GROUP TESTING AND SENSE OF BELONGING IN REFORM-BASED CALCULUS: EQUITABLE FOR ALL WOMEN?

by

Candice M. Quinn

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Mathematics and Science Education

Middle Tennessee State University

July 22, 2021

Dissertation Committee:

Dr. Jennifer J. Kaplan

Dr. Susan D. Carver

Dr. James Hart

Dr. Katherine Mangione

Dr. Amy J. Phelps

DEDICATION

This work is dedicated to all of those who *try*.

I want you to know I see you; I hear you; and I value you more than you will ever know.

I have tried. I have tried to be what you wanted me to be. I have tried to conform to your needs. I have tried to fit into your scheme. I have tried to maintain your status quo.

All the while, I have tried.

I have tried to maintain my individuality. I have tried to be myself in your world. I have tried to find a place where I can fit into your scheme. I have tried to fight for equity in your status quo.

I have tried, though I have failed.

I have failed. I have failed to find my place. I have failed to meet your needs. I have failed to meet my own needs within your scheme. I have failed to be recognized for my fight finding a place in your world.

I have failed to belong, to be noticed, to be valued, to be heard. I have failed, but at least I have tried.

ACKNOWLEDGEMENTS

First and most importantly, I want to thank my parents, Bob and Kathy Quinn for their support throughout this tough journey. You two gave me the strength to leave Ohio, and to believe in myself. Mom, I would not have survived these last years without our (almost) weekly video calls. Even if we went a week or two without talking, just knowing you were there for me emotionally helped me persevere and finally reach this finish line. Dad, your encouragement and support especially through the dark parts of this journey helped to keep me strong. Thank you both for supporting me in whatever I attempt and for never letting me give up on myself. I love you both bushels and baskets full.

Secondly, to my siblings, Becky Bryda and Michael Quinn, thank you for reminding me to have fun once in a while as I was on this journey. You two are some of the smartest people I know, and you should know I never thought I could hold a candle to either of you. Thank you for believing in me ever since I started college. You both treated me as equals even when I did not believe in myself, all the way back in my undergraduate years. I love and miss you both very much.

Thank you to my wonderful committee, Dr. Kaplan, Dr. Carver, Dr. Hart, Dr. Mangione, and Dr. Phelps. Thank you for not only pushing my work forward with your insight, but for also appreciating and encouraging my future work after seeing the results. All of your insight helped to make this document a masterpiece in which I am proud.

I would like to give a special thanks to Dr. Kaplan, this document would not exist without you! Your wizardry and wordsmithing helped this story come to life. You truly did help me create a document that *I am proud* to call mine. You were able to see through my dyslexic writing tendencies in ways no one else could, and helped me craft the most powerful sentences. You will always be more than my chair. You are also my confidant, my pandemic buddy, and my friend. I do not know if I could ever thank you enough for your emotional help through the PTSD that I had from the past two years. I know we had our moments, but I always felt you truly believed in me and my work and I thank you for that support. You truly are a special gift to the MSE program.

I would like to give another special thanks to Dr. Joshua Reid. You never let me give up on myself and I thank you. Our journey together throughout the past five years has been amazing. I know you will go much farther in this academic world, and I cannot wait to see what you accomplish. Just know both me and McQuiddy support you and Jocelyn's endeavors, whatever they may be.

To my peers at MTSU, thank you for helping me feel like I belonging in this program, especially Drs. Lucy Watson, Zhigang "Jacob" Jia, Kristen Hartland, Angela Google, and Sam Reed. Thank you all for giving me a sense of family throughout these rough years and for the great conversations, drinks, tears, and cheers.

To the OpSTEM team and my wonderful friends and colleagues at CSU, especially Dr. John P. Holcomb, Dr. Susan D. Carver, and Dr. Ivan Soprunov. You three have given me so much strength to believe in myself and I do not think I can ever express how thankful I am for each and every one of you.

Finally, a special thanks to my boys: Juan Amador, Chris Tuttle, Justin Simmerer, Marcus Jones, and Joshua Jucutan. You boys are the reason I started down this path, and the reason I do what I do so passionately. You have helped me discover my purpose through our time together. You are all remarkable young men, and I know you all have amazing futures in front of you. I love you all so dearly, and miss you every day.

ABSTRACT

There continues to be a national concern about the shortage of science, technology, engineering, and mathematics (STEM) graduates needed to fill current and anticipated STEM workforce positions. There is an additional concern regarding attracting and maintaining women in the STEM disciplines. In this study, I investigated one part of the STEM workforce problem by focusing on women in Calculus I courses who intend to major in a STEM discipline. Calculus presents a unique leak in the STEM pipeline for women because women who perform poorly in calculus are more likely to leave STEM fields than men who perform poorly in calculus. The main focus of this study is women's sense of belonging in calculus, as sense of belonging is one reason women leave the historically male-dominated STEM fields. Specifically, this study investigates the communication women have with their peers in a group test environment and how these interactions may influence the women's sense of belonging to a mathematics community. The main purpose of my study was to examine the social interactions of the group members during a group test, the differences in the interactions for groups with different compositions, and how the interactions might explain changes in the women's sense of belonging to mathematics in a reform Calculus I course. Women's communication around a group testing activity in Calculus I is significant because it has the potential to change assessment practices and promote retention of women in STEM by increasing women's collaboration and sense of belonging in a STEM community in a course required of all STEM majors. I used an exploratory descriptive mixed methods design to investigate female students' sense of belonging and their social interactions with group members in a Calculus I reform class. The results of the study revealed: (1) white women performed better when tested individually than in groups, (2) students who are under-represented minorities by race performed better when tested in groups than individually, (3) groups with more women had more equitable social interactions around the mathematics, (4) groups with only women felt more frustrated and asked more questions around the mathematics, (5) when studying communication and sense of belonging in mathematics, researchers may need to consider the intersectionality of race and gender.

