these new variables in our investigation. Figure 5.1 illustrates the hypothesized relationship between demographics and score. We hypothesize that the demonstrated relationships between demographics and score will be partially mediated by our composite variables. Our compos-

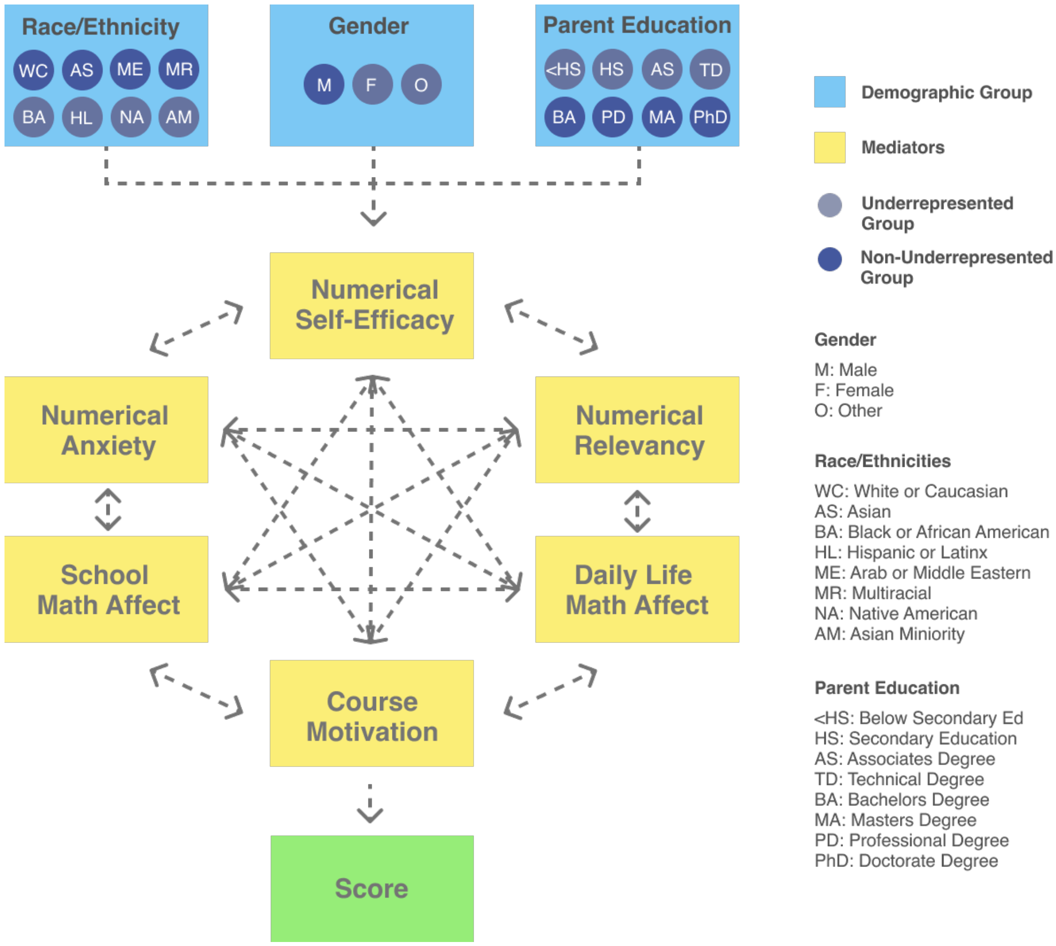


Figure 5.1: Proposed relationship of variables. Blue boxes indicate our broad demographic groups. Each contain specific sub-groups that we combine into UR and Non-UR groups, indicated by gray and navy circles. The composite variables that we suspect to mediate the relationship between UR demographic groups and score are indicated by the yellow boxes. Suspected relationships between variables are indicated by a dashed line.

ite variables also have complex interactions with each other, as demonstrated in Section 4.1.4.

Recall that we grouped specific demographics into either underrepresented (UR) or non-underrepresented (Non-UR) categories, as listed in Table 3.4 to simplify our analysis. We will investigate each variable independently, but must recognize that these variables are correlated with one another. Recall in Section 4 we expected our composite variables to be correlated to one another, and intentionally created oblique factors, which are not independent of each other.

First, we will give a general overview of distributions of variables constructed from factor analysis using QuaRCS Lt.2 assessments[1]. Then, we will analyze disparities between underrepresented and non-underrepresented groups. These results will guide our subsequent development of linear regression models and mediation analysis, which we use to examine the extent to which our composite variables account for the differences in score between UR and non-UR groups[2].

5.1 Exploratory Plots

General Sample

The overlapping histograms in Figure 5.2 show the distributions of the composite variables for our entire sample. Medians are indicated by the dashed line; means are indicated by the solid line (at zero because these are standardized values). A table quantifying these values can be found in Table B.6 in the Appendix.

First, we note that most variables are not normally distributed[3]. This

[1] or data set S3 from Table 3.3

[2] As a general note, we fit many models in this chapter, but adjustments for multiple comparisons were not made

[3] While medians are the proper statistic to report in such cases because they are a more

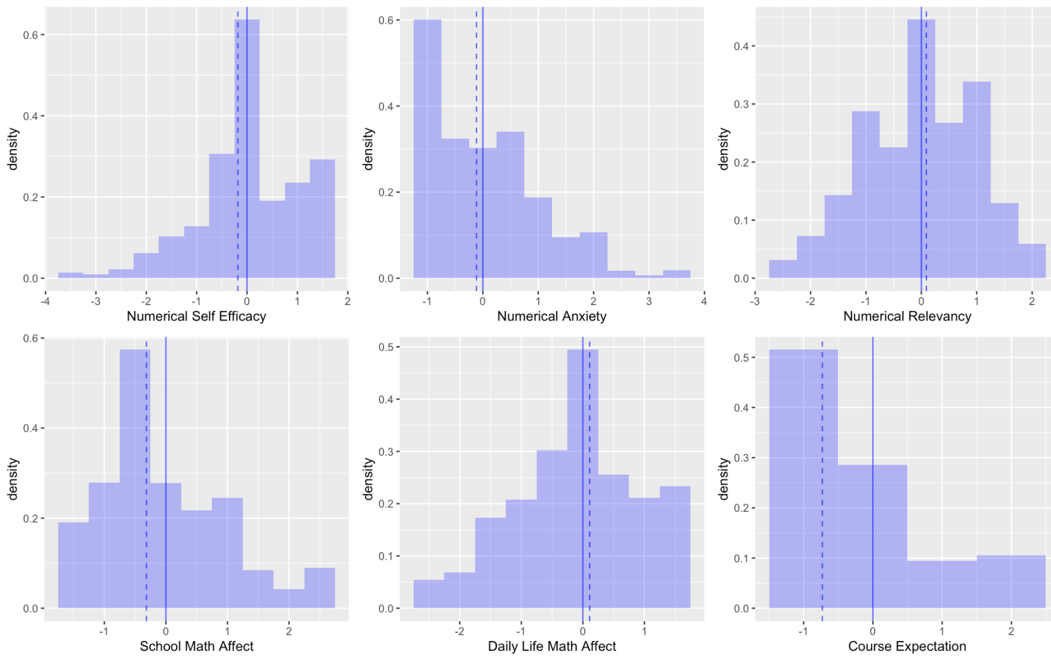


Figure 5.2: Distributions of composite variables. Medians are marked by a dashed line; means are marked with a solid line.

provides a caveat to subsequent analyses that 'require' normally distributed samples, for instance regression analysis and t-tests. However, we will proceed with caution[4] and keep this limitation in mind. Second, we note that the composite variables are not entirely continuous as the questions from which these variables are calculated are ordinal or binary and can only take a range of integer values; there are a finite number of possible values for each variable (with the exception of Course Motivation which can only take 8 discrete values). Nonetheless, we treat the variables as continuous because the number of

reliable measure of central tendency, means are also included as we will utilize them in later analyses.

[4]'Proceeding with caution', is not an unusual decision to make in statistical analyses, where it is quite common to encounter non-normal data

possible values is sufficient[5].

Generally, a considerable peak for each variable exists in the distribution of responses. It indicates the most common combination of responses to the non-quantitative questions from which the composite variable is calculated. More normal distributions are preferred, as one common response lessens our ability to distinguish differences between students.

However, there are distinct, although skewed, distributions about these peaks. Numerical Self-Efficacy appears bimodal, or two peaked, with a long tail towards negative self-efficacy. Similarly, Numerical Anxiety has a tail towards higher levels of anxiety. School Math Affect, Daily Life Math Affect, and Numerical Relevancy have comparatively more symmetrical about the mean. Factor scores for Course Expectations can take on only a narrow range of values because its derived from a small number of binary variables, illustrated by the large bin sizes. It is heavily skewed to the left, suggesting that most students are not intrinsically motivated to take their particular gen-ed course, but rather are taking it as prerequisite or required for their major.

By Underrepresented Groups

Next, we break down the results of score, effort, and our composite variables by each group defined in Table 3.4 to understand qualitatively how the underrepresented (UR) groups compare with non-underrepresented groups (Non-UR). Overlapping histograms of variables for UR and Non-UR groups are presented

[5] Ordinal values that range from 1-5 are often considered continuous in statistical analyses. We have more than 5 possible values, so it would not be unreasonable to treat our variables as continuous

in Figures 5.3 and 5.4. As students' parents' education is highly correlated with socioeconomic status, we will refer to the group of students who report having parents with postsecondary education as the "high SES" group and students who report having parents without post secondary education as the "low SES" group[6].

Means are identified by vertical lines. To determine whether means between groups are significantly different[7], we performed a two-sample t-test[8]. Dashed lines indicate that means between UR and Non-UR groups are not statistically different, while solid lines indicate that they are statistically different. Medians, means, and p-values for the t-tests can be found in Tables B.7 through B.9 of the Appendix.

In all cases, we see that the distribution of effort is similar between groups, yet there is a significant difference in the means of score between UR and Non-UR groups, consistent with and reflective of the achievement gap in quantitative literacy. UR gender and race/ethnicity groups score 7% and 12% below their Non-UR counterparts. Low SES students score 13% below High SES students. These results are consistent with the existing literature, as discussed in Section 2.2.

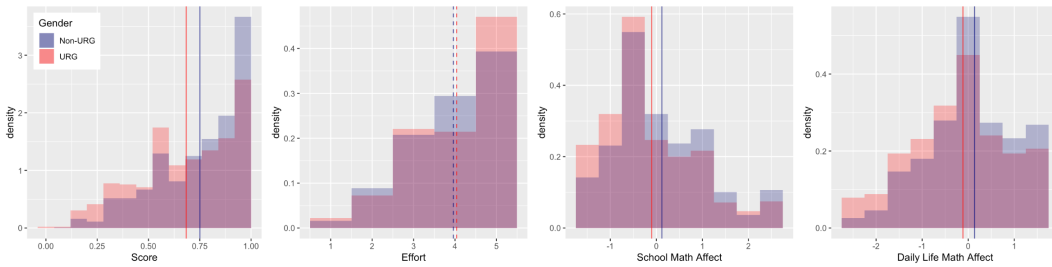
The means for the composite variables are also consistently lower for underrepresented groups, with the exception of Numerical Anxiety where it is

[6]alternatively, we can think of these two groups as first generation and non-first generation students

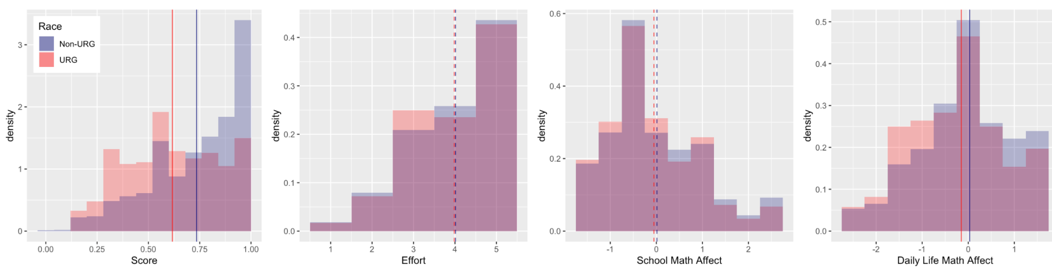
[7]significance in the context of statistics indicates whether or not a statistic or result is likely to occur by chance, if the null hypothesis is true. In other words, the statistic or value is meaningful and not arbitrary. There are different metrics to define significance. In this thesis, if a p-value is <0.01, then we deem a statistic to be significant

[8]While two-sample t-tests assume normally distributed samples, it is fairly robust measure even when this condition is not met [53]

A. Gender



B. Race



C. Socioeconomic Status

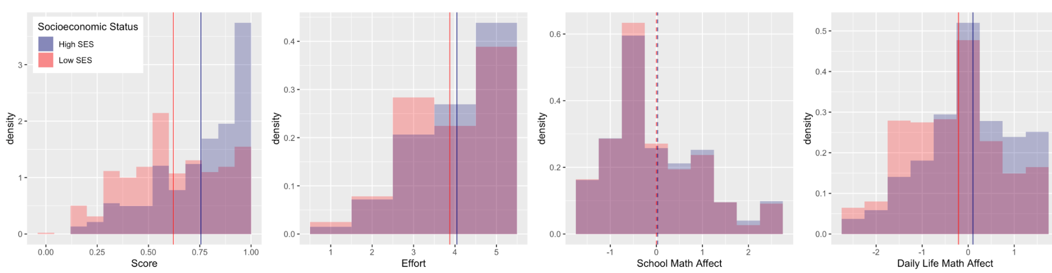


Figure 5.3: Distributions of Score, Effort, School Math Affect, Daily Life Math Affect by Gender (top), Race (middle), and SES (bottom). UR groups are in red, Non-UR groups are in blue. Vertical lines identify the mean of each distribution. Vertical solid lines indicate that means are significantly different between UR and Non-UR distributions. Dashed lines indicate that means are not significantly different.