Table of Contents

List of Tables

List of Figures

CHAPTER ONE: INTRODUCTION

Introduction and Rationale

There continues to be a national concern regarding the shortage of science, technology, engineering, and mathematics (STEM) graduates to fill current and anticipated STEM workforce positions (National Science Foundation, 2017; President's Council of Advisors on Science and Technology [PCAST], 2012; United States Department of Education, 2016). Many educational programs have been created to help reduce the deficit in the workforce by increasing student interest in STEM (Honey et al., 2014; Roberts et al., 2018; United States Department of the Interior, 2013) and retention rates in STEM at the university level (Bénéteau et al., 2016; Carver et al., 2017; Gayles & Ampaw, 2016). Despite the push for promoting STEM education, fewer than half of the students who enter college with the intent of obtaining a degree in STEM fields are successful (Chen, 2013; Chen 2015; PCAST, 2012; Young et al., 2011). Additionally, the shortage of women in many STEM fields is a notable concern (Beede et al., 2011; Blackburn, 2017; Chavatzia, 2017; Diekman et al., 2015; Hewlett et al., 2014) as they consist of less than 25% of the STEM workforce in the United States (Beede et al., 2011; Diekman et al., 2015).

In this study, I investigated one part of the STEM workforce problem by focusing on women in Calculus I courses who intend to major in a STEM discipline. In this chapter, I provide the rationale for the study by explaining *why* I focus on three main areas: women, calculus, and group testing. After I address each focus, I provide the reader with the purpose of the study, which includes my three research questions, the significance of the study, and definitions of key terms that will be used throughout this document.

Why Women?

Although women are being encouraged to pursue STEM degrees (Corbett & Hill, 2015), and currently do represent a majority of the workforce in biology and life sciences, they are still largely underrepresented in mathematically-intensive STEM fields such as computer science, engineering, mathematics, and physics (Cheryan et al., 2017). Historically, currently, and consistently these mathematically-intensive STEM fields are dominated by men (Cheryan et al., 2017; Hill et al., 2010). Additionally, there is a retention deficit of the women who do enter the STEM workforce; nearly one-third of women who enter a STEM field in the United States indicated intentions of leaving their position within the first year of employment due to a feeling of isolation (Hewlett et al., 2014). Overall, women are less likely to enter the STEM workforce and are more likely to leave once there (Chavatzia, 2017). This gender gap in the STEM workforce helps reinforce the stereotype of the STEM fields as male domains. Reinforcement of this stereotype can lead women who do enter STEM fields feeling isolated in their field of choice (Hewlett et al., 2014; O'Brien et al., 2015). This feeling of isolation as well can occur for women when taking courses in their major long before entering the workforce (Ellis et al., 2016). One such course is Calculus I, in which women's sense of belonging is a main contributor toward their desire to continue studying mathematics (Good et al., 2012).

A feeling of isolation, or a sense of alienation, can be described as a lack of feeling that one belongs (i.e., sense of belonging; Strayhorn, 2019), which is an essential component that fosters interest and motivation for women in STEM (Diekman et al., 2015; Good et al., 2012; Murphy et al., 2007; Walton & Cohen, 2007; Walton et al., 2012). This feeling of not belonging can occur for women who major in a STEM discipline when a majority of students in their mathematics courses are men. Essentially, this lack of sense of belonging for women decreases their interest in pursuing STEM fields (Diekman et al., 2015; Good et al., 2012; Murphy et al, 2007; Walton & Cohen, 2007; Walton et al., 2012), performance in mathematics courses (Good et al., 2012), and interest in working with others in a collaborative environment (Diekman et al., 2015). The decreased sense of belonging may also lead to a lack of communication. That is, if someone does not feel they are a valued member of the group (i.e., that they belong), why would they continue to offer ideas, thoughts, and knowledge with that group? Understanding that one's sense of belonging impacts how a student communicates in certain social settings may help educators understand how to alleviate women's lower perceived performance and confidence in mathematics. Therefore, this study aims to understand the link between women's sense of belonging and the nature of their social interactions that occur within a group setting during their Calculus I course. To understand this link, I studied communication among students while taking a group test.

Why Calculus?

There are many leaks in the STEM pipeline[1] (see Figure 1) specific to women and start as early as middle school. One of the first leaks, associated with middle school, is attributed to stereotypical gender roles. Young women can be detoured from pursuing

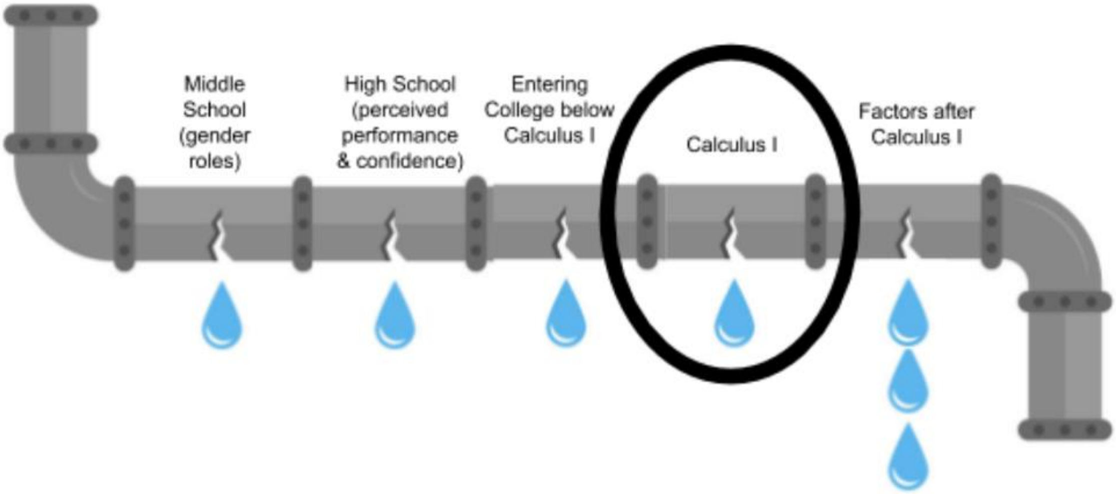
---

[1] Although I have recently argued that we need to reconsider the STEM pipeline metaphor (Quinn et al., 2020), I maintain this metaphor currently is still best for noticing the "leaks" in which women are diverted from reaching the STEM workforce.

STEM fields in middle school because stereotypes of gender roles are still a prominent influence on students' future career aspirations (Shapiro et al., 2015). One common stereotype that affects women's desire to enter STEM fields is that men are better at mathematics than women (Cadaret et al., 2017; Spencer et al., 1999). This stereotype can diminish women's confidence and performance in mathematics and science courses starting in high school (Blickenstaff, 2005; Good et al., 2008). Young women who maintain interest in STEM past middle school, but have mathematical skills below the Calculus I level by the time they enter college, have additional hurdles to overcome. The first hurdle for women who enter college below the Calculus I level is the need to take lower-level prerequisite mathematics courses in order to get to their required calculus courses. This additional hurdle is another leak in the pipeline because students who enter college below the Calculus I level are more likely to stop pursuing a STEM degree than students who enter at the Calculus I level (Bressoud, 2015; Ellis et al., 2016; Good et al., 2012; Sanabria & Penner, 2017; Shapiro et al., 2015). The next hurdle is passing Calculus I, a course that has a historically low pass rate nationally (i.e., under 60%; Bressoud, 2015). The final hurdle is maintaining interest in STEM after passing Calculus I (Ellis et al., 2016; Rasmussen & Ellis, 2013; Sanabria & Penner, 2017). A particularly interesting area to study with regard to women in STEM is Calculus I, as this course is one of the first places women leave the STEM pipeline *after* initially intending to earn a degree in a STEM field (Ellis et al., 2016; Sanabria & Penner, 2017).

Figure 1.

The Leaky STEM Pipeline for Women.



Although Calculus I is considered a gatekeeper for both male and female STEM majors, failure in Calculus I does not deter male students from their original intentions in the same way it does female students (Sanabria & Penner, 2017). Recently Sanabria and Penner (2017) examined the data from the National Educational Longitudinal Study (NELS) Postsecondary Transcript Study (PETS): 1988–2000 consisting of data about 25,000 individual students. Based on this nationally representative data set, passing or failing calculus was not associated with men's degree intentions (i.e., if they failed once they intended to retake the course and still earn a STEM degree). Women who did not pass calculus, however, were 32% less likely to earn a STEM degree than women who did pass the course (Sanabria & Penner, 2017). Additionally, women were more likely than men to switch out of a STEM major after taking Calculus I even if they passed the course (Ellis et al., 2016). Given the deterrents women overcame before reaching Calculus I, this required course should not pose a new barrier to overcome. Two noted reasons women switched their major after passing Calculus I were perceived lower performance in mathematics and lack of confidence (Ellis et al., 2016; Good et al., 2012). These two reasons have been attributed to the competitive nature of many STEM courses (Blackburn, 2017), which can weaken women's already lowered sense of belonging (Seymour & Hewitt, 1997; Vogt et al., 2007).

Retaining women in their pursuit of STEM degrees at the completion of their calculus education could increase the number of women in the STEM workforce by 20% (Ellis et al., 2016). Unfortunately, our current educational techniques (i.e., lecture, whole class discussion, and traditional assessment) have not been shown to increase women's interest in and persistence for obtaining STEM degrees and entering the workforce (Gonzalez & Kuenzi, 2012). In Chapter 2, I provide further expansion on these educational techniques. With a change in the way students learn mathematics, educators must also consider a change in the way student success in learning is assessed (Berry & Nyman, 2002).

Why Group Testing?

In 1993, the National Research Council's (NRC) recommended that mathematics educators teach mathematics using reform methods (i.e., group work and student collaboration), and suggested that assessments mirror the learning environment as "test[s] should measure what's worth learning, not what's easy to measure" (p.4). In 1995, Webb discussed the four main purposes of assessment of learning: (1) an individual student's proficiency of the subject matter in knowledge and skills (i.e., measured by traditional individual testing methods); (2) an individual student's performance after working within a group (i.e., measured through traditional individual testing methods in a class that utilizes group work); (3) group effectiveness and productivity (i.e., a group's performance with classroom tasks); and (4) students' interpersonal skills (i.e., effective group membership; Webb, 1995). In mathematics classrooms, however, students are usually formally tested only on individual expertise of the subject matter knowledge (i.e. purpose number one and two above). Additionally, it is important to consider that students' learning is influenced by the assessment methods used in a classroom (Hattie & Timperley, 2007). The testing of individual knowledge and skills perpetuates the competitive nature of learning mathematics. The increased use of student-centered teaching strategies creates a mismatch between the learning and assessing contexts: students are asked to work in groups to learn together cooperatively, but are tested individually and potentially competitively.

If the learning environment is to match the testing environment, students' learning must be measured using both individual and collaborative assessments (i.e., group assessments). Group assessment methods can help reduce the perceived competitive nature of STEM courses and allow students to focus more on learning than on the grade they can achieve on the test (Paterson et al., 2013; Pedersen & Liu, 2003). Group testing is a student-centered assessment strategy that can help increase social interaction, student's perception of learning (Revere et al., 2008), and students' confidence in mathematics (Goetz, 2005). Studies that have explored the use of collaborative testing methods will be expanded on in Chapter 2. Specifically, Chapter 2 contains a literature review around group testing in various disciplines. In this chapter, the reader will see that group testing has been sparsely used in mathematics classrooms, and that the interest in this assessment technique has increased in the past decade. This study aims to extend the literature base on group testing, specifically focusing on the link between how group testing may enhance women's sense of belonging and the nature of their social interactions during their Calculus I course (i.e., group testing will be used as a catalyst for social interactions).

Purpose of Study

The purpose of this study is to investigate women's sense of belonging and social interactions in a reformed Calculus I course while they are engaged in group testing. Specifically, I investigated the following three research questions:

1. In a reform Calculus I course, how do students' performance on group tests compare with their performance on individual tests; and how do gender, and the intersection of race and gender contribute to the differences, if at all?
2. How does the sense of belonging to a mathematics community change for female students after engaging in group testing in a reform Calculus I course, if at all?
3. How are social interactions within the group composition different based on the composition of the group, if at all, and can any differences be used to explain differences in the female students' sense of belonging?