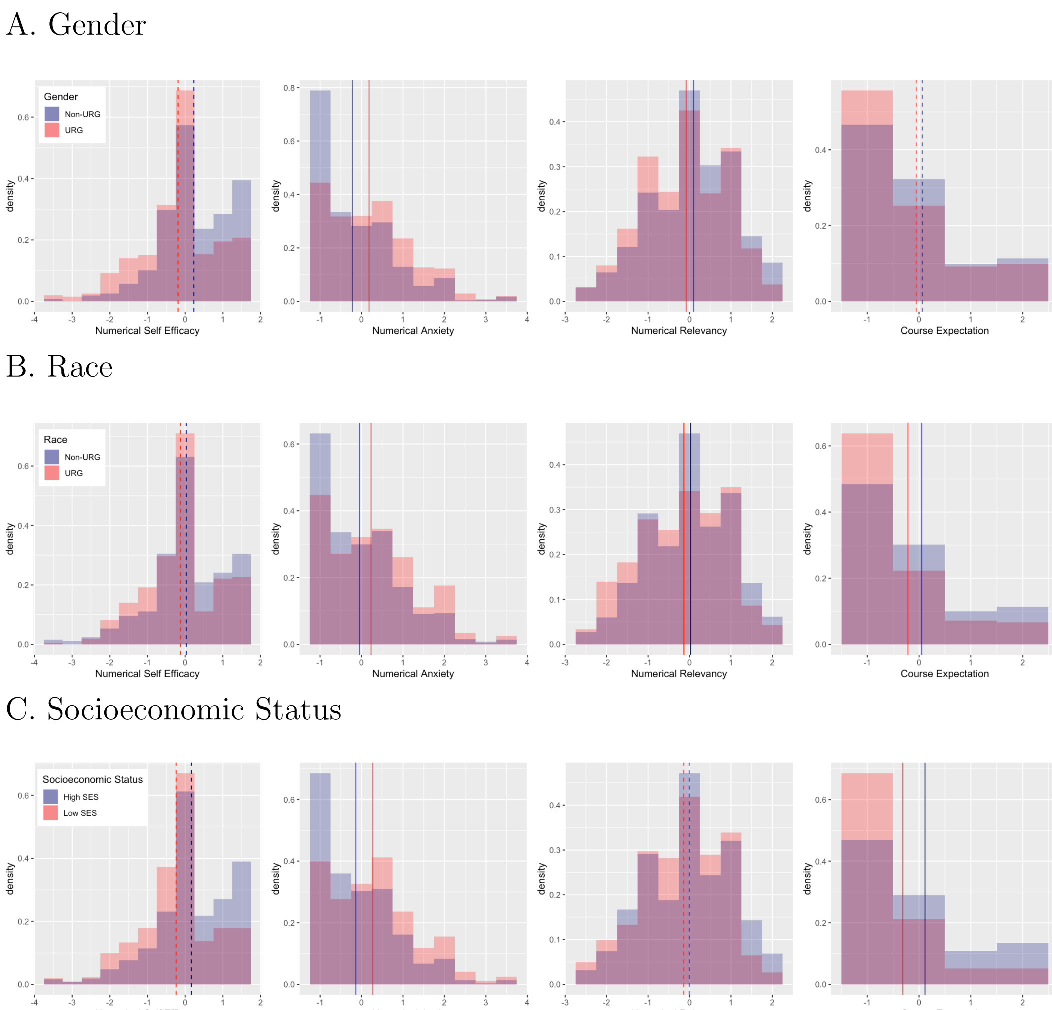


Figure 5.4: Distributions by Gender, Race, and SES: Numerical Self-Efficacy, Numerical Anxiety, and Numerical Relevancy. UR groups are in red, Non-UR groups are in blue. Vertical lines identify the mean of each distribution. Vertical solid lines indicate that means are significantly different between UR and Non-UR distributions. Dashed lines indicate that means are not significantly different.

consistently higher. The differences in means between UR Gender and Non-UR gender are highly significant for School Math Affect, Numerical Anxiety, Numerical Self-Efficacy, Daily Life Math Affect, and Numerical Relevancy. Means between UR and Non-UR Race are significantly different for Numerical Anxiety, Numerical Self-Efficacy, Daily Life Math Affect, Numerical Relevancy and Course Motivation. Means between Low SES and High SES are highly significant for Numerical Anxiety, Numerical Self-Efficacy, Daily Life Math Affect, and Course Motivation. The significant differences suggest that factors that influence attitude and affect contribute to UR groups' lower QuaRCS scores.

5.2 Linear Regression Models

The clear differences in score and composite variables between UR and Non-UR groups raise the question of whether these variables might explain differences in score. Moreover, as our variables are related to one's attitude and affect, we can expect these variables to provide some explanation to differences as previously discussed in Section 2.2.

We explore this question using a variety of linear regression models[9]. First we investigate whether our composite variables have a relationship with score, and whether this relationship differs between UR and Non-UR groups, by creating linear regression models of the form:

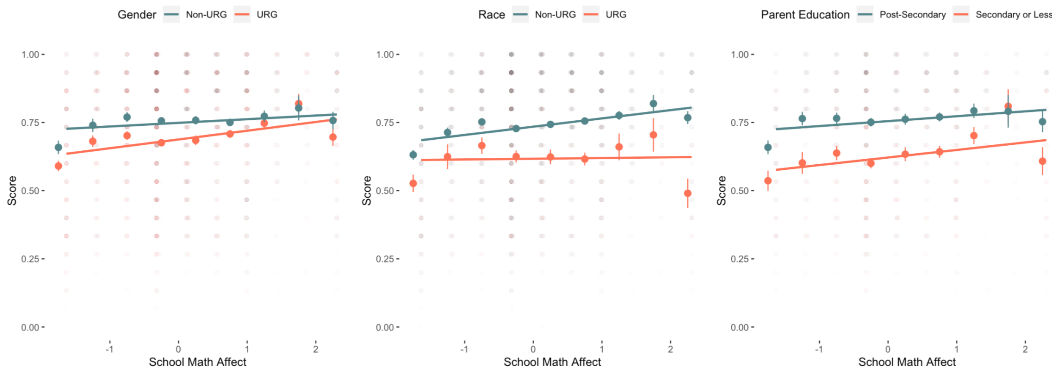
[9]Linearity of relationships was confirmed by applying a smooth line in R which fits a flexible function to the data

$$Score \sim \beta_0 + \beta_1 Z + \beta_2 X + \beta_3 Z * X \tag{5.1}$$

X is a binary variable that indicates the UR status of a demographic group (Gender, Race or SES), Z is a composite variable, and X*Z is the interaction term between them. The interaction term gives us insight into how the composite variable may behave differently depending on UR status. Quantitatively, it changes the steepness of the linear regression line depending on the UR group.

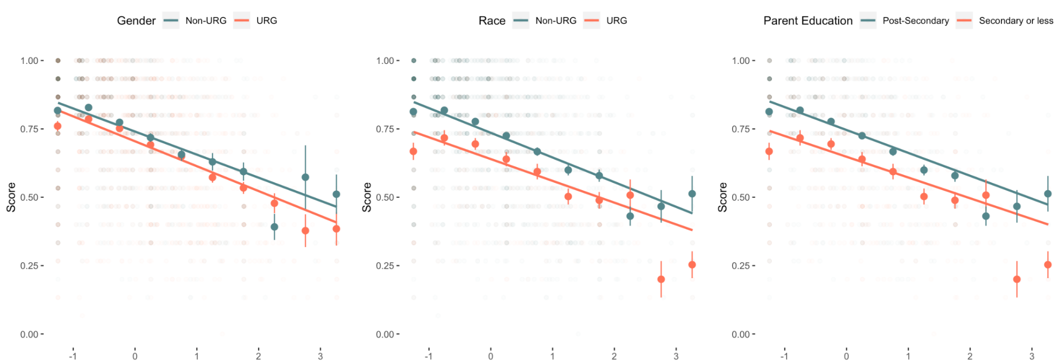
Statistical significance of the β coefficients indicate whether a variable is predictive of score. If β_1 is significant, then our composite variable is predictive of score. If β_2 is significant, then the demographic group is predictive of score. If β_3 is significant, the composite variable behaves differently when predicting score for different demographic groups.

Figures 5.5 to 5.10 provide a visualization of the data and linear regression models. They include tables that give regression coefficients and p-values. Regression lines for each level (UR and Non-UR) are overlaid with binned means with error bars. Binned means provide a more simple visualization and illustrate the trends in the data more easily, compared to a scatter plot. Note that the regression lines are derived from the entire data set, not the binned means.



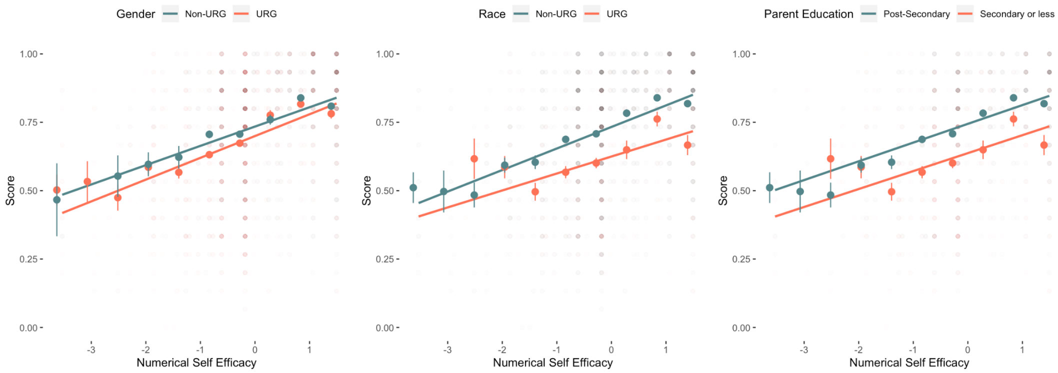
URG	β_0	p-value	β_1	p-value	β_2	p-value	β_3	p-value	R^2
Gender	0.749	<0.001	0.013	0.047	-0.061	<0.001	0.019	0.038	0.034
Race	0.735	<0.001	0.030	<0.001	-0.118	<0.001	-0.028	0.022	0.057
Education (SES)	0.755	<0.001	0.018	0.020	-0.133	<0.001	0.010	0.432	0.094

Figure 5.5: Score vs. School Math Affect



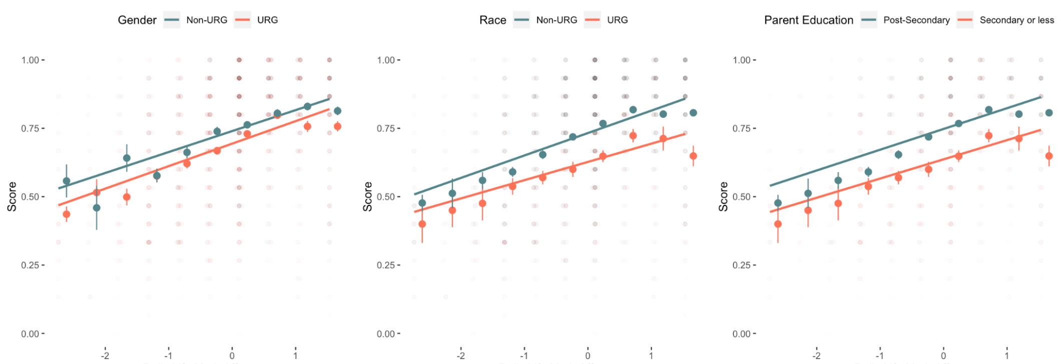
URG	β_0	p-value	β_1	p-value	β_2	p-value	β_3	p-value	R^2
Gender	0.740	<0.001	-0.084	<0.001	-0.036	<0.001	-0.007	0.438	0.179
Race	0.736	<0.001	-0.090	<0.001	-0.096	<0.001	0.011	0.315	0.199
Education (SES)	0.746	<0.001	-0.084	<0.001	-0.098	<0.001	0.008	0.498	0.206

Figure 5.6: Score vs. Numerical Anxiety



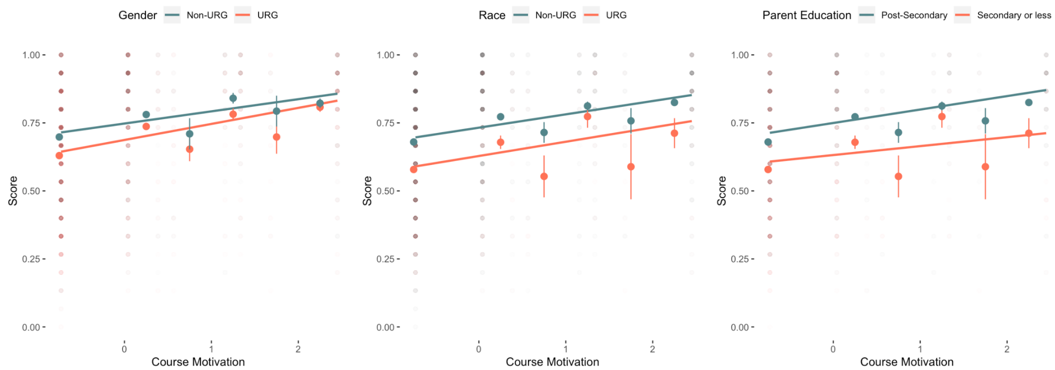
URG	β_0	p-value	β_1	p-value	β_2	p-value	β_3	p-value	R^2
Gender	0.734	<0.001	0.071	<0.001	-0.035	<0.001	0.009	0.298	0.133
Race	0.733	<0.001	0.079	<0.001	-0.108	<0.001	-0.017	0.148	0.159
Education (SES)	0.744	<0.001	0.069	<0.001	-0.107	<0.001	-0.003	0.805	0.175

Figure 5.7: Score vs. Numerical Self-Efficacy



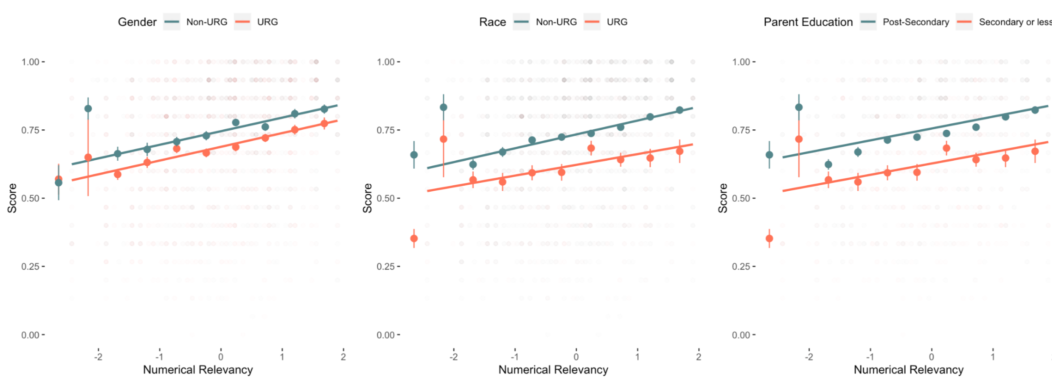
URG	β_0	p-value	β_1	p-value	β_2	p-value	β_3	p-value	R^2
Gender	0.740	<0.001	0.077	<0.001	-0.046	<0.001	0.006	0.512	0.149
Race	0.733	<0.001	0.082	<0.001	-0.106	<0.001	-0.015	0.160	0.168
Education (SES)	0.748	<0.001	0.076	<0.001	-0.111	<0.001	-0.006	0.629	0.187

Figure 5.8: Score vs. Daily Life Math Affect



URG	β_0	p-value	β_1	p-value	β_2	p-value	β_3	p-value	R^2
Gender	0.748	<0.001	0.045	<0.001	-0.061	<0.001	0.014	0.103	0.077
Race	0.732	<0.001	0.049	<0.001	-0.104	<0.001	0.003	0.791	0.090
Education (SES)	0.750	<0.001	0.049	<0.001	-0.118	<0.001	-0.017	0.224	0.123

Figure 5.9: Score vs. Course Motivation



URG	β_0	p-value	β_1	p-value	β_2	p-value	β_3	p-value	R^2
Gender	0.746	<0.001	0.050	<0.001	-0.057	<0.001	0.001	0.947	0.071
Race	0.734	<0.001	0.051	<0.001	-0.112	<0.001	-0.011	0.313	0.089
Education (SES)	0.756	<0.001	0.044	<0.001	-0.128	<0.001	-0.002	0.864	0.122

Figure 5.10: Score vs. Numerical Relevancy

Almost all β_1 and β_2 coefficients are statistically significant. In other words, our composite variables are predictive of score for both UR and non-UR groups, with the exception of School Math Affect for gender and SES groups; and a student's status as an underrepresented group is predictive of a lower score compared to a non-underrepresented group. Qualitatively, these relationships are illustrated by the magnitude and direction of the slope and the offset of the regression lines between UR and non-UR. For example, regression lines for School Math Affect and Numerical Self-Efficacy generally have positive slopes, which indicate that higher self-efficacy or School Math Affect correlates to higher score. However, the slopes of School Math Affect are shallower than the slopes of Numerical Self-Efficacy. This suggests that Numerical Self-Efficacy is a stronger predictor of score than School Math Affect. We also note the offset of the regression lines between UR and Non-UR groups. Regression lines for UR groups consistently lie below Non-UR groups. The width of this offset is related to the differences in means of score between groups, a reflection of the achievement gap, as demonstrated in the previous section.

On the other hand, all β_3 coefficients are not statistically significant. β_3 corresponds to the interaction term between a demographic and a composite variable, which adjusts the slope of the regression line depending on the group. Is UR status 'a moderator' of the relationship between a composite variable and score? In other words, does the effect of a composite variable on score vary according to whether a student is an underrepresented minority? The insignificance of this term demonstrates that the effect of a composite variable on score between UR and non-UR groups does not differ. We only see that

UR groups have, on average, higher Numerical Anxiety, lower self-efficacy, etc.

R^2, a measure of how well our regression line explains the variance in score, ranges from 3% to 20%. School Math Affect, Course Motivation, and Numerical Relevancy individually explain less than ~10% of the variance, while Numerical Anxiety, Numerical Self-Efficacy, and Daily Life Math Affect explain more than 15% each. Qualitatively, a lower R^2 is associated with a more dispersed distribution of data about the regression line; a higher R^2 is associated with a more concentrated distribution. While regressions with higher R^2 values are ideal, all our beta coefficients prove to be significant. Therefore, our linear regression models may not be *precise*, but are still *accurate* and demonstrate that there are relationships between score, the composite variables, and demographics. In particular, Numerical Anxiety, Numerical Self-Efficacy, and Daily Life Math Affect, comparatively, have the strongest relationships with score.

To summarize, we investigated the relationship between the composite variables and score by UR and Non-UR groups using regression. The results of the regression models show that our composite variables, which are factors that influence students' attitudes and affect, are able to predict 3% to 20% of the variance in score. Therefore, we can continue to investigate whether differences in the composite variables we observed in the previous section *explains* the demographic differences in score. Because we also observed that the ability of each variable in predicting score (the slope of the regression lines) is the same between UR and Non-UR groups, we do not need to consider differing effects of our variables on score for each group.

5.3 Mediation Analysis

Now that we have observed how the composite variables predict score, can they explain the relation between underrepresented groups and their lower scores in quantitative literacy? To investigate this crucial aspect of our investigation, we utilize a statistical procedure called Mediation Analysis.

5.3.1 Overview and Procedure of Mediation Analysis

When there is a statistically significant relationship between an independent and dependent variable that does not seem to be warranted, one might suspect that there is another variable driving this relationship. If this other variable explains all or some of the relationship between the initial independent and dependent variable, we say that the variable is a "mediator" [54][55].

Mediation analysis is a process involving a series of linear regression models from which we can determine whether a given variable mediates a relationship between two others (Figure 5.11). Three linear regression models are utilized:

$$Y \sim \beta_1 X + \beta_a \tag{5.2}$$

$$M \sim \beta_2 X + \beta_b \tag{5.3}$$

$$Y \sim \beta_3 M + \beta_4 X + \beta_c \tag{5.4}$$

Model 1 describes the relationship between the independent variable X and dependent variable Y. Model 2 describes the relationship between the independent variable X and potential mediating variable, M. The significance

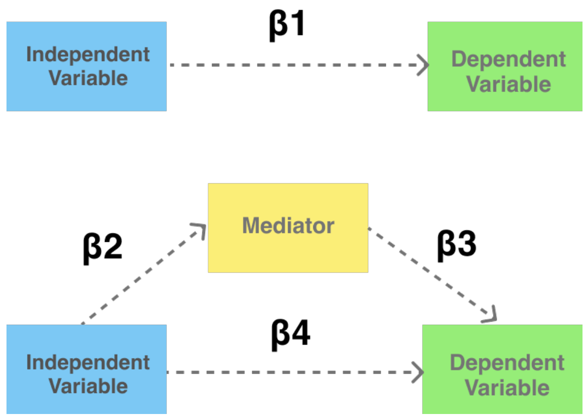


Figure 5.11: General Mediation Model

of Models 1 and 2 motivate the undertaking of mediation analysis. If there is no relationship between the potential mediator and independent variable, it cannot act as a mediator even if there is a significant relationship between it and the dependent variable. Model 3 is a multiple regression model that describes how the effect of our independent variable changes with the inclusion of the mediator. If β_4 loses its significance entirely, then the mediator explains all of the relationship between the independent and dependent variable; there is a "full mediation". If β_4 is still significant, but is lesser in magnitude, then the mediator explains some of the relationship between these variables; there is a "partial mediation". Essentially, the *difference* in the significance and magnitudes of β_1 and β_4, or "mediation effect", determines the conclusion of our mediation analysis.

5.3.2 Results of Mediation Analysis

In our implementation of mediation analysis, the dependent variable is QuaRCS score, the independent variable is UR status (gender, race, or socioeconomic), and the mediator is a composite variable (Figure 5.12). Our independent variable is a binary variable, 0 for Non-UR group and 1 for UR group. Therefore, the β_1, β_2 and β_4's in our models indicate the average difference in the QuaRCS score or the composite variable between Non-UR and UR groups (we will soon see that these quantities are negative, which indicate that UR-groups tend to have lower QuaRCS score/composite variable scores than Non-UR groups). Again, our aim is to determine whether our composite variables can mediate the relationship between certain demographic groups and score, and if/how these mediation effects differ between different demographic groups.

In R, we perform mediation analysis on all combinations of demographic groups and composite variables and note the statistically significant results[10] in Table 5.1. There are two quantities to note: the mediation effect (ME) and the proportion mediated (PM). Mediation effect, is the difference between β_1 and β_4, or a measure of the decrease in the achievement gap after accounting for the mediator. The proportion mediated (PE) is the proportion of the mediation effect relative to the initial gap in score, $(\beta_1 - \beta_4)\ /\beta_1$. In other words, it describes how much the mediator was able to 'explain' the initial gap between a demographic group and score.

Most of our composite variables have a significant mediation effect, explain-

[10]There is a function in R, mediate() that determines whether or not the differences in β_1 and $\beta 4$ are significant. If the difference is significant (p<0.001) then we say we have a significant mediation effect

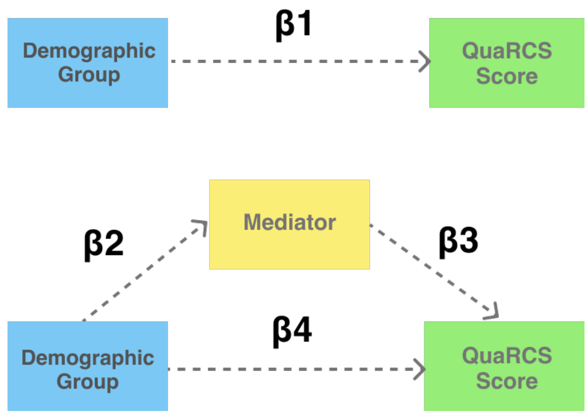


Figure 5.12: Our Mediation Model

Table 5.1: Mediation Analysis Results

URG	Mediator	β_1	β_2	β_3	β_4	ME†	PM†
Gender	School Math Affect	-0.067	-0.222	0.024	-0.061	-0.005	0.080
Gender	Numerical Anxiety	-0.067	0.401	-0.088	-0.035	-0.035	0.501
Race	Numerical Anxiety	-0.118	0.280	-0.088	-0.094	-0.025	0.208
Education (SES)	Numerical Anxiety	-0.134	0.408	-0.080	-0.097	-0.033	0.252
Gender	Numerical Self-Efficacy	-0.067	-0.418	0.076	-0.034	-0.032	0.481
Race	Numerical Self-Efficacy	-0.118	-0.156	0.076	-0.106	-0.012	0.100
Education (SES)	Numerical Self-Efficacy	-0.134	-0.401	0.067	-0.107	-0.027	0.202
Gender	Daily Life Math Affect	-0.067	-0.252	0.080	-0.046	-0.020	0.306
Race	Daily Life Math Affect	-0.118	-0.182	0.079	-0.104	-0.014	0.122
Education (SES)	Daily Life Math Affect	-0.134	-0.314	0.073	-0.111	-0.023	0.172
Gender	Course Motivation	-0.067	-0.118	0.052	-0.060	-0.006	0.093
Race	Course Motivation	-0.118	-0.266	0.049	-0.105	-0.013	0.111
Education (SES)	Course Motivation	-0.134	-0.430	0.045	-0.115	-0.019	0.144
Gender	Numerical Relevancy	-0.067	-0.176	0.050	-0.057	-0.009	0.134
Race	Numerical Relevancy	-0.118	-0.161	0.049	-0.110	-0.008	0.066

† ME: Mediation Effect, PM: Proportion Mediated

ing between 6% and 50% of the relationship between a demographic group and score. There were only a handful of mediations that were not significant, specifically: School Math Affect for race and SES, and Numerical Relevancy for SES. In these cases, the β_2 coefficient (the relationship between the demographic group and the mediator) was not statistically significant.

Gender Effects

All composite variables were successfully able to partially mediate the relationship between gender and score. The most notable are Numerical Anxiety and Self-Efficacy, which each explain roughly half the relationship. Therefore, the achievement gap narrows by half when accounting for Numerical Anxiety or Numerical Self-Efficacy. Of the composite variables that were not present in the literature and emerged in factor analysis, Daily Life Math Affect explained the largest portion. It mediated 30% of the relationship between gender and score, while School Math Affect, Course Motivation, and Numerical Relevancy mediated 8%, 9%, and 13%, respectively.

Socioeconomic Status/ Parental Education Effects

Most composite variables mediated the relationship between SES (Parent Education) and score, with the exception of School Math Affect and Numerical Relevancy. Similar to gender, Numerical Anxiety, Numerical Self-Efficacy, and Daily Life Math Affect explained larger portions of the score gap, 25%, 20%, and 17%, respectively. Interestingly, Course Motivation explained 14% of the relationship; its mediation effect is larger for SES than for gender.

Mediations for underrepresented gender group are higher than SES. However, this is largely due to the fact that gender has the narrowest score gap. Differences in score between UR Gender and Non-UR Gender is 6.7% while the difference is 13.4% between low and high SES. However, mediation effect sizes are similar for both groups–around 3% of score is explained by Numerical Anxiety and Self Efficacy, and 2% by Daily Life Math Affect. Proportions mediated are naturally lower for the low SES group simply because the initial gap between SES groups is much larger. In other words, since the proportion mediated is relative to the size of the gap, a mediation effect of the same magnitude represents a proportionally larger increase in score for the wider achievement gap.

Race/Ethnicity Effects

Mediations of the composite variables on score for the UR Race group are lower–no greater than 20%. These results suggest that the composite variables do not explain racial/ethnic differences in score as well as they do for gender and SES. However, this is likely due to the diverse make-up of the UR race/ethnicity group. In the literature, particularly for racial/ethnic demographics, the relationship between self-efficacy and performance in mathematics varies. In one study, Asian students were shown to have lower self-efficacy but high math performance, while Black/African American students had higher self-efficacy and lower math performance [27]. Therefore, relationships between the composite variables and score vary within the single "underrepresented" race group, and are not as clear between UR and Non-UR race groups as for gender, leading to lower mediation effects. This also accounts for the lack of significance in the differences in means observed in Section 5.1.

While UR and Non-UR groups were made initially for simplicity's sake, for the case of race, it seems especially important to examine specific racial subgroups. Therefore, we performed mediation analysis, using White students as our baseline group, for three of the largest racial/ethnic groups: Black or African American (AA), Asian, and Hispanic/Latinx (Table 5.2). While Asian students were categorized in the non-UR group, they are still notably distinct from White students and underperform on the QuaRCS assessment. There is a significant difference in score between Asian students and White students, quantified by β_1 (though to a lesser degree than for Black/AA or Hispanic/Latinx students).

Table 5.2: Mediation Analysis Results: Subgroups

Race/Ethnicity	Mediator	β_1	β_2	β_3	β_4	ME†	PM†
Asian	Numerical Anxiety	-0.079	0.388	-0.076	-0.048	-0.029	0.382
Asian	Numerical Self-Efficacy	-0.079	-0.267	0.063	-0.063	-0.017	0.210
Asian	Course Motivation	-0.079	-0.421	0.034	-0.065	-0.014	0.181
Black or African American	Numerical Anxiety	-0.193	0.390	-0.082	-0.165	-0.032	0.162
Black or African American	Course Motivation	-0.193	-0.520	0.029	-0.178	-0.015	0.079
Black or African American	Numerical Relevancy	-0.193	-0.338	0.043	-0.180	-0.015	0.075
Hispanic or Latinx	Numerical Anxiety	-0.170	0.504	-0.083	-0.130	-0.042	0.242
Hispanic or Latinx	Numerical Self-Efficacy	-0.170	-0.415	0.062	-0.145	-0.026	0.151
Hispanic or Latinx	Daily Life Math Affect	-0.170	-0.378	0.070	-0.145	-0.026	0.154
Hispanic or Latinx	Course Motivation	-0.170	-0.482	0.034	-0.154	-0.016	0.096
Hispanic or Latinx	Numerical Relevancy	-0.170	-0.160	0.046	-0.164	-0.007	0.043

† ME: Mediation Effect, PM: Proportion Mediated

What is most striking about these analyses are the differences in which mediators are significant for the various racial/ethnic groups. For instance, only three composite variables are significant mediators for Asian and Black/AA students, while five are significant for Hispanic/Latinx. Daily Life Math Affect was a strong mediator in the cases of underrepresented gender and SES groups, but not in the cases of Asian or Black/AA groups. Numerical Self-Efficacy is also not significant mediator for Black/AA.

Additionally, the mediation effect of Numerical Anxiety for Hispanic/Latinx students is larger than the effect for Asian and Black/AA, explaining 1% more of the difference in score. Numerical Self-Efficacy also has a larger mediation effect for the Hispanic/Latinx students, explaining 1% more than Asian students. Whether this difference is meaningful or statistically significant is unknown, however. But if significant, it suggests that there may, in fact, be differences between specific races/ethnicities.