Significance of the Study

This study is significant in at least two ways. First, it offers the field a compelling reason to reexamine student-centered instructional and assessment strategies. Second, this study highlights retention methods focused on maintaining women in the STEM pipeline, increasing student collaboration, and women's sense of belonging in a mathematics community. Therefore, this study will contribute to a body of knowledge on women's sense of belonging in mathematics (Good et al., 2012) and the use of group testing in mathematics (Berry & Nyman, 2002; Goetz, 2005; Paterson et al., 2013; MacArthur, 2019) as an equitable student-centered assessment.

Definitions

Throughout this study, I use the key terms listed and defined below. The following section is intended to provide clarity of the terms.

Authority

Authority as used in this dissertation is defined as the relation that "exists when one person (or group of people) tend[s] to obey, act on, or accept without question the statements or commands of another person (or group of people or any other entity capable of producing statements or commands)" (Amit & Fried, 2005, p. 147). Although authority can have a broader meaning and not only consist of people (i.e., a textbook, testing materials, or even technology can be seen as an authority), I am narrowing the definition for the purpose of this dissertation as the students did not have a textbook to use for this course. The people or group of people whose statements are accepted without question hold the authority. People of authority in the study can include the students (when engaged in group work), the STEM Peer Teachers (SPTs, defined below), and the instructor. This study will only consider the authority of the students themselves while they are taking a group test.

Cooperative and Collaborative Work

Both adjectives *cooperative* and *collaborative* refer to instances when students work together to complete a task. The difference in the two terms is subtle. For the purpose of this study, they are defined as follows: Cooperative work is used for situations when students are allowed time to work alone first and then confer their mathematical ideas with a partner or group. Collaborative work is used for situations when students work on a problem together without individual work time to build their mathematical ideas together.

Engagement

Engagement is defined as a person's participation within a certain context and becoming involved in a specific context (Merriam Webster, 2019). For this study, engagement of a student is defined as their participation with their group. The context in which the students are participating is the group work and in the creation of the norms within their group. This type of engagement requires social interactions such as speech and actions. Speech can be considered either on-task (i.e., offering mathematical ideas of questions) or off-task (i.e., speech not directly concerning the mathematics) (Langer-Osuna, 2016). Researchers have noted the benefits of student engagement specific to increased retention (Carver et al., 2017; Chen, 2013; Kuh et al., 2008; Ohland et al., 2008).

Group Work

Group work refers to a group of three or four students working together during instructional class time. This does not include work outside of the regularly scheduled class. During group work, students are working together toward a common goal (e.g., solving a given problem).

Student-Centered Activities and Pedagogies

Student-centered activities refer to a wide variety of pedagogical approaches and activities in the classroom that are centered on student-to-student discussions of mathematical ideas (Adler, 1997; Ball, 1991; Le Roux, 2011). Such activities include groups of students working through mathematics problems, investigations, or projects and includes methods such as Inquiry-Based Learning (IBL) and Inquiry-Oriented Instruction (IOI). For this study, I consider the recent commentary from Laursen and Rasmussen (2019) in which they argued that all student-centered pedagogical approaches share four common pillars that form Inquiry-Based Mathematics Education (IBME). These pillars are (1) student engagement in meaningful mathematics, (2) student collaboration for sense making, (3) instructor inquiry into student thinking, and (4) equitable instructional practice to include all in rigorous mathematical learning and mathematical identity-building (Laursen & Rasmussen, 2019). Therefore, student-centered activities and pedagogies are defined as student-centered pedagogical approaches that exhibit the four pillars of IBME.

Student Centered Assessments

Student-centered assessments are evaluative methods that mirror the way students learn while working in groups (Johnson & Johnson, 2004). If students generally work in pairs during a mathematics course, then a student-centered assessment would have the same pairs of students working through test questions together. If students generally work in groups of three or four peers, then a student-centered assessment would have the three or four students working together to answer questions throughout the test.

STEM Peer Teachers

STEM Peer Teachers (SPTs) are supplemental teacher assistants and tutors who have been trained with classroom management and active learning techniques (Carver et al., 2017). These students have all previously completed Calculus I. The SPTs were present in the classroom every class meeting of the course used in this study to assist with the implementation of the course.

Stereotype

A stereotype is a generalized belief about a person or group of people (Cardwell, 2014). These beliefs are generalized because of assumptions that the stereotype is true for similar people or members of the group (Cardwell, 2014).

Sense of Belonging

The sense of belonging, lack of sense of isolation, or lack of sense of alienation are used interchangeably throughout this document and refer to the feeling of being a contributing and respected member of a certain group in a given context (Good et al., 2012; Strayhorn, 2019). Good et al. (2012) classify a sense of belonging as the feeling of membership, acceptance, trust, and desire to remain a member of a certain community. There are dual contexts for the community in study: the mathematics classroom and the group in which the student worked during group work settings.

Social Interaction

Social interactions can be broadly defined as an exchange of ideas between an individual and another person via written or verbal communication (Sfard, 2003). In sociology and psychology, social interaction is defined as a dynamic sequence of interaction and responses between two or more individuals (VandenBos, 2007). Social interaction in the psychological context includes social roles among members in a group, dynamics of group behavior, as well as leadership and conformity among individuals in a group (VandenBos, 2007). Although the term social interaction itself can have a broad definition as defined by Sfard (2003) for this study, I refer to social interactions as defined in sociology and psychology by VandenBos (2007) as I am referring to the dynamics of students while in groups.

Underrepresented Minority

The National Science Foundation (2017) defines underrepresented minorities (URMs) as groups of students who are historically underrepresented in STEM disciplines. This definition includes students that are African Americans, Alaskan Natives, American Indians, Hispanic Americans, Native Hawaiians, and Native Pacific Islanders (National Science Foundation, 2017).

Chapter Summary

In conclusion, there is a need for more women to enter and remain in the STEM workforce. However, many are lost in the “leaky pipeline” due to negative stereotypes and women’s lack of sense of belonging in the STEM domains. Understanding how men and women communicate and the impact of these communications on women’s sense of belonging during their undergraduate education can help generate methods at the university level to retain more women in STEM. Calculus I is a vital course for all STEM majors and many women are lost in the pipeline due to this course even if they pass. With the current trend in mathematics education using student-centered pedagogies that often have students working together in the classroom, group testing may help eliminate the competitive nature of this particular STEM course, increase communication among the women in the course, and perhaps foster a sense of belonging to mathematics or mathematically-intensive fields for women.

CHAPTER TWO: LITERATURE REVIEW

> "It is tiresome and counterproductive to argue about the relative merits of male and female mathematicians, for we have no precise method of quantifying or comparing their individual accomplishments" (Osen, 1974, p.2).

Introduction

The sentiment from the comment above should be carried as a reminder throughout this dissertation. That is, this study is not meant to compare men and women in undergraduate mathematics classrooms, but instead to illuminate and alleviate the threats to women in the Calculus I classroom. As discussed in Chapter 1, there are many factors that need to be addressed to attract and retain more women in the STEM pipeline. I focused on two main factors: stereotype threat and sense of belonging of women in mathematics. The choice to focus on these two factors specifically is due to the perceived competitive nature of Calculus I and reforms in pedagogy used in Calculus I courses that increase the use of group environments, which may increase the feeling of isolation and/or alienation for women in the classroom.

In this chapter, four main areas of research are addressed that are centrally relevant to this study: women in mathematics, undergraduate calculus reform, equitable education techniques, and the use of group assessments in undergraduate classrooms. Through my review of the literature, I will summarize several important findings from research on stereotype threat with regard to women in mathematics, and the importance of sense of belonging with regards to the affective domain. In the review of undergraduate calculus reform, I will discuss the shift in pedagogy, and the importance of using student centered learning environments in the calculus classroom. Next, I will focus on characteristics of an equitable classroom and current equitable instruction techniques that include group work. Finally, I will discuss forms of group assessments used in other fields such as nursing and psychology, the few studies in mathematics, and the effects group assessments have on certain influential factors of learning. The chapter ends with the explication of the theoretical and conceptual frameworks that underlie the design of the research presented.

Women in Mathematics

In this section, I focus on two main factors that need to be addressed in order to retain women in STEM fields: stereotype threat and sense of belonging. I chose to address these two factors, stereotype threats and sense of belonging to mathematics, as they can both hinder women's performance in mathematics, and hence keep women out of STEM fields (Chavatzia, 2017; Good et al., 2012; Pronin et al., 2004; Spencer et al., 1999). When it comes to understanding the lack of women in STEM and highly intensive mathematics fields, it is first useful to understand how stereotype threats affect women, specifically with regards to their identity and performance in mathematics (Pronin et al., 2004), and understand how students' sense of belonging in STEM is directly related to gender-based STEM stereotypes (Dasgupta, 2011; Master & Meltzoff, 2020).

Stereotype Threat

A stereotype is a generalized belief about a person or group of people (Cardwell, 2014) and stereotype threat is "the social-psychological threat that occurs when one is in a situation or doing something for which a negative stereotype about one's group applies" (Steele, 1997; p. 614). A common stereotype that affects women in mathematics classes is the statement that men are better at mathematics than women (Cadaret et al., 2017;

Spencer et al., 1999). Spencer et al. (1999) explained that this burden is one placed on women alone and stated:

> Thus, in situations where math skills are exposed to judgment—be it a formal test, classroom participation, or simply computing the waiter's tip—women bear the extra burden of having a stereotype that alleges a sex-based inability. This is a predicament that others, not stereotyped in this way, do not bear. (p. 6)

Therefore, this stereotype can threaten women's performance in mathematics and is still happening in classrooms today. For instance, in 2019 Thorson and colleagues found that women working in pairs with a male partner performed worse than the men even if there was not an explicit threat of being stereotyped. This may be explained by Cadaret and colleagues (2017) who suggested that "a person exposed to stereotype threat feels threatened about being negatively stereotyped, judged, treated stereotypically, or having to conform to the stereotype" (p. 42). Additionally, Thorson et al. (2019) found that women working in pairs with another woman asked more questions and were more socially engaged than women working with a male partner or alone. In order to better understand the role of stereotype threat on women in mathematics, I present a review of the two seminal studies in which this phenomenon was grounded.

There are two seminal studies by Steele (1997) and Spencer et al. (1999) regarding the stereotype threat that affects women with regard to their performance in mathematics. In the first study in 1997, the researchers conducted experiments with men and women that examined gender stereotype threat versus genetic limitations (i.e., intellectual ability) that defined the risk of people judging women to have weaker mathematical skills as *stereotype threat*. In 1999, Spencer and his colleagues continued to investigate this phenomenon and conducted multiple studies regarding the effect of stereotype threat on women's performance on mathematics tests. The results published in Spencer et al. (1999) indicated this threat is real and can affect women's performance with regard to mathematics. Recall the negative stereotype many women may face in STEM fields is the sentiment that 'women are not as good at math as men' (Spencer et al., 1999). In the following paragraphs, I provide the details on the two studies conducted in 1999 by Spencer and colleagues.

The first study conducted by Spencer et al. (1999) consisted of 28 men and 28 women from the University of Michigan who had all completed one semester of calculus and received a B or better in the course. Participants took a mathematics test that was labeled as either difficult or easy[2] to investigate the stereotype that "women underperform on difficult tests but perform just as well on easier test[s]" (Spencer et al., 1999, p. 9). Participants arrived in groups of three to six consisting of both men and women, and all participants were allotted 30-minutes to complete the test. A two-way ANOVA was conducted on the data and women in the study did in fact perform comparable to men on the 'easy' test and considerably worse than men on the 'difficult' test (Spencer et al., 1999). Since the test itself may have led to the gender difference in the scores (Spencer et al., 1999) a second study was conducted.

The second study conducted by Spencer et al. (1999) consisted of 24 men and 30 women with the same criteria as their first study. Participants were split into two different

---

[2] The labeling of difficult or easy was not shared with the participants. Questions for the difficult mathematics test consisted of questions from the Graduate Record Examination (GRE) advanced mathematics exam where the easy test consisted of questions from the quantitative section of the general GRE exam (Spencer et al., 1999). Students in the course were no more than one semester past their calculus course and did not have the advanced mathematics formal education to complete the questions on the GRE advanced mathematics exam.

groups, each group for this second study took the difficult test (i.e., GRE advanced mathematics exam questions). The difference of the groups in the study was how the test was presented to them. In the first group, the participants (consisting of both men and women) were told the test they were about to take was confirmed to have gender differences. The comparison group, the participants (consisting of both men and women) were told the test they were about to take had no gender differences. A two-way ANOVA was conducted on the data and on the scores of men and women in the two groups (Spencer et al., 1999). Women in the gender difference group had lower scores than the men. Women in the no gender difference group, however, performed similarly to the men (Spencer et al., 1999). After this finding, the researchers conducted a 2X2 ANOVA and revealed the main effect of gender and testing condition ($p < .05$). That is, women's performance on the test was no different from their male counterpart when the stereotype threat was reduced (i.e., the first group), and if the stereotype treat was present (i.e., the comparison group) then it influenced the women's performance on mathematics tests (Spencer et al., 1999).