In the case similar to the UR Gender group, the proportion mediated by Numerical Anxiety, Self-efficacy and Course Motivation is higher for Asian students than other groups. For instance, Numerical Anxiety mediates 38% of the differences in score for Asian students, but 16% for Black/AA, and 24% for Hispanic/Latinx. Again, this is entirely related to a smaller score gap, between Asian and White students. On average, there is a 7.9% score difference between Asian and White students, 17% between Hispanic/Latinx, and 19.3% difference between Black/AA students (quantified by β_1). While the effect sizes are comparable between groups, Black/AA and Hispanic/Latinx students on average score more than 10% lower than White students. Therefore, the proportion mediated is smaller.

The score gap between Black/AA and White is the largest, yet only three of six composite variables mediate the difference in score. Therefore, our composite factors cannot easily explain the lower performance of Black/AA students. This suggests that there are other factors not measured by the QuaRCS instrument, that may underlie these differences.

Generally, the variables with the highest mediation effect are Numerical Anxiety, Numerical Self-Efficacy, and Daily Life Math Affect. While the explanatory power of Numerical Anxiety and self-efficacy follows consistently from the literature, our constructed variables, like Daily Life Math Affect, require more consideration. Recall that Daily Life Math Affect describes how easy/hard, stressful/not stressful, or confusing/not confusing using math is in daily life, and contrast this with School Math Affect which describes similar aspects but in the context of school. When understanding quantitative literacy performance, student affect surrounding school math does not explain differences in performance, but affect surrounding the use of mathematics in daily life does. The QuaRCS questions themselves are more reflective of daily life experiences (as they are questions set in everyday contexts). Therefore, it makes sense that Daily Life Math Affect is a stronger mediator.

What is interesting is that Daily Life Affect can explain more of the difference in QuaRCS score than Numerical Relevancy. Recall that Numerical Relevancy is a measure of how frequently students believe numerical situations are encountered in everyday life. Therefore, students with higher Numerical Relevancy are likely to understand the value of developing quantitative skills. So, what we observe is that feelings elicited from experiences developing numerical skills predict score more strongly than a more positive value judgement of the utility of developing skills. Even Course Motivation, the degree to which a student's motivation in taking a course was more intrinsic, had a higher mediation effect than Numerical Relevancy.

While these mediators ostensibly explain the relationship independent of other composite variables, it is important to remember that these variables are correlated with one another. For instance, Numerical Self-Efficacy and Daily Life Math Affect are high correlates of Numerical Anxiety and can have overlapping explanatory power. Therefore, we cannot claim that, for example, Numerical Anxiety explains 50%, Numerical Self-Efficacy explains another 50%, for a total of 100% of the relationship between gender and score explained. However, our composite variables are distinct constructs that can uniquely explain part of the relationship, even if they explain partially overlapping phenomena[11].

[11] Unfortunately, extracting the unique explanatory power requires additional statistical analyses, which is beyond the scope of this investigation

Chapter 6

Conclusion

Quantitative literacy or numeracy is the ability to utilize, apply, and reason with mathematical, statistical, and numerical skills. Science, health, and financial literacy are contingent upon numeracy; circumstances in daily life such as understanding statistics presented in the news or knowing the price of an item during a sale require the use of numerical skills. Consequently, we recognize the importance of developing quantitative literacy, yet, our numerical skills are comparatively lower than other developed countries [10]. Additionally, in the US, we see that there are significant disparities in quantitative literacy among demographic groups.

To explore why these disparities exist, we utilized data from the Quantitative Reasoning for College Science (QuaRCS) study. It measures students' performance on a quantitative literacy assessment and gathers responses to non-quantitative questions probing the affects and attitudes of students in addition to demographic information. Biological variables do not contribute to innate differences in ability, but there is evidence that psychological variables like self-efficacy and math anxiety do contribute to differences in performance [17][21][22][20]. Equipped with this dataset, we investigated whether factors that influence affect and attitude can explain differences in score.

To prepare our data for analysis, we performed Factor Analysis, a data reduction technique to extract underlying variables in our data. From this procedure, several composite variables emerged: Numerical Self-Efficacy, Numerical Anxiety, Numerical Relevancy, School Math Affect, Daily Life Math Affect, and Course Motivation. Using these composite variables, along with score and demographic information, we proceeded with our investigation of understanding the relationships of demographics and score.

Using linear regression and mediation analysis, we determined that our composite variables, which are related to one's attitude and affect, are predictive of performance on the QuaRCS assessment. Generally, Numerical Self-Efficacy, Numerical Anxiety, and Daily Life Math Affect were more highly correlated with score than the other composite variables, namely School Math Affect, Numerical Relevancy, and Course Motivation. Nevertheless, the differences in means of these variables between underrepresented (UR) and non-underrepresented (Non-UR) groups partially explained the achievement gap of UR groups. We found that certain composite variables were able to partially mediate the relationship between a UR group and score by as much as 50%. Composite variable(s) that acted as mediators varied among UR groups, however, which emphasizes that the story explaining differences is unique to each group. There are many factors that are at play, some of which have not been uncovered or investigated in this thesis. Additional psychological variables may underlie differences in score. And, we must also consider other factors that cause differences in score such as structural inequalities or educational debt that cannot be easily measured.

Nevertheless, we do know that psychological variables such as Numerical Self-Efficacy, Numerical Anxiety and Daily Life Math Affect play a significant role in impacting student performance. Thus, we–teachers, professors, students–must be able to identify and be cognizant of our (and other's) affect, attitudes, and beliefs. We must be aware of the idiosyncratic circumstances and experiences that shape them. We must recognize that high or low academic performance is not dependent on innate or biological characteristics, rather, dependent on extrinsic factors surrounding the individual. In other words, given conducive environments, proper resources, adequate support, etc. students can and will be academically successful.

There are several limitations and caveats to our analysis to consider and improve on. First, our sample is not representative of all college students in the US. Participants are recruited by contacting professors of gen-ed courses at various universities and colleges. Of those who have agreed to enroll their courses in this study, many professors teach at a liberal arts college in the Northeast. While we do have a large proportion of students from several research universities, our sample is not representative of all college students in the nation. Whether a student is enrolled in a liberal arts college, a research university, or a community college in different areas of the country may factor into how the student performs on the QuaRCS assessment. However, as the

QuaRCS study matures and establishes more robust infrastructure, we can expand our outreach and gather more representative data. With a larger and more diverse sample size, we can investigate various demographic groups like Native American or LGBTQ+ students for future studies.

Secondly, our data is not normal. Many statistical procedures assume normality as a condition, yet we continued with our analysis. This is not an uncommon practice, but also not an ideal one. There may be some transformations on data to investigate, however, working with non-normal data may be the only option.

Thirdly, we also did not consider removing or understanding the outliers in our data. Outliers can skew data and interfere with our interpretations. This is a typical procedure that was neglected and should be considered in future analyses.

While the investigation of this thesis concludes here, there are a plethora of future directions to explore. An immediate direction is an extraction of individual/unique effects of our composite variables. We saw that composite variables partially mediated relationships between UR status and score. However, composite variables are all correlated with one another. This creates difficultly in making absolute claims regarding how composite variables explain relationships independently of one another. Another direction to explore is using multiple mediators in mediation analysis. If we combine several of our variables in mediation, do we see a larger mediation effect? This analysis may address the uniqueness or degree of overlap of each variable.

We did not consider certain demographic groups, such as disability status and academic major. More specific demographics, such as students at the intersection of multiple demographic groups, are worth exploring as well, sample size permitting. For example, we could investigate differences between White women and Black women, or low SES White students and low SES Hispanic students. As the theory of intersectionality explains, there is importance in considering the multiple identities of an individual which carry varying combinations of experiences. One identity is not sufficient in generalizing an individual, but rather requires multiple identities to explain the complex relationships between identities [56]. The literature on numeracy does not cover such groups extensively, therefore, it may prove to be worthwhile.

We also did not consider the effort variable in our analysis. Ostensibly, effort seems like a simple variable to work with. However, because it can be treated as either an independent or dependent variable, a control or experimental variable, it adds a level of complexity to the analysis. Do we want to compare high effort and low effort students as a 'demographic' group? Do we want to investigate how our composite variables predict effort, which in turn predicts score? Do we want to remove students of certain effort levels from analysis? It's a complicated variable that can be utilized in different ways, therefore is worth its own separate investigation.

There are many more directions that we can take with QuaRCS data, for instance, looking at causal patterns/directionality of variables, changes in QuaRCS score between Pre- and Post-assessments, and identification of exemplary instructors. The opportunities of exploration are rich.

Appendix A

Supplements to OECD and NAAL

OECD
BETTER POLICIES FOR BETTER LIVES

PIAAC **Numeracy – sample items**

In the Survey of adult skills (PIAAC). numeracy is defined as the ability to use, apply, interpret, and communicate mathematical information and ideas. It is an essential skill in an age when individuals encounter an increasing amount and wide range of quantitative and mathematical information in their daily lives. Numeracy is a skill parallel to reading literacy, and it is important to assess how these competencies interact, since they are distributed differently across subgroups of the population.

The items are presented in the form delivered by the computer-based version of the assessment. To answer the questions, respondents need to click in the appropriate box, and/or type figures in the space provided.

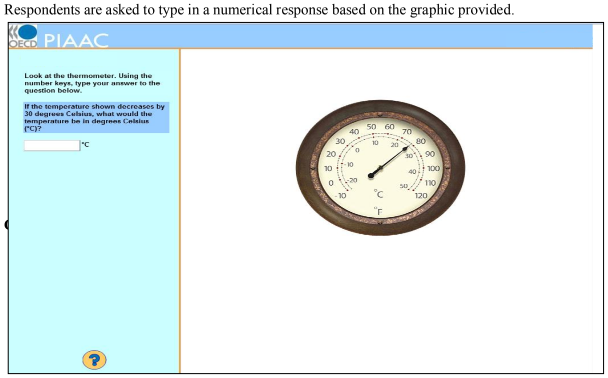
Numeracy - Sample Items

Sample Item 1: Thermometer

This item (of low difficulty) focuses on the following aspects of the numeracy construct:

Content	Dimension and shape
Process	Act upon, use (measure)
Context	Every day or work

Respondents are asked to type in a numerical response based on the graphic provided.



Sample Item 2: Wind power stations

This sample item (of medium difficulty) focuses on the following aspects of the numeracy construct:

Content	Quantity and Number
Process	Act upon, use (compute)
Context	Community and society

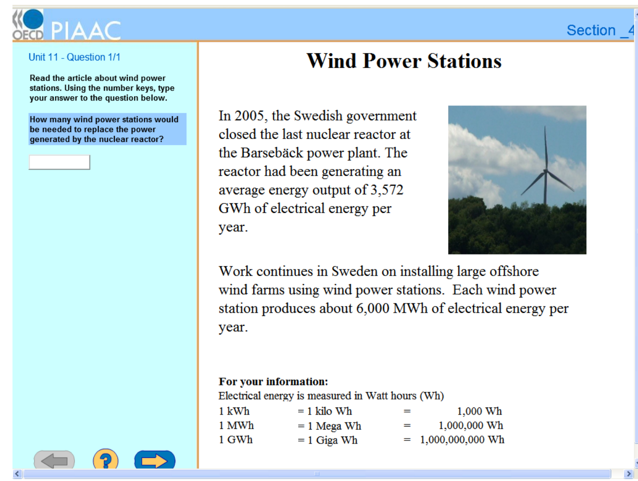


Figure A.1: Sample Questions From OECD

Table 1-2. Overview of the literacy levels

Level and definition	Key abilities associated with level
Below Basic indicates no more than the most simple and concrete literacy skills. Score ranges for *Below Basic*: Prose: 0–209 Document: 0–204 Quantitative: 0–234	Adults at the *Below Basic* level range from being nonliterate in English to having the abilities listed below: ■ locating easily identifiable information in short, commonplace **prose** texts ■ locating easily identifiable information and following written instructions in simple **documents** (e.g., charts or forms) ■ locating numbers and using them to perform simple **quantitative** operations (primarily addition) when the mathematical information is very concrete and familiar
Basic indicates skills necessary to perform simple and everyday literacy activities. Score ranges for *Basic:* Prose: 210–264 Document: 205–249 Quantitative: 235–289	■ reading and understanding information in short, commonplace **prose** texts ■ reading and understanding information in simple **documents** ■ locating easily identifiable **quantitative** information and using it to solve simple, one-step problems when the arithmetic operation is specified or easily inferred
Intermediate indicates skills necessary to perform moderately challenging literacy activities. Score ranges for *Intermediate*: Prose: 265–339 Document: 250–334 Quantitative: 290–349	■ reading and understanding moderately dense, less commonplace **prose** texts as well as summarizing, making simple inferences, determining cause and effect, and recognizing the author's purpose ■ locating information in dense, complex **documents** and making simple inferences about the information ■ locating less familiar **quantitative** information and using it to solve problems when the arithmetic operation is not specified or easily inferred
Proficient indicates skills necessary to perform more complex and challenging literacy activities. Score ranges for *Proficient*: Prose: 340–500 Document: 335–500 Quantitative: 350–500	■ reading lengthy, complex, abstract **prose** texts as well as synthesizing information and making complex inferences ■ integrating, synthesizing, and analyzing multiple pieces of information located in complex **documents** ■ locating more abstract **quantitative** information and using it to solve multi-step problems when the arithmetic operations are not easily inferred and the problems are more complex

NOTE: Although the literacy levels share common names with the National Assessment of Educational Progress (NAEP) levels, they do not correspond to the NAEP levels.
SOURCE: Hauser, R.M, Edley, C.F. Jr., Koenig, J.A., and Elliott, S.W. (Eds.). (2005). *Measuring Literacy: Performance Levels for Adults, Interim Report*. Washington, DC: National Academies Press; White, S. and Dillow, S. (2005). *Key Concepts and Features of the 2003 National Assessment of Adult Literacy* (NCES 2006-471). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Figure A.2: Literacy Levels on NAAL

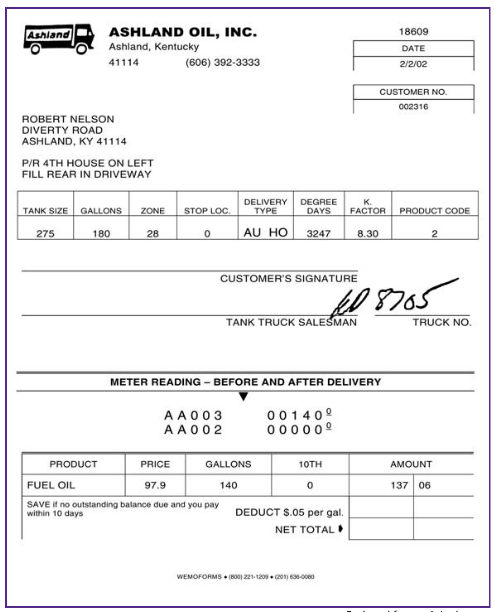
Appendix A: Sample Assessment Questions

Quantitative Literacy Question

Suppose that you had your oil tank filled with 140.0 gallons of oil, as indicated on the bill, and you wanted to take advantage of the five cents ($.05) per gallon deduction.