As described above, stereotype threats can affect women's performance on mathematics tests. In addition, stereotype threats can influence a woman's sense of belonging in mathematics (Good et al., 2012; Master & Meltzoff, 2020) as well as create a physiological response (i.e., stress response) for women who feel a threat on their intellectual ability due to such stereotypes (Cadaret et al., 2017). If one does not feel as if they belong to a certain situation (i.e., women's lower mathematical sense of belonging) then the threats of stereotyping could additionally affect their performance in that situation (i.e., performance on mathematics tests). One way suggested in the research to reduce stereotype threat for women in mathematics is by having a role model (i.e., a successful woman in mathematics) in the classroom (Marx et al., 2005; Shapiro & Williams, 2012; Master & Meltzoff, 2020). In the next section, I discuss the relevant literature regarding the sense of belonging for women in mathematics.

Sense of Belonging

The need to belong and to form interpersonal social attachments is a fundamental human need and is a powerful motivator of one's social behavior (Baumeister & Leary, 1995). For students in academic settings, a sense of belonging is their own internalized representation of how they see themselves "fitting in" with a certain group (Master et al., 2016; Master & Meltzoff, 2020; Strayhorn, 2019). That is, the sense of belonging can be context specific for some college students, it can change from class to class, from groups within each class, and can be salient at times (Lewis & Hodges, 2015; Strayhorn, 2019). Students who do have a strong sense of belonging in academia and have social attachments have shown to have positive changes in both motivation and achievement (Osterman, 2000).

A sense of belonging is a human need for a person to feel as if they are a valuable member of a community and "when sense of belonging is protected by learning environments that convey a malleable view of intelligence, students may be less vulnerable to the impact of negative stereotypes on achievement and intention to remain in the domain" (Good et al., 2012, p. 714). That is, Good and colleagues (2012) described how sense of belonging to mathematics in learning environments is not only impacted by negative stereotypes but impacted as well as by the entity or incremental theories of intelligence[3] of their instructors. Since I acknowledge that I have an incremental view of intelligence (i.e., I truly believe anyone can learn calculus) this topic will not be explored in this study. Additionally, note that learning environments can be different and this need to belong is context specific and can affect both men and women in an academic setting (Good et al., 2012; Master & Meltzoff, 2020; Strayhorn, 2019). Additionally, many researchers have examined how one's sense of belonging is a strong predictor of interest for that context (i.e., STEM; Cech et al., 2011; Cheryan et al., 2009; Cheryan & Plaut, 2010; Cheryan et al., 2017; Good et al., 2012; Walton & Cohen, 2007) and for college women in STEM their sense of belonging is positively correlated to their persistence in their major more than it is for men (Banchefsky et al., 2019; Good et al., 2012; London et al., 2011; Stout & Blaney, 2017).

Women in male-dominated fields do not always have a sense that they belong to that group (Good et al., 2012; Lewis et al., 2016; Master et al., 2016; Master & Meltzoff, 2020; Moudgalya et al., 2021; Murphy et al., 2007; Walton & Cohen, 2007). Considering STEM fields, such as computer science and physics, the lack of women in college classes can further widen the gender gap due to women feeling a lack of social belonging to the group (i.e., not represented in numbers within the group of people; Lewis et al., 2017; Sax et al., 2018; Stout & Blaney, 2017). Additionally, women have further threats to their sense of belonging in mathematically demanding fields due to negative stereotypes on the

---

[3] The entity theory, or fixed mind set, view of intelligence corresponds to views that intelligence is a fixed internal characteristic. The incremental theory, or open mind set, of intelligence corresponds to the view that intelligence is malleable and can be increased (Dweck, 1999).

perceived lower mathematical ability of women versus men (Dar-Nimrod & Heine, 2006; Diekman et al., 2015; Good et al., 2008; Good et al., 2012; Spencer et al., 1999).

The need to feel as though one belongs is an important component to learning mathematics (Sfard, 2003) especially for women in mathematics (Diekman et al., 2015; Good et al., 2012). One's sense of belonging to *mathematics* as defined and measured by Good et al. (2012) encapsulates the essences of the affective domain by considering how one interprets their membership within a group, their acceptance with that group, their feelings about being part of that group (i.e., positively and negatively), their trust of authority figures of the group (i.e., teacher or instructor and testing materials), and their desire to maintain membership in that group. In 2012 Good and colleagues conducted two studies to create and validate a scale to measure one's sense of belonging to mathematics. Good et al. (2012) created questions aimed at five components of belongingness: membership, affect or feelings, acceptance, trust, and desire to remain in that domain. Details regarding the survey are provided in Chapter 3. Considering how stereotype threat and sense of belonging play large roles in women's learning in mathematics courses, it is now imperative to understand how mathematics teaching has evolved over the last few decades. In the next section, I provide a review of the literature centered around calculus reform.

Calculus Reform

> "Reform raises questions about the core beliefs of mathematics education, moving to restructure thinking about the nature of mathematics, how it is taught, how it is learned, and, ultimately, what constitutes success in learning it." (Ellis & Berry, 2005, p.8).