1. Figure out how much the deduction would be if you paid the bill within 10 days. Enter the amount of the deduction on the bill in the space provided.

___



ASHLAND OIL, INC.
Ashland, Kentucky
41114 (606) 392-3333

18609

DATE
2/2/02

CUSTOMER NO.
002316

ROBERT NELSON
DIVERTY ROAD
ASHLAND, KY 41114

P/R 4TH HOUSE ON LEFT
FILL REAR IN DRIVEWAY

TANK SIZE	GALLONS	ZONE	STOP LOC.	DELIVERY TYPE	DEGREE DAYS	K. FACTOR	PRODUCT CODE
275	180	28	0	AU HO	3247	8.30	2

CUSTOMER'S SIGNATURE

TANK TRUCK SALESMAN 8705 TRUCK NO.

METER READING – BEFORE AND AFTER DELIVERY

AA003 00140
AA002 00000

PRODUCT	PRICE	GALLONS	10TH	AMOUNT	
FUEL OIL	97.9	140	0	137	06
SAVE if no outstanding balance due and you pay within 10 days		DEDUCT $.05 per gal.			
		NET TOTAL			

WEMOFORMS • (800) 221-1209 • (201) 636-0080

Reduced from original copy

Correct answer
$7.00

Percentage of adults who answered the question correctly, 2003

All Adults	*Below Basic*	*Basic*	*Intermediate*	*Proficient*
52	1	40	92	100

NOTE: Adults are defined as people 16 years of age and older living in households or prisons. Adults who could not be interviewed because of language spoken or cognitive or mental disabilities (3 percent in 2003) are excluded from these data.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2003 National Assessment of Adult Literacy.

87

Figure A.3: Sample Question from NAAL

Appendix B

Supplemental Tables and Figures

Table B.1: QuaRCS Quantitative Questions

#	Question	Skill Assessed	Version
1	You have a rectangular fish tank that's 10 inches tall, 20 inches wide, and 15 inches deep. If the volume of one gallon of water is 231 cubic inches, then how many gallons are required to fill the tank?	AR AV UD	Full 1.0
2	Your grocery store has a 20 ounce jar of peanut butter for $4.00, and a 45 ounce jar for $9.00. Which purchase will get you the best price per ounce?	PR AR	Full 1.0 Lite 2.0
3	A college that typically has 50,000 students experiences an increase in enrollment to 55,000 students. By what percentage did enrollment increase?	AR PC	Full 1.0
4,5,6	According to the graph, what was the approximate population of City X in 1980? If the current population growth rate continues, which is the best estimate for the population of City X in the year 2050? Based on this graph, compare the population growth rates (i.e. increase in number of people per year) before and after 1970	GR ES PR AR	Full 1.0 Lite 2.0
7,8,9	Imagine you have already filled a measuring cup (like the one shown above) with the amount of peanut butter in the recipe and you want to add the correct amount of shortening on top of it. Which line on the measuring cup should you fill to with shortening?If your measuring cup has ounces on the side instead of cups, which line should you fill to when measuring the flour? There are 8 ounces in 1 cup. You have only a half-Tablespoon measuring spoon. How much should you fill it to get the correct amount of baking soda? There are 3 teaspoons in 1 tablespoon.	AR UD PR	Full 1.0
10,11,12	How many total injuries (including deaths) were sustained at Resort Y during this time period?What were the chances of a randomly-selected skier sustaining an injury of any kind (minor, severe or death) while at Resort Y during this time period?What proportion of severely injured skiers at Resort Y during this time period were intermediate skiers?	TR AR SP PC PR	Full 1.0 Lite 2.0
13,14	The graph above shows the predicted viewership of three television shows in two cities based on a poll of a small number of residents in each city. The poll has a reported error of 25%, shown as vertical error bars. Which of the following statements about the predicted viewers of Show A is most accurate?Which of the following predictions can be made based on the information (including errors) shown in the graph?Prediction 1: In City 2, more people will watch Show B than Show C Prediction 2: In City 1, Show C will have the smallest viewership Prediction 3: None of the three shows (A, B or C) will be equally popular in Cities 1 and 2	GR ER SP	Full 1.0
15,16	You purchased 100 square feet of solar panels for your roof. However, your local Homeowner's Association requires that solar panels not be visible from the road. You decide to put solar panels on the roof of your shed instead. The shed has a flat 5 foot by 5 foot roof. Complete the following sentence: "To produce the same amount of power as your original design, you need to buy panels that produce ______ more power per unit area than your original panels."If you cover the shed with your original panels, how many more of the same size sheds would you have to put up in your backyard in order to fit the rest of the panels?	AR PR AV SP	Full 1.0
17	Your cable bill is $36 per month from January 1 through September 30 and then doubles to $72 per month starting October 1. What is your average monthly bill over the course of the entire calendar year (January-December)?	AR PR SP	Full 1.0 Lite 2.0

18	If you place $10 under your mattress every day for the next 40 years, approximately how much money will you have?	AR	Full 1.0 Lite 2.0
19,20	A newspaper conducts a survey and predicts that in the local election between Candidates A and B, Candidate A will receive 60% of the votes. The newspaper estimates the error in this prediction to be 5%.If the newspaper conducts another survey with 400 participants, how many people can report that they will vote for Candidate A for the result to be consistent with the original conclusion (that Candidate A will receive 60% of the votes with 5% error)?Several days later, the newspaper conducts another survey with 300 new participants. What is the minimum number of votes that Candidate A can receive in this new survey in order to be consistent with the original prediction (that Candidate A will receive 60% of the votes with 5% error)?	ER AR PC SP	Full 1.0
21,22	You want to carpet a 15 foot by 20 foot room. You have two carpet options to choose from. One is $1.50 per square foot and the other is $3.00 per square foot. How much more will your total bill be if you choose the more expensive carpet rather than the cheaper one?To carpet your 15 foot by 20 foot room and a hallway that is 4 feet by 12 feet, how much total carpet do you need?	AV AR PR	Full 1.0
23	If one scoop of lemonade powder is needed for every 12 ounces of water, then how many scoops should you add to three gallons of water to make it into lemonade?16 ounces = 1 Pint 2 Pints = 1 Quart 4 Quarts = 1 Gallon	UD AR PR PC ES	Full 1.0
24	A sweater that was originally $100 is on sale for 30% off. Which of the following coupons should you use to get the lowest final price?	AR PC PR	Full 1.0 Lite 2.0
25	You drove 200 miles on 11 gallons of gas. Which of these is closest to the number of miles per gallon that you got?	AR ES PR UD	Full 1.0 Lite 2.0
26,27	You purchased 100 square feet of solar panels for your roof. However, your local Homeowner's Association requires that solar panels not be visible from the road. You decide to put solar panels on the roof of your shed instead. The shed has a flat 5 foot by 5 foot roof. Complete the following sentence: "To produce the same amount of power as your original design, you need to buy panels that produce ______ more power per unit area than your original panels."If you cover the shed with your original panels, how many more of the same size sheds would you have to put up in your backyard in order to fit the rest of the panels?		Lite 2.0
28,29	You want to carpet a 15 foot by 20 foot room. You have two carpet options to choose from. One is $1.50 per square foot and the other is $3.00 per square foot. How much more will your total bill be if you choose the more expensive carpet rather than the cheaper one?To carpet your 15 foot by 20 foot room and a hallway that is 4 feet by 12 feet, how much total carpet do you need?		Lite 2.0

Figure B.1: Distribution of responses to the effort question on QuaRCS Full Spring 2015

Table B.4: QuaRCS Demographic Questions

Demographic Variable	Question	Version
PRE_MAJOR_NEW	Please select your major or majors from the list below. Arts & Humanities Social & Behavioral Sciences Biological & Health Sciences Physical Sciences Engineering, Mathematics, or Computer Science Education Business-related Journalism Undecided Other	Lt 1.0, Lt 2.0, Lt 2.1
PRE_GENDER	With which gender identity do you most identify? Male Female Other	Fl 1.0, Fl 2.0, Lt 1.0, Lt 2.0, Lt 2.1
PRE_RACE	With which racial or ethnic group(s) do you identify? Choose all that apply. African American Asian Caucasian Hispanic/Latino Native American Other Prefer not to specify	Fl 1.0, Fl 2.0, Lt 1.0, Lt 2.0
PRE_RACE_V2	With which racial or ethnic group(s) do you identify? Choose all that apply. Asian or Pacific Islander (East Asian, South Asian, Southeast Asian, Native Hawaiian or Pacific Islander) Arab or Middle Eastern Black or African American White or Caucasian (Non-Hispanic) Hispanic of Latinx Native American Other: Prefer not to specify	Lt 2.1
PRE_DIS	Have you ever been diagnosed with any of the following? Please select all that apply. A physical disability A cognitive disability A learning disability I prefer not to specify None	Fl 1.0, Fl 2.0, Lt 1.0, Lt 2.0, Lt 2.1
PRE_MAXEDU	What is the highest level of education obtained by a parent or guardian in the household where you grew up? Informal Education (not including homeschooling) Early Elementary (K-3rd grade) Elementary (4th -5th grade) Middle School (6th - 8th grade) Some High School High School Diploma or GED Career, Trade, or Technical Certification Associate's Degree (Community College) Bachelor's Degree (College or University) Master's Degree Professional Degree (JD, RN, MD) Doctorate (PhD)	Lt 2.1

Table B.5: Non-Quantitative QuaRCS Questions

Variable	Question	Version
PRE_FREQEN	In your everyday life, how frequently do you encounter situations similar to problems in this survey?	Fl 1.0, Fl 2.0, Lt 1.0, Lt 2.0, Lt 2.1
PRE_DAILYM	How frequently do you do calculations in your everyday life?	Fl 1.0, Fl 2.0, Lt 1.0, Lt 2.0, Lt 2.1
PRE_DAILYG	How frequently do you encounter graphs and tables in your daily life?	Fl 1.0, Fl 2.0, Lt 1.0, Lt 2.0, Lt 2.1
PRE_ATT_SC_1 PRE_ATT_SC_2 PRE_ATT_SC_3 PRE_ATT_SC_4 PRE_ATT_SC_5 PRE_ATT_DL_1 PRE_ATT_DL_2 PRE_ATT_DL_3 PRE_ATT_DL_4 PRE_ATT_DL_5	Think of several instances in which you have had to use math to solve a problem for school/in your daily life. How do these experiences usually rate on the following scales between two opposite adjectives? (on a Likert scale from 1-4) School: Easy - Hard Fun - Boring Useful - Pointless Straightforward - Confusing Not Stressful - Stressful Daily life: Easy - Hard Fun - Boring Useful - Pointless Straightforward - Confusing Not Stressful - Stressful	Lt 1.0, Lt 2.0, Lt 2.1
PRE_CF_MEAN	(averages of) How confident are you in the answer that you just chose?	Lt 1.0, Lt 2.0, Lt 2.1
PRE_LK1	Rate the degree to which you agree with the following statement: "I feel confident using numbers in my non-math courses"	Fl 1.0, Fl 2.0, Lt 1.0, Lt 2.0, Lt 2.1
PRE_LK2	Rate the degree to which you agree with the following statement: "I feel confident using numbers in my everyday life"	Fl 1.0, Fl 2.0, Lt 1.0, Lt 2.0, Lt 2.1
PRE_LK5	Rate the degree to which you agree with the following statement: "I am satisfied with my current level of numerical/mathematical skill"	Fl 1.0, Fl 2.0, Lt 1.0, Lt 2.0, Lt 2.1
PRE_ANX.1_1 PRE_ANX.1_2 PRE_ANX.1_3 PRE_ANX.1_4	Please indicate the level of your anxiety in the following situations. (Not at all - a little - a fair amount - a lot) When a teacher uses an equation to explain something When I have to extract information from a graph in order to answer a homework problem When I have to calculate a tip When someone uses numbers or statistics to make an argument that I disagree with	Lt 1.0, Lt 2.0, Lt 2.1
PRE_WHYMAJ_1 PRE_WHYMAJ_2 PRE_WHYMAJ_3 PRE_WHYMAJ_4 PRE_WHYMAJ_5 PRE_WHYMAJ_6 PRE_WHYMAJ_7 PRE_WHYMAJ_8	I chose (or will choose) my major because: Check all that apply I like the subject I feel that it will help me get a job that I enjoy after graduation I feel that it will help me get a well-paying job after graduation I am good at it I chose a major that would avoid math as much as possible I chose a major that would avoid writing as much as possible I'm not sure yet Other	Fl 1.0, Fl 2.0, Lt 1.0, Lt 2.0, Lt 2.1
PRE_WHYCS_1 PRE_WHYCS_2 PRE_WHYCS_3 PRE_WHYCS_4 PRE_WHYCS_5 PRE_WHYCS_6 PRE_WHYCS_7 PRE_WHYCS_8	Why did you choose to take this course? Check all that apply. It is a prerequisite for courses in my major To fulfill a university general education requirement It sounded interesting It sounded easy I heard the class was good I heard the instructor was good I'm not sure yet Other	Fl 1.0, Fl 2.0, Lt 1.0, Lt 2.0, Lt 2.1
PRE_EFFORT	Knowing that this survey is being used for research to try to improve courses like yours and that your answer to this question will not be shared with your instructor, please honestly describe the amount of effort that you put into this survey. I just clicked through and chose randomly to get the participation credit I didn't try very hard I tried for a while and then got bored I tried pretty hard I tried my best on all or most of the questions	Fl 1.0, Fl 2.0, Lt 1.0, Lt 2.0, Lt 2.1

Table B.6: General: Medians and Means

	Median	Mean
Score	0.767	0.714
Effort	4.000	4.002
SCHMATH	-0.317	0.000
NUMANX	-0.114	0.000
SELFEFF	-0.181	0.000
DLMATH	0.110	0.000
MOT	-0.730	0.000
NUMREL	0.090	0.000

Table B.7: Gender: Medians and Means

Variables	Median		Mean		T.test
	UR	Non-Ur	UR	Non-UR	
Score	0.733	0.800	0.684	0.751	<0.001
Effort	4.000	4.000	4.039	3.961	0.071
SCHMATH	-0.321	0.112	-0.102	0.120	<0.001
NUMANX	0.039	-0.431	0.181	-0.220	<0.001
SELFEFF	-0.181	-0.022	-0.191	0.228	<0.001
DLMATH	0.110	0.110	-0.113	0.138	<0.001
MOT	-0.730	0.041	-0.054	0.064	0.004
NUMREL	-0.062	0.149	-0.078	0.098	<0.001

Table B.8: Race: Medians and Means

Variables	Median		Mean		T.test
	UR	Non-Ur	UR	Non-UR	
Score	0.667	0.800	0.617	0.735	<0.001
Effort	4.000	4.000	3.983	4.013	0.600
SCHMATH	-0.321	-0.299	-0.053	0.012	0.213
NUMANX	0.110	-0.161	0.230	-0.050	<0.001
SELFEFF	-0.181	-0.181	-0.126	0.030	0.003
DLMATH	0.110	0.110	-0.151	0.031	0.001
MOT	-0.730	0.041	-0.216	0.050	<0.001
NUMREL	-0.122	0.090	-0.135	0.026	0.004

Table B.9: Socioeconomic Status: Medians and Means

Variables	Median		Mean		T.test
	UR	Non-Ur	UR	Non-UR	
Score	0.600	0.800	0.621	0.755	<0.001
Effort	4.000	4.000	3.874	4.045	0.005
SCHMATH	-0.321	-0.317	0.001	0.027	0.652
NUMANX	0.232	-0.384	0.270	-0.138	<0.001
SELFEFF	-0.181	-0.181	-0.240	0.162	<0.001
DLMATH	0.075	0.110	-0.211	0.103	<0.001
MOT	-0.730	0.041	-0.314	0.115	<0.001
NUMREL	-0.122	0.090	-0.140	-0.006	0.018

Appendix C

Validating Factor Analysis in R

The first iteration of factor analysis was performed on all full version assessments given between Fall 2015 and Fall 2016 semesters using SPSS Statistics software. As I am unfamiliar with SPSS syntax, all statistical analyses were performed in R. To validate the factor analysis procedures taken in R, we compared the factor loadings generated in the QuaRCS 2 paper using SPSS Statistics software with the factor loadings generated using R C.1. The loadings are not identical, however, they differ by no more than 0.01. The small differences we attribute to the differences in the nature of SPSS and R code. Therefore, we will subsequently proceed with our analyses in R.

	Factor					
	1		2		3	
Variable	SPSS	R	SPSS	R	SPSS	R
PRE_LK1	0.889	0.886				
PRE_LK2	0.834	0.830				
PRE_ATT_3	0.602	0.590				
PRE_DIFF	0.602	0.592				
CONF_AVG	-0.547	-0.530				
PRE_DAILYM			0.721	0.729		
PRE_DAILYG			0.647	0.655		
PRE_FREQEN			0.583	0.587		
PRE_ATT_1			-0.509	-0.508		
PRE_LK4			-0.436	-0.430		
PRE_ATT_2			-0.405	-0.399		
PRE_WHYMAJ_LIKEJOB					0.597	0.599
PRE_WHYMAJ_GOOD					0.528	0.528
PRE_WHYMAJ_MONEY					0.493	0.493
PRE_WHYMAJ_LIKE					0.444	0.445
PRE_WHYMAJ_NOTSURE					-0.428	-0.429

Figure C.1: Factor Analysis in SPSS and R

Appendix D

Code Snippets

D.1 Wrangling

```
```{r, packages}
require(mosaic)
require(forcats)
require(tidyverse)
require(leaps)
```
```{r, echo = F}
# read in data
full <- read.csv("Master_data_v5_race_edit.csv")

# selecting relevant variables

vars <- c("PRE_PCT_TOTAL", "PRE_EFFORT",

          "PRE_NUMANX", "PRE_NUMREL",
          "PRE_SELFEFF", "PRE_SCHMATH", "PRE_DLMATH", "PRE_MOT",

          "PRE_GENDER",
          "PRE_MAXEDU",
```
```

```
          "PRE_DIS_1", "PRE_DIS_2", "PRE_DIS_3", "PRE_DIS_4",

          "PRE_MAJOR_NEW_1", "PRE_MAJOR_NEW_2", "PRE_MAJOR_NEW_3", "PRE_MAJOR_NEW_4",
          "PRE_MAJOR_NEW_5", "PRE_MAJOR_NEW_6", "PRE_MAJOR_NEW_7", "PRE_MAJOR_NEW_8",
          "PRE_MAJOR_NEW_9", "PRE_MAJOR_NEW_10", "PRE_MAJOR_NEW_12","PRE_MAJOR_NEW_13",

          "PRE_RACE_1", "PRE_RACE_2", "PRE_RACE_3", "PRE_RACE_4", "PRE_RACE_5",
          "PRE_RACE_6", "PRE_RACE_7",

          "PRE_RACE_V2_1", "PRE_RACE_V2_2", "PRE_RACE_V2_3", "PRE_RACE_V2_4",
          "PRE_RACE_V2_5", "PRE_RACE_V2_6", "PRE_RACE_V2_7", "PRE_RACE_V2_8",

          "PRE_RACE_ASIAN_1", "PRE_RACE_ASIAN_2", "PRE_RACE_ASIAN_3", "PRE_RACE_ASIAN_4",
          "PRE_RACE_ASIAN_5")

v1 <- dplyr::select(full, vars)

#Identifying URG groups for later use
URG_RACE <- c('black_afram', 'hisplat', 'nat_american', 'southeast_asian', 'hawaiian_pi', 'other_asian', 'other', 'ur_asian', 'URG multiracial')
URG_GENDER <- c('female', 'other')
URG_DIS <- c('learning', 'cognitive')
URG_EDU <- c('<HS', 'HS','TC', 'AD') #URG for education isn't correct term but for the sake of consistency, will use
LOW_EFF <- c('1', '2')
HIGH_EFF <- c('4','5')

NURG_RACE <- c('asian', 'white', 'arab_me', 'multiracial')
NURG_EDU <- c('BD','MD','PD', 'PhD')

# Renaming Variables
```

```
v2 <- rename(v1,
                       SCORE = PRE_PCT_TOTAL,
                       EFFORT = PRE_EFFORT,

                       GENDER = PRE_GENDER,

                       MAXEDU = PRE_MAXEDU,

                       DIS_PHYS = PRE_DIS_1,
                       DIS_COG = PRE_DIS_2,
                       DIS_LEARN = PRE_DIS_3,

                       MAJ_ARTHUM = PRE_MAJOR_NEW_1,
                       MAJ_SOCSCI = PRE_MAJOR_NEW_2,
                       MAJ_BIOSCI = PRE_MAJOR_NEW_3,
                       MAJ_PHYSCI = PRE_MAJOR_NEW_4,
                       MAJ_ENGMATHCS = PRE_MAJOR_NEW_5,
                       MAJ_EDUC = PRE_MAJOR_NEW_6,
                       MAJ_BUSI = PRE_MAJOR_NEW_7,
                       MAJ_JOURN = PRE_MAJOR_NEW_8,
                       MAJ_UNDEC = PRE_MAJOR_NEW_9,
                       MAJ_OTHER = PRE_MAJOR_NEW_10,

                       RACEV1_AFRAM = PRE_RACE_1,
                       RACEV1_ASIPAC = PRE_RACE_2,
                       RACEV1_CAUC = PRE_RACE_3,
                       RACEV1_HISPLAT = PRE_RACE_4,
                       RACEV1_NATAM = PRE_RACE_5,
                       RACEV1_OTHER = PRE_RACE_6,
                       RACEV1_NAN = PRE_RACE_7, #Prefer not to specify

                       RACEV2_ASIPAC = PRE_RACE_V2_1,
                       RACEV2_BLAFR = PRE_RACE_V2_2,
                       RACEV2_WHITE = PRE_RACE_V2_3,
                       RACEV2_HISPLAT = PRE_RACE_V2_4,
                       RACEV2_ARABME = PRE_RACE_V2_5,
                       RACEV2_NATAM = PRE_RACE_V2_6,
```

```
                       RACEV2_OTHER = PRE_RACE_V2_7,
                       RACEV2_NAN = PRE_RACE_V2_8, #Prefer not to specify

                       RACE_EASIA = PRE_RACE_ASIAN_1,
                       RACE_SASIA = PRE_RACE_ASIAN_2,
                       RACE_SEASIA = PRE_RACE_ASIAN_3,
                       RACE_NHPI = PRE_RACE_ASIAN_4,
                       RACE_OASIA = PRE_RACE_ASIAN_5,

                       )

# Recoding and collapsing
v3 <- mutate(v2,
             # Recode effort
             EFFORT = as.factor(EFFORT),

             # Recode gender
             GENDER = fct_recode(as.factor(GENDER),
                               male ="1",
                               female = "2",
                               other = "3"),

             # Recode Parent Education
             MAXEDU = fct_recode(as.factor(MAXEDU),
                                 "<HS" = '1',
                                 "<HS" = '2',
                                 "<HS" = '3',
                                 "<HS" = '4',
                                 "<HS" = '5',
                                 'HS' = '6', # High school
                                 'TC' = '7', # Technical Certification
                                 'AD' = '8', # Associate's
                                 'BD' = '9', # Bachelor's
                                 'MD' = '10', # Master's
                                 'PD' = '11', # Professional (e.g. JD, RN)
                                 'PhD' = '12'),

```

```
         # Collapsing and clenaing DISABILITY variables
         ## for identifying mulitple disabilities
         DIS_SUM = rowSums(tibble(DIS_PHYS, DIS_COG, DIS_LEARN), na.rm = TRUE),

         ## recoding numbers into strings
         DIS_PHYS = fct_recode(as.factor(DIS_PHYS), physical = '1'),
         DIS_COG = fct_recode(as.factor(DIS_COG), cognitive = '1'),
         DIS_LEARN = fct_recode(as.factor(DIS_LEARN), learning = '1'),

         ## concatenating disability columns
         DIS_DETAILED =  str_trim(gsub("NA", "",
                                   paste(DIS_PHYS, DIS_COG, DIS_LEARN, sep = " "))),

         ## recode race variable as a single race vs. two or more races
         DISABILITY = ifelse(DIS_SUM < 2, DIS_DETAILED, "two or more disabilities"),

         ## recode all blanks as "NA"
         DISABILITY = fct_recode(factor(DISABILITY), "None" = ""),

         # Collapsing and cleaning MAJOR variables
         ## for identifying double+ majors
         MAJOR_SUM = rowSums(tibble(MAJ_ARTHUM, MAJ_SOCSCI, MAJ_BIOSCI, MAJ_PHYSCI,
                                    MAJ_ENGMATHCS, MAJ_EDUC, MAJ_BUSI, MAJ_JOURN,
                                    MAJ_OTHER, MAJ_UNDEC), na.rm = TRUE),

         # recode numbers into strings
         MAJ_ARTHUM = fct_recode(as.factor(MAJ_ARTHUM), arts_humanities = '1'),
         MAJ_SOCSCI = fct_recode(as.factor(MAJ_SOCSCI), social_sciences = '1'),
         MAJ_BIOSCI = fct_recode(as.factor(MAJ_BIOSCI), bio_sciences = '1'),
```

```r
         MAJ_PHYSCI = fct_recode(as.factor(MAJ_PHYSCI), physical_sciences = '1'),
         MAJ_ENGMATHCS = fct_recode(as.factor(MAJ_ENGMATHCS), engs_math_cs = '1'),
         MAJ_EDUC = fct_recode(as.factor(MAJ_EDUC), education = '1'),
         MAJ_BUSI = fct_recode(as.factor(MAJ_BUSI), business = '1'),
         MAJ_JOURN = fct_recode(as.factor(MAJ_JOURN), journalism = '1'),
         MAJ_OTHER = fct_recode(as.factor(MAJ_OTHER), other = '1'),
         MAJ_UNDEC = fct_recode(as.factor(MAJ_UNDEC), undecided = '1'),

         # concatenate major columns
         MAJOR_DETAILED =  str_trim(gsub("NA", "",
                                    paste(MAJ_ARTHUM, MAJ_SOCSCI, MAJ_BIOSCI,
                                          MAJ_PHYSCI, MAJ_ENGMATHCS,MAJ_EDUC, MAJ_BUSI,
                                          MAJ_JOURN, MAJ_OTHER, MAJ_UNDEC, sep = " "))),

         ## recode race variable as a single race vs. two or more races
         MAJOR = ifelse(MAJOR_SUM < 2, MAJOR_DETAILED, "two or more majors"),
         ## recode all blanks as "NA"
         MAJOR = fct_recode(factor(MAJOR), NULL = ""),

         # Collapsing and cleaning RACE variables
         ## for identifying multiracial folks
         # RACE_SUM = rowSums(tibble(RACEV1_AFRAM, RACEV1_ASIPAC, RACEV1_CAUC, RACEV1_HISPLAT,
         #                           RACEV1_NATAM, RACEV1_OTHER, RACEV2_BLAFR, #RACEV2_ASIPAC,
         #                           RACEV2_WHITE, RACEV2_HISPLAT, RACEV2_ARABME, RACEV2_NATAM,
         #                           RACEV2_OTHER, RACE_EASIA, RACE_SASIA, RACE_SEASIA, RACE_NHPI, RACE_OASIA),na.rm = TRUE),
         RACE_NURG_SUM = rowSums(tibble(RACEV1_ASIPAC, RACEV1_CAUC,RACEV2_WHITE, RACE_EASIA, RACE_SASIA),na.rm = TRUE),
```

```
          RACE_URG_SUM = rowSums(tibble(RACEV1_AFRAM,RACEV1_HISPLAT, RACEV1_NATAM, RACEV1_OTHER,
                                        RACEV2_BLAFR,  RACEV2_HISPLAT, RACEV2_ARABME, RACEV2_NATAM,
                                        RACEV2_OTHER, RACE_SEASIA, RACE_NHPI, RACE_OASIA),na.rm = TRUE),
          RACE_URGNURG_SUM = RACE_NURG_SUM + RACE_URG_SUM,

          ## recoding numbers into strings
          RACEV1_AFRAM = fct_recode(as.factor(RACEV1_AFRAM), black_afram = "1"),
          RACEV1_ASIPAC = fct_recode(as.factor(RACEV1_ASIPAC), asian = "1"),
          RACEV1_CAUC = fct_recode(as.factor(RACEV1_CAUC), white = "1"),
          RACEV1_HISPLAT = fct_recode(as.factor(RACEV1_HISPLAT), hisp_latino = "1"),
          RACEV1_NATAM = fct_recode(as.factor(RACEV1_NATAM), nat_american = "1"),
          RACEV1_OTHER = fct_recode(as.factor(RACEV1_OTHER), other = "1"),

          #RACEV2_ASIPAC = fct_recode(as.factor(RACEV2_ASIPAC), asian_pacific = "1"),
          RACEV2_BLAFR = fct_recode(as.factor(RACEV2_BLAFR), black_afram = "1"),
          RACEV2_WHITE = fct_recode(as.factor(RACEV2_WHITE), white = "1"),
          RACEV2_HISPLAT = fct_recode(as.factor(RACEV2_HISPLAT), hisp_latino = "1"),
          RACEV2_ARABME = fct_recode(as.factor(RACEV2_ARABME), arab_me = "1"),
          RACEV2_NATAM = fct_recode(as.factor(RACEV2_NATAM), nat_american = "1"),
          RACEV2_OTHER = fct_recode(as.factor(RACEV2_OTHER), other = "1"),

          RACE_EASIA = fct_recode(as.factor(RACE_EASIA), asian = "1"),
          RACE_SASIA = fct_recode(as.factor(RACE_SASIA), asian = "1"),
          RACE_SEASIA = fct_recode(as.factor(RACE_SEASIA), ur_asian = "1"),
```

```
            RACE_NHPI = fct_recode(as.factor(RACE_NHPI), ur_asian = "1"),
            RACE_OASIA = fct_recode(as.factor(RACE_OASIA), ur_asian = "1"),

            ## concatenate race columns
            RACE_DETAILED =  str_trim(gsub("NA", "",
                                        paste(RACEV1_AFRAM, RACEV1_ASIPAC, RACEV1_CAUC, RACEV1_HISPLAT,
                                              RACEV1_NATAM, RACEV1_OTHER,

                                              RACEV2_BLAFR, RACEV2_WHITE, RACEV2_HISPLAT, #RACEV2_ASIPAC,
                                              RACEV2_ARABME, RACEV2_NATAM, RACEV2_OTHER,

                                              RACE_EASIA, RACE_SASIA, RACE_SEASIA, RACE_NHPI, RACE_OASIA,

                                              sep = " "))),

            ## recode race variable as a single race vs. two or more races
            RACE = ifelse(RACE_URGNURG_SUM < 2, RACE_DETAILED, ifelse(RACE_URG_SUM < 2, "multiracial", "URG multiracial")),

            ## recode all blanks as "NA"
            RACE = fct_recode(factor(RACE), NULL = ""),

            #Further collapsing demographic columns into under-represented group and non-URG
            ## identifying URG
            # Collapsing demographic columns further into Under-represented group and non-URG
            RACE_URG = ifelse(RACE %in% URG_RACE, 'URG Race', ifelse(is.na(RACE), NA, 'Non-URG Race')),
            GENDER_URG = ifelse(GENDER %in% URG_GENDER, 'URG Gender', ifelse(is.na(GENDER), NA, 'Non-URG Gender')),
            DIS_URG = ifelse(DISABILITY %in% URG_DIS, 'URG Disability', ifelse(is.na(DISABILITY), NA, 'Non-URG Disability')),
```

~~~r
            EDU_URG = ifelse(MAXEDU %in% URG_EDU, 'URG Education',ifelse(is.na(MAXEDU), NA, 'Non-URG Education') ),
            EFF_STAT = ifelse(EFFORT %in% HIGH_EFF, 'High Effort',ifelse(is.na(EFFORT), NA, ifelse(EFFORT %in% LOW_EFF, 'Low Effort', 'Lapsed Effort')))
            )

#removing students with only PST scores/score = N
v4 <- v3[!is.na(v3$SCORE),]
```