Since Calculus I is an essential course for many students pursuing degrees in STEM (Ellis et al., 2016; Sanabria & Penner, 2017) and with its historically low pass rate (Bressoud, 2015), changes in the way undergraduate calculus courses are taught are being made (Ellis et al., 2016; Saxe et al., 2015). Researchers have long been exploring how students learn mathematics, their perceptions of the classroom environment, and the benefits of collaborative learning (Cohen, 1994; Draskovic et al., 2004; Esmonde, 2009; Johnson & Johnson, 1999; Kearns & Bolyard, 2018; Webb, 1995). Teacher-centered forms of instruction (i.e., traditional lecture) generally have lower pass rates than other instructional formats (Freeman et al., 2014; Saxe et al., 2015). Furthermore, research has shown that active learning can help retain STEM majors (Freeman et al., 2014) and small-group collaborative learning has been shown to have a positive impact on STEM majors' persistence, achievement, and attitudes (Draskovic et al., 2004; Kearns & Bolyard, 2018; King & Behnke, 2005; Micari et al., 2010; Smith et al., 2014; Springer et al., 1999). The amount of research has increased over the past decade, and is aligned with the shift in mathematics education that Ellis & Berry described in 2005: a shift away from a traditional paradigm that is procedurally focused to a cognitive cultural paradigm that focuses on learning that incorporates the culture, beliefs, and students' sense of self.

The calculus reform movement started in the late 1980's and is still occurring today especially due to the shortage of STEM majors. The themes of the reform include studies about a shift in focus from rote memorization and procedural knowledge to conceptual understanding (Cavanagh, 1996; Heid, 1988; Judson, 1988; White & Mitchelmore, 1996), the effect of technology in calculus classrooms (Heid, 1988; Hickernell & Proskurowski, 1985; Hurley et al., 1999; Judson, 1990; Palmiter, 1991;

Tall, 1986), a focus on students' understanding of specific central topics in calculus (i.e., limits; Aydin, & Mutlu, 2013; Bezuidenhout, 2001; Fernandez-Plaza & Simpson, 2016; Ferrini-Mundy & Guether-Graham, 1991), changing curricular materials (Ferrini-Mundy & Graham, 1991; Lithner, 2004; Lovric & Burazin, 2018), students' understanding and proficiency in prerequisite mathematics (Ilhan et al., 2011; Kabael, 2014; Kidron, 2015; Ogden et al., 2018; Pettersson & Scheja, 2008; Roh, 2008; Scheja, & Pettersson, 2010; Stewart et al., 2018), and the shift from lecture style and teacher-centered pedagogies to active learning and student-centered pedagogies (Cory & Martin, 2018; Ferrini-Mundy & Graham, 1991; Joiner et al., 2002). Since the focus of this study is student-centered pedagogies and the benefits of these pedagogies specific to women in the STEM fields, I provide a literature review regarding the current state of calculus reform with respect to a shift in pedagogy impacts on student learning in calculus.

Reform in Pedagogy

A national call to move away from traditional lecture style in undergraduate calculus courses started nearly forty years ago and continues today (Ganter, 2001; McGarvey, 1981; Saxe et al., 2015). Research strongly supports the notion that calculus students do not need to learn procedures first in order to understand key concepts in calculus (Heid, 1988; Judson, 1988) and that the learning of concepts can be accomplished using student-centered instruction (Kogan & Laursen, 2014; Laursen et al., 2014; Rasmussen et al., 2006). Student-centered instruction, such as Inquiry-Based Learning (IBL) and Inquiry-Based Mathematics Education (IBME), has been shown to significantly increase student learning outcomes (Boaler & Greeno, 2000; Bressoud, 2011; Freeman et. al, 2014; Hassi & Laursen, 2015; Kogan & Laursen, 2014).

Additionally, student-centered learning environments have been shown to improve low-achieving students' performance in subsequent classes, as well as increase persistence in high-achieving students (Hassi & Laursen, 2015; Kogan & Laursen, 2014). Failure rates have been shown to be up to 55% higher in lecture-based courses compared to those using active learning student-centered pedagogies (Douglas, 1986; Freeman et al., 2014). Although student-centered learning has shown positive results in many classrooms, recently, a national survey conducted by the Mathematical Association of America (MAA) concluded that traditional lecture is still the prominent form of instruction in undergraduate calculus courses (Bressoud et al., 2015).

In 2015, Bressoud and Rasmussen identified seven characteristics of effective Calculus I programs from 17 universities that had reported success in their calculus reform. A Calculus I program consists of more than just the course itself as seen in the characteristics. The seven characteristics shared by the effective programs included:

1. Regularly examining the pass rates and retention rates of students each year to guide future reform;
2. Examining and improving the placement exam for incoming students to ensure appropriate placement into their first undergraduate mathematics course;
3. Coordinating instruction and increasing the communication between all instructors who are scheduled to teach calculus;
4. Maintaining an academically challenging and engaging calculus course;

5. Creating and maintaining training programs for graduate teaching assistants who are involved in calculus courses as either teaching assistants, supplemental help, or instructors;
6. Having student support services available for all students, monitoring and providing help to foster academic and social integration for at-risk students, and maintaining a willingness to change the programs designed to help all students;
7. Using student-centered pedagogies and active learning strategies that require students to explain their thinking in an effort to help students understand what it means to engage with mathematics (Bressoud & Rasmussen, 2015).

Note first that the characteristics indicate a successful Calculus I program consists of more than just the course itself, and includes regular evaluation of the program through continual examination of pass rates and retention rates. Focusing on the course-related characteristics, note that student-centered pedagogies are mentioned explicitly, but lecturing is not. Nevertheless, researchers today are still questioning whether the learning gains in active learning classrooms are truly different from those that can be seen in traditional classrooms (Cory & Martin, 2018; Gerasimova et al., 2017). This topic will be discussed in the following section along with equitable education techniques. The list above also presents the need for the training of all instructors who teach calculus to maintain programs that will help students succeed in student-centered classrooms (Reinholz, 2017). This dissertation study takes into consideration the training of the course instructor, graduate students, and SPTs with the following educational techniques that have been identified for equitable instruction in high school teachers.