# Standardizing Aff/Att variables
```{r}
schmath.mean <- mean(v4$PRE_SCHMATH, na.rm = T)
schmath.sd <- sd(v4$PRE_SCHMATH, na.rm = T)
numanx.mean <- mean(v4$PRE_NUMANX, na.rm = T)
numanx.sd <- sd(v4$PRE_NUMANX, na.rm = T)
selfeff.mean <- mean(v4$PRE_SELFEFF, na.rm = T)
selfeff.sd <- sd(v4$PRE_SELFEFF, na.rm = T)
dlmath.mean <- mean(v4$PRE_DLMATH, na.rm = T)
dlmath.sd <- sd(v4$PRE_DLMATH, na.rm = T)
mot.mean <- mean(v4$PRE_MOT, na.rm = T)
mot.sd <- sd(v4$PRE_MOT, na.rm = T)
numrel.mean <- mean(v4$PRE_NUMREL, na.rm = T)
numrel.sd <- sd(v4$PRE_NUMREL, na.rm = T)

sdz <- function(value, mean, sd){
                value.sdz <- (value - mean) / sd
                return(value.sdz)
}


v4$SCHMATH <- lapply(v4$PRE_SCHMATH, sdz, mean=schmath.mean, sd = schmath.sd)
v4$NUMANX <- lapply(v4$PRE_NUMANX, sdz, mean=numanx.mean, sd = numanx.sd)
v4$SELFEFF <- lapply(v4$PRE_SELFEFF, sdz, mean=selfeff.mean, sd = selfeff.sd)
v4$DLMATH <- lapply(v4$PRE_DLMATH, sdz, mean=dlmath.mean, sd = dlmath.sd)
v4$MOT <- lapply(v4$PRE_MOT, sdz, mean=mot.mean, sd = mot.sd)
~~~

```
v4$NUMREL <- lapply(v4$PRE_NUMREL, sdz, mean=numrel.mean, sd = numrel.sd)

v4$SCHMATH <- unlist(v4$SCHMATH)
v4$NUMANX <- unlist(v4$NUMANX)
v4$SELFEFF <- unlist(v4$SELFEFF)
v4$DLMATH <- unlist(v4$DLMATH)
v4$MOT <- unlist(v4$MOT)
v4$NUMREL <- unlist(v4$NUMREL)

```


#Export data

```{r}
#selecting only relevant columns

vars2 <- c("SCORE", "EFFORT",
           "NUMANX", "NUMREL", "SELFEFF", "SCHMATH", "MOT", "DLMATH",
           "GENDER", "MAXEDU",
           'DISABILITY', 'DIS_DETAILED',
           'MAJOR', 'MAJOR_DETAILED',
           'RACE', 'RACE_DETAILED',
           'RACE_URG', 'GENDER_URG', 'DIS_URG', 'EDU_URG', 'EFF_STAT'
          )

data <- dplyr::select(v4, vars2)

write.csv(data,'Master_Data_v5_race_edit_wrangled2.csv', row.names = F)
```

Collapsing column code graciously provided by Prof. Bailey
```
```

D.2 Factor Analysis

```
```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)

require(mosaic)
require(foreign)
require(tidyverse)
require(magrittr)
require(psych)
require(haven)
require(car)
require(dplyr)
require(paran)
options(max.print=1000000)
```

Import data

```{r}
full = read.csv("Master_data_EFA_cut.csv")

# Filling in NaN values of binaries with 0
## vector with binary variables

binary <- c("PRE_WHYMAJ_1","PRE_WHYMAJ_2","PRE_WHYMAJ_3","PRE_WHYMAJ_4",
         "PRE_WHYMAJ_5","PRE_WHYMAJ_6","PRE_WHYMAJ_7","PRE_WHYMAJ_8",
         'PRE_WHYCS_1','PRE_WHYCS_2','PRE_WHYCS_3','PRE_WHYCS_4','PRE_WHYCS_5
    ','PRE_WHYCS_6',
         'PRE_WHYCS_7' )

full[,binary][is.na(full[,binary])] <- 0

#selecting relevant variables
allvars <- c('PRE_ATT_SC_1', 'PRE_ATT_SC_2',
              'PRE_ATT_SC_3','PRE_ATT_SC_4','PRE_ATT_SC_5',
```
```

````

              'PRE_ATT_DL_1','PRE_ATT_DL_2','PRE_ATT_DL_3',
              'PRE_ATT_DL_4','PRE_ATT_DL_5',

              'PRE_CF_MEAN', 'PRE_LK1', 'PRE_LK2', 'PRE_LK5',

              'PRE_DAILYM', 'PRE_DAILYG','PRE_FREQEN',

              'PRE_ANX.1_1','PRE_ANX.1_2','PRE_ANX.1_3','PRE_ANX.1_4',

              'PRE_WHYMAJ_1', 'PRE_WHYMAJ_2','PRE_WHYMAJ_3', 'PRE_WHYMAJ_4',
              'PRE_WHYMAJ_5', 'PRE_WHYMAJ_6', 'PRE_WHYMAJ_7','PRE_WHYMAJ_8',

              'PRE_WHYCS_1',
              'PRE_WHYCS_2',
              'PRE_WHYCS_3', 'PRE_WHYCS_4',
              'PRE_WHYCS_5', 'PRE_WHYCS_6',
              'PRE_WHYCS_7'
          )

data <- full[,c(allvars)]

# cleaning
p.clean <- data[complete.cases(data),]

```

#correlations histogram
```{r, fig.height = 2, fig.width = 3}
cor<- as.data.frame(cor(p.clean))
cord <- data.frame(cor[1])
for (i in 2:36){
    df <- data.frame(PRE_ATT_SC_1 = cor[,i], check.names = F)
    cord <- bind_rows(cord, df)
}

f4 <- ggplot(cord, mapping = aes(PRE_ATT_SC_1, )) +
````

```
            geom_histogram(data = subset(cord, cord != 1), stat = 'bin',
    binwidth = .1, fill = 'darkgreen', alpha = 0.5) +
            xlab('Correlation coefficient')
f4
```

Principle Component Analysis
Scree Plot
```{r}
Scree_Plot <- princomp(p.clean, cor=TRUE)
plot(Scree_Plot,type="lines", ylim = c(1,7))
summary(Scree_Plot)
```

Parallel Analysis
```{r}
paran(p.clean, iterations = 1000, centile = 0, quietly = FALSE,
    status = TRUE, all = TRUE, cfa = TRUE, graph = TRUE, color = TRUE,
    col = c("black", "red", "blue"), lty = c(1, 2, 3), lwd = 1, legend = TRUE
    ,
    file = "", width = 640, height = 640, grdevice = "png", seed = 0)
```


EFA
Select variables of interest + cleaning
```{r}
#selecting relevant variables for generating factor loadings
## variables that did not load or cross loaded are commented out

vars <- c('PRE_ATT_SC_1',
          # 'PRE_ATT_SC_2', 'PRE_ATT_SC_3',
          'PRE_ATT_SC_4','PRE_ATT_SC_5',

          'PRE_ANX.1_1','PRE_ANX.1_2','PRE_ANX.1_3','PRE_ANX.1_4',

          'PRE_LK1', 'PRE_LK2', 'PRE_LK5',
```

```
          # 'PRE_CF_MEAN',

          'PRE_ATT_DL_1',
          # 'PRE_ATT_DL_2','PRE_ATT_DL_3',
          'PRE_ATT_DL_4','PRE_ATT_DL_5',

          'PRE_WHYCS_3','PRE_WHYCS_5', 'PRE_WHYCS_6',
          # 'PRE_WHYCS_1', 'PRE_WHYCS_2', 'PRE_WHYCS_4', 'PRE_WHYCS_7'

          'PRE_DAILYM', 'PRE_DAILYG','PRE_FREQEN'

          # 'PRE_WHYMAJ_1','PRE_WHYMAJ_2','PRE_WHYMAJ_3', 'PRE_WHYMAJ_4',
          # 'PRE_WHYMAJ_5', 'PRE_WHYMAJ_6', 'PRE_WHYMAJ_7','PRE_WHYMAJ_8',

          )

efa.clean <- p.clean[,vars]

```

Factor Analysis (Spearman)
```{r}
## Generating a correlation matrix
cor <- cor(efa.clean, use="complete.obs", method = "spearman")

## Oblimin Factor Analysis with Promax rotation
factorsol <-  fa(r=cor, nfactors = 6,fm="ml", rotate ="promax",  oblique.
    scores = TRUE, max.iter = 25)

print.psych(factorsol, cut = 0.30 ,sort = T)
fa2latex(factorsol)
```

Export

```{r}
# Extracting the loadings into a table for export
```

````r
loadings <- factorsol$loadings[c(1:ncol(efa.clean)),]
loadings <- apply(loadings, 2, function(x) ifelse(abs(x)>0.4,x, vector()))

#export factor loadings for use in Python
write.csv(loadings,'factorloadings_v5lab.csv', row.names = T)
```
````

Listing D.1: EFA, PCA, and Parallel Analysis

````r
```{r setup, include=FALSE}

require(mosaic)
require(psych)
require(haven)
require(lavaan) # for CFA

```

# Import data

```{r}
full = read.csv("Master_data_CFA_cut.csv")
```

# Cleaning
```{r}
# Filling in NaN values of binaries with 0
## vector with binary variables

binary <- c("PRE_WHYMAJ_1","PRE_WHYMAJ_2","PRE_WHYMAJ_3","PRE_WHYMAJ_4",
         "PRE_WHYMAJ_5","PRE_WHYMAJ_6","PRE_WHYMAJ_7","PRE_WHYMAJ_8",
         'PRE_WHYCS_1','PRE_WHYCS_2','PRE_WHYCS_3','PRE_WHYCS_4','PRE_WHYCS_5','PRE_WHYCS_6',
         'PRE_WHYCS_7' )

full[,binary][is.na(full[,binary])] <- 0
````

```

#selecting relevant variables

vars <- c('PRE_ATT_SC_1',
          # 'PRE_ATT_SC_2', 'PRE_ATT_SC_3',
          'PRE_ATT_SC_4','PRE_ATT_SC_5',

          'PRE_ANX.1_1','PRE_ANX.1_2','PRE_ANX.1_3','PRE_ANX.1_4',

          'PRE_LK1', 'PRE_LK2', 'PRE_LK5',
          # 'PRE_CF_MEAN',

          'PRE_ATT_DL_1',
          # 'PRE_ATT_DL_2','PRE_ATT_DL_3',
          'PRE_ATT_DL_4','PRE_ATT_DL_5',

          'PRE_WHYCS_3','PRE_WHYCS_5', 'PRE_WHYCS_6',
          # 'PRE_WHYCS_1', 'PRE_WHYCS_2', 'PRE_WHYCS_4', 'PRE_WHYCS_7'

          'PRE_DAILYM', 'PRE_DAILYG','PRE_FREQEN'

          # 'PRE_WHYMAJ_1','PRE_WHYMAJ_2','PRE_WHYMAJ_3', 'PRE_WHYMAJ_4',
          # 'PRE_WHYMAJ_5', 'PRE_WHYMAJ_6', 'PRE_WHYMAJ_7','PRE_WHYMAJ_8',

          )

# selecting relevant variables
data <- full[,c(vars)]

# Cleaning
d.clean <- data[complete.cases(data),]

```

```{r}
# creating factor model

```

```
cfa.model <- 'schmath =~ PRE_ATT_SC_1 + PRE_ATT_SC_4 + PRE_ATT_SC_5
              selfeff =~ PRE_LK1 + PRE_LK2 + PRE_LK5
              numanx =~ PRE_ANX.1_1 + PRE_ANX.1_2  + PRE_ANX.1_3  + PRE_ANX.1
    _4
              numrel =~ PRE_DAILYM + PRE_DAILYG + PRE_FREQEN
              expect =~ PRE_WHYCS_3 +  PRE_WHYCS_5 + PRE_WHYCS_6
              dlmath =~ PRE_ATT_DL_1 + PRE_ATT_DL_4 + PRE_ATT_DL_5'

# fitting model
fit <- cfa(cfa.model, d.clean)
summary(fit, fit.measures=T)

semPlot::semPaths(fit, "std")

```

Listing D.2: CFA

```
import pandas as pd
import numpy as np

def factor_analysis(data, semester, year, season):

    FAloadings_v1 = pd.read_csv("factorloadings_v1.csv")
    FAloadings_v5 = pd.read_csv("Factorloadings_v5.csv")

    ## Excluded some code to minimize length ##

            ld_att_sc1 = FAloadings_v5['ML2'][0]
            ld_att_sc4 = FAloadings_v5['ML2'][1]
            ld_att_sc5 = FAloadings_v5['ML2'][2]

            ld_anx1 = FAloadings_v5['ML5'][3]
            ld_anx2 = FAloadings_v5['ML5'][4]
            ld_anx3 = FAloadings_v5['ML5'][5]
            ld_anx4 = FAloadings_v5['ML5'][6]
```

```
        ld_lk1 = FAloadings_v5['ML1'][7]
        ld_lk2 = FAloadings_v5['ML1'][8]
        ld_lk5 = FAloadings_v5['ML1'][9]

        ld_att_dl1 = FAloadings_v5['ML6'][10]
        ld_att_dl4 = FAloadings_v5['ML6'][11]
        ld_att_dl5 = FAloadings_v5['ML6'][12]

        ld_dailym = FAloadings_v5['ML4'][16]
        ld_dailyg = FAloadings_v5['ML4'][17]
        ld_freqen = FAloadings_v5['ML4'][18]

        #assigning variables involved in calculating factor scores from student responses
        range = np.arange(0, len(data), 1)
        for i in range:

            att_sc1 = data[semester +'_ATT_SC_1'][i]
            att_sc4 = data[semester +'_ATT_SC_4'][i]
            att_sc5 = data[semester +'_ATT_SC_5'][i]

            lk1 = data[semester +'_LK1'][i]
            lk2 = data[semester +'_LK2'][i]
            lk5 = data[semester +'_LK5'][i]

            anx1 = data[semester +'_ANX#1_1'][i]
            anx2 = data[semester +'_ANX#1_2'][i]
            anx3 = data[semester +'_ANX#1_3'][i]
            anx4 = data[semester +'_ANX#1_4'][i]

            dailym = data[semester +'_DAILYM'][i]
            dailyg = data[semester +'_DAILYG'][i]
            freqen = data[semester +'_FREQEN'][i]

            att_dl1 = data[semester +'_ATT_DL_1'][i]
            att_dl4 = data[semester +'_ATT_DL_4'][i]
```

```
            att_dl5 = data[semester +'_ATT_DL_5'][i]

            # checking for nan values, then calculating factor score

            ## setting preliminary variables
            schmath = 0
            dlmath = 0
            selfeff = 0
            numrel = 0
            numanx = 0

            ## calculating school math score
            if pd.isnull(att_sc1) or pd.isnull(att_sc4) or pd.isnull(att_sc5):
                schmath = np.nan
            else:
                schmath = ld_att_sc1*att_sc1 + ld_att_sc4*att_sc4 + ld_att_sc5*att_sc5

            ## calculating numerical self efficacy score
            if pd.isnull(lk1) or pd.isnull(lk2) or pd.isnull(lk5):
                selfeff = np.nan
            else:
                selfeff = ld_lk1*lk1 + ld_lk2*lk2 + ld_lk5*lk5

            ## calculating math anxiety score
            if pd.isnull(anx1) or pd.isnull(anx2) or pd.isnull(anx3) or pd.isnull(anx4):
                numanx = np.nan
            else:
                numanx = ld_anx1*anx1 + ld_anx2*anx2 + ld_anx3*anx3 + ld_anx4*anx4

            ## calculating numerical relevancy score
            if pd.isnull(dailym) or pd.isnull(dailyg) or pd.isnull(freqen):
                numrel = np.nan
```

```
            else:
                numrel = ld_dailym*dailym + ld_dailyg*dailyg + ld_freqen*
    freqen

            ## calculating daily life math score
            if pd.isnull(att_dl1) or pd.isnull(att_dl4) or pd.isnull(
    att_dl5):
                dlmath = np.nan
            else:
                dlmath = ld_att_dl1*att_dl1 + ld_att_dl4*att_dl4 +
    ld_att_dl5*att_dl5

            ## filling in data with factor scores
            data.loc[i, semester +'_SCHMATH'] = schmath
            data.loc[i, semester +'_DLMATH'] = dlmath
            data.loc[i, semester +'_SELFEFF'] = selfeff
            data.loc[i, semester +'_NUMREL'] = numrel
            data.loc[i, semester +'_NUMANX'] = numanx

    return data
```

Listing D.3: Calculating Factor Scores

D.3 Mediation Analysis

```
mr_schmath.gen0 <- mediate(schmath_gen.lm0, score_schmath_gen.lm0, treat='
    GENDER_URG', mediator='SCHMATH', boot=TRUE, sims=100)
summary(mr_schmath.gen0)
```

Listing D.4: Mediation Code Example

Bibliography

[1] P. Lockhart, *A Mathematician's Lament*.

[2] L. Steen, "Mathematics and Democracy: The Case for Quantitative Literacy," Mathematics and Computer Education **36** (2002).

[3] P. Cohen, *A Calculating People: The Spread of Numeracy in Early America* (Routledge, 1999), ISBN 9780415925785, URL `https://books.google.com/books?id=SpyYKC-pNGoC`.

[4] G. Meisels, "Science Literacy: Hand in Glove with Numeracy," Numeracy **3**, 2 (2010).

[5] K. Skagerlund, T. Lind, C. Strömbäck, G. Tinghög, and D. Västfjäll, "Financial literacy and the role of numeracy–How individuals' attitude and affinity with numbers influence financial literacy," Journal of Behavioral and Experimental Economics **74**, 18 (2018), ISSN 22148051, URL `https://doi.org/10.1016/j.socec.2018.03.004`.

[6] R. L. Rothman, V. M. Montori, A. Cherrington, and M. P. Pignone, "The Role of Numeracy in Health Care," Journal of Health Communication **13**, 37 (2008).

[7] E. Peters, D. Västfjäll, P. Slovic, C. K. Mertz, K. Mazzocco, and S. Dickert, "Numeracy and decision making," Psychological Science **17**, 407 (2006), ISSN 09567976.

[8] D. Widdowson and K. Hailwood, "Financial literacy and its role in promoting a sound financial system," Reserve Bank of New Zealand Bulletin **70** (2007).

[9] J. Almenberg and O. Widmark, "Numeracy, Financial Literacy and Participation in Asset Markets," SSRN Electronic Journal pp. 1–40 (2012).

[10] M. Kankaraš, G. Montt, M. Paccagnella, G. Quintini, and W. Thorn, *Skills Matter: Further Results from the Survey of Adult Skills. OECD Skills Studies* (2016), ISBN 9789264258051.

[11] S. Parsons and J. Bynner, "Does Numeracy Matter More ?" National Research and Development Centre for Adult Literacy and Numeracy pp. 1–37 (2005).

[12] M. F. Charette and R. Meng, "The Determinants of Literacy and Numeracy , and the Effect of Literacy and Numeracy on Labour Market Outcomes Author (s): Michael F . Charette and Ronald Meng Source : The Canadian Journal of Economics / Revue canadienne d ' Economique , Vol . 31 , No ." **31**, 495 (2016).

[13] R. Nadeau, R. G. Niemi, and J. Levine, "Innumeracy About Minority Populations," The Public Opinion Quarterly **57**, 332 (1993), ISSN 0033362X, 15375331, URL http://www.jstor.org/stable/2749094.

[14] J. C. Coronel, S. Poulsen, and M. D. Sweitzer, "Investigating the generation and spread of numerical misinformation : A combined eye movement monitoring and social transmission approach," **00**, 1 (2019).

[15] M. Planty, W. Hussar, T. Snyder, and G. Kena, "The Condition of education 2007," (2019), URL `http://www.voced.edu.au/content/ngv10426`.

[16] M. Kutner, E. Greenberg, Y. Jin, B. Boyle, Y.-c. Hsu, and E. Dunleavy, "Literacy in Everyday Life: Results from the 2003 National Assessment of Adult Literacy. NCES 2007-490," National Center for Education Statistics (2007).

[17] S. Ansell, "Achievement Gap," (2011), URL `https://www.edweek.org/ew/issues/achievement-gap/index.html`.

[18] M. Kuhfeld, E. Gershoff, and K. Paschall, "The development of racial/ethnic and socioeconomic achievement gaps during the school years," Journal of Applied Developmental Psychology **57**, 62 (2018), ISSN 01933973.

[19] URL `https://www.nationsreportcard.gov/ndecore/landing`.

[20] J. Lee, "Racial and Ethnic Achievement Gap Trends: Reversing the Progress Toward Equity?" Educational Researcher **31**, 3 (2002), ISSN 0013189X.

[21] W. T. Dickens, C. L. Schultze, and T. J. Kane, "Does The Bell Curve Ring True? A Closer Look at a Grim Portrait of American Soci-

ety," Brookings (2016), URL http://www.brookings.edu/articles/does-the-bell-curve-ring-true-a-closer-look-at-a-grim-portrait-of-american-so

[22] R. Nisbett, *Race, Genetics, and IQ* (Brookings, 1998), p. 86–102.

[23] G. Ladson-Billings, "From the Achievement Gap to the Education Debt: Understanding Achievement in U.S. Schools," Educational Researcher **35**, 3 (2006), ISSN 0013189X, 1935102X, URL http://www.jstor.org/stable/3876731.

[24] J. Marsh, "Neighborhood Poverty and Household Financial Security," (2016).

[25] A. Flores, "Examining Disparities in Mathematics Education: Achievement Gap or Opportunity Gap?" The High School Journal **91**, 29 (2007), ISSN 1534-5157.

[26] A. Cilluffo and R. Kochhar, "Income Inequality in the U.S. is Rising Most Rapidly Among Asians," Pew Research Center (2018).

[27] S. Kotok, "Unfulfilled Potential: High-Achieving Minority Students and the High School Achievement Gap in Math," The High School Journal **100**, 183 (2017), ISSN 1534-5157.

[28] G. Clore and S. Schnall, "The Influence of Affect on Attitude." (2005).

[29] "APA Dictionary of Psychology," URL https://dictionary.apa.org/affect.

[30] "APA Dictionary of Psychology," URL https://dictionary.apa.org/affect.

[31] S. Grigg, H. N. Perera, P. McIlveen, and Z. Svetleff, "Relations among math self efficacy, interest, intentions, and achievement: A social cognitive perspective," Contemporary Educational Psychology **53**, 73 (2018), ISSN 10902384.

[32] "Considering the role of affect in learning: Monitoring students' self-efficacy, sense of belonging, and science identity," CBE Life Sciences Education **13**, 6 (2014), ISSN 19317913.

[33] L. J. Sax, "Mathematical self-concept: How college reinforces the gender gap," Research in Higher Education **35**, 141 (1994), ISSN 1573-188X, URL https://doi.org/10.1007/BF02496699.

[34] "Math anxiety: Personal, educational, and cognitive consequences," Current Directions in Psychological Science **11**, 181 (2002), ISSN 09637214.

[35] N. Durrani and V. Tariq, "Relationships between undergraduates' mathematics anxiety and their attitudes towards developing numeracy skills and perceptions of numerical competence," pp. 787–794 (2009), URL http://library.iated.org/view/DURRANI2009REL.

[36] J. Lee, "Universals and specifics of math self-concept, math self-efficacy, and math anxiety across 41 PISA 2003 participating countries," Learning and Individual Differences **19**, 355 (2009), ISSN 1041-6080,

large-scale Cross-cultural Studies of Cognitive and Noncognitive Constructs, URL `http://www.sciencedirect.com/science/article/pii/S104160800800112X`.

[37] A. Devine, K. Fawcett, D. Szucs, and A. Dowker, “Gender differences in mathematics anxiety and the relation to mathematics performance while controlling for test anxiety,” Behavioral and Brain Functions **8**, 1 (2012), ISSN 17449081.

[38] S. A. Hart and C. M. Ganley, “The nature of math anxiety in adults: Prevalence and correlates,” Journal of Numerical Cognition **5**, 122 (2019).

[39] K. Follette, D. McCarthy, E. Dokter, S. Buxner, and E. Prather, “The Quantitative Reasoning for College Science (QuaRCS) Assessment, 1: Development and Validation,” Numeracy **8** (2015).

[40] K. Follette, S. Buxner, E. Dokter, D. McCarthy, B. Vezino, L. Brock, and E. Prather, “The Quantitative Reasoning for College Science (QuaRCS) Assessment 2: Demographic, Academic and Attitudinal Variables as Predictors of Quantitative Ability,” Numeracy **10** (2017).

[41] A. B. Frymier and G. M. Shulman, ““What’s in it for me?”: Increasing content relevance to enhance students’ motivation,” Communication Education **44**, 40 (1995), `https://doi.org/10.1080/03634529509378996`, URL `https://doi.org/10.1080/03634529509378996`.

[42] A. Wigfield and J. S. Eccles, “Expectancy-value theory of achievement

motivation," Contemporary Educational Psychology **25**, 68 (2000), ISSN 0361476X.

[43] *Statistical Abstract of the United States, 2011* (U.S. Census Bureau, 2011).

[44] "Postsecondary Attainment: Differences by Socioeconomic Status," (2015), URL https://nces.ed.gov/programs/coe/indicator_tva.asp.

[45] "A Practical Introduction to Factor Analysis: Exploratory Factor Analysis," URL https://stats.idre.ucla.edu/spss/seminars/introduction-to-factor-analysis/a-practical-introduction-to-factor-analysis/.

[46] M. Matsunaga, "How to Factor-Analyze Your Data Right: Do's Don'ts, and How-To's," International Journal of Psychological Research, ISSN 2011-7922, Vol. 3, N^{o}. 1, 2010, pags. 97-110 **3** (2010).

[47] J. Nerbonne, "PDF," URL https://www.let.rug.nl/nerbonne/teach/rema-stats-meth-seminar/Factor-Analysis-Kootstra-04.PDF.

[48] "Determining the Number of Factors with Parallel Analysis in R," (2016), URL http://equastat.com/determining-number-factors-parallel-analysis-r/.

[49] R. Hoffman, "Indices of Complexity and InsrpretatiOn: Their Computation and Uses in Factor Analysis," (1973), URL https://files.eric.ed.gov/fulltext/ED074155.pdf.

[50] A. Yong and S. Pearce, "A Beginner's Guide to Factor Analysis: Focusing on Exploratory Factor Analysis," Tutorials in Quantitative Methods for Psychology **9**, 79 (2013).

[51] D. Kenny, "Measuring Model Fit," URL http://www.davidakenny.net/cm/fit.htm.

[52] "A Practical Introduction to Confirmatory Factor Analysis," IDRE Statistical Consulting URL https://stats.idre.ucla.edu/spss/seminars/introduction-to-factor-analysis/a-practical-introduction-to-factor-analysis-confirmatory-factor-analysis/.

[53] L. L. Havlicek and N. L. Peterson, "Robustness of the T Test: A Guide for Researchers on Effect of Violations of Assumptions," Psychological Reports **34**, 1095 (1974), https://doi.org/10.2466/pr0.1974.34.3c.1095, URL https://doi.org/10.2466/pr0.1974.34.3c.1095.

[54] B. Kim, "Introduction to Mediation Analysis," (2016), URL https://data.library.virginia.edu/introduction-to-mediation-analysis/.

[55] A. J. Fairchild and D. P. MacKinnon, "A general model for testing mediation and moderation effects," Prevention science : the official journal of the Society for Prevention Research **10**, 87 (2009), ISSN 1573-6695, URL https://pubmed.ncbi.nlm.nih.gov/19003535https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2908713/.

[56] A. L. Coleman, "What Is Intersectionality? A Brief His-

tory of the Theory," (2019), URL https://time.com/5560575/intersectionality-theory/.