In August of 2019, a month after I collected the data for this study, Hagman added one more characteristic to the list above: diversity, equity, and inclusion practices. She further defined each practice separately. The diversity practice includes actions within the program that leads to including more students with diverse backgrounds (Hagman, 2019). She noted this need came from the lack of diverse students in the original 17 universities examined for the first seven characteristics. That is, Hagman (2019) noted those universities in the first study had predominantly white and Asian male populations. Equity practices included actions related to access, achievement, identity, and power as discussed by Gutiérrez (2013). A successful program, therefore, needs to acknowledge the barriers each student has to face in order to be successful and help in dismantling the barriers for all students (Hagman, 2019). Finally, inclusion practices included actions that support the full participation of all students within the program (Hagman, 2019). This dissertation study fully considered the eight characteristics of successful calculus programs (i.e., the seven characteristics first presented and the eighth characteristic diversity, equity, and inclusion characteristics). The program in which the study was conducted was built to support a diverse population of STEM students. As the course instructor, I considered equitable testing methods, and most importantly the focus with these testing methods was on how the women felt included or not in the classroom and within their group. In the following section I discuss current equitable educational techniques that are present in student-centered classrooms. This section considers more than just calculus classrooms. Foundational literature from successful undergraduate student-centered classrooms is included to better understand the possible effects of these equitable classroom techniques.

Equitable Education Techniques

There have been numerous studies over the past 30 years that have examined classrooms that de-emphasize rote memorization and procedural skills, and instead emphasize student's active involvement and responsibility for creating and learning mathematical concepts (Freeman et al., 2014). Considering calculus alone, the results of these studies indicate such classrooms can increase student achievement, knowledge, or skills (Bookman & Friedman, 1994, 1999; Ganter & Jirovtek, 2000; Hurley et al., 1999; Meel, 1998; Penn, 1994; Roddick, 2003; Schwingendorf et al., 2000; Smith & Star, 2007; Williams, 1998), student attitudes and beliefs (Bookman & Friedman, 1994, 1999; Smith & Star, 2007), and have long-term effects such as persistence (Bookman & Friedman, 1994, 1999; Hurley et al., 1999; McDonald et al., 2000; Schwingendorf et al., 2000). Not all classrooms look the same when observed, however, so how would one know if they are observing equitable active learning classroom instruction? The National Research Council's report in 2012 called for identification of these instructional approach features so researchers can explore ways to help minoritized groups of students in mathematics classrooms.

For this study, consider Perry (2013; 2018) who explored three early career high school teachers. From the examination of these teachers, Perry (2018) identified seven characteristics of equitable classroom spaces:

1. Instructors use high cognitive demanding tasks in the classroom.

2. Students are given clear expectations for how they are expected to engage with each task.
3. Students are provided with resources that will help them learn.
4. Students are given opportunities to publicly share their mathematical work in both verbal and written forms.
5. Students are given opportunities to explain their thinking and justify their mathematics using both verbal and written methods.
6. Instructors answer students' questions with guiding questions to help move the students thinking forward.
7. Instructors model the use of high-level reasoning skills during lessons.

These characteristics can be seen in many student-centered classrooms with the use of collaborative groups. One such student-centered approach previously analyzed in mathematics courses that has been shown to help women succeed is Inquiry-Based Learning (IBL). In 2011, Laursen and colleagues conducted one of the largest studies on the use of IBL in over 100 different courses across four U.S. universities. The courses included first year honors mathematics courses, upper-level courses for mathematics majors, and elementary and secondary mathematics courses for future teachers. Each instructor for all of the courses in their study was recruited by the individual universities and trained to implement IBL in their assigned course. The IBL process deemphasizes rote memorization and procedural skills and instead seeks to help students develop critical thinking skills by having them work on ill-defined problems, apply logic to situations, and make and analyze mathematical arguments (Laursen et al., 2011). The use of IBL also can increase student creativity and persistence with regards to learning mathematics (Laursen et al., 2011).

In 2014, Laursen and colleagues disseminated more of the results from the 2011 report regarding gender differences. Specifically, using a quasi-experimental design they examined both student learning and affective outcomes of the large IBL study. Forty-two of the 100 courses were observed and compared to non-IBL courses. The observed courses contained either mathematics majors or preservice teachers. Data for the study were only analyzed for students who were majoring in mathematics or mathematically intensive STEM fields. Students were asked to take a pre- and post-course survey of their attitudes, beliefs, and approaches to learning mathematics (Laursen et al., 2011). Students took this survey again at the end of the course in addition to the Student Assessment of their Learning Gains survey, which is a validated self-report instrument (Seymour et al., 2000). The authors of the learning gains survey used a five-point Likert scale consisting of questions regarding three types of learning gains. Specifically, the survey measured cognitive gains (i.e., understanding and thinking), affective gains (i.e., confidence, persistence, and positive attitude about mathematics), and social gains (i.e., collaboration and comfort when working with others, seeking help, and appreciating different perspectives). Data were collected and compared for 358 students (250 men and 108 women) in non-IBL courses and 544 students (366 men and 178 women) in IBL courses. Results showed no statistical difference in cognitive and affective gains for women and men in the IBL courses. Women in the non-IBL courses had statistically lower gains than men on both the cognitive and affective measures. Additionally, from the pre-survey to post-survey on attitude, beliefs, and approach to learning mathematics, women in the non-IBL courses had a substantial decrease in their confidence and desire to take more mathematics. It is interesting to note that, although it was not statistically significant, men in the non-IBL courses also had a decrease in confidence and intent to pursue more mathematics. Contrary to the non-IBL results, women in the IBL courses had an increase in intent to pursue mathematics that was similar to their male counterparts. Finally, women in the IBL courses had a significant increase in their confidence in doing mathematics that surpassed the increase men had in the same course. The Lauren study is important for the purposes of my study because the increase in the affective domain for women was the largest outcome for women in IBL classrooms. My study attempts to add to this literature base by offering an assessment method for student-centered classrooms that could further increase other areas of women's affect domain (i.e., membership and acceptance to a group).

Even with the evidence of students' learning in student-centered classrooms from both secondary school research (Boaler, 1998; Boaler & Staples, 2008; Clarke et al., 2004; Slavin et al., 2009) and undergraduate research (Bressoud, 2011; Carver et al., 2017; Freeman et. al, 2014; Hassi & Laursen, 2015; Kogan & Laursen, 2014; Laursen et al., 2014) the movement to use such Inquiry-Based Mathematics Education (IBME) activities in undergraduate mathematics has been very slow (DeHaan, 2005; Fairweather, 2008; Laursen et al., 2014; Walczyk et al., 2007). In addition to the slow movement, currently several researchers still question whether the learning gains in these types of student-centered classrooms are truly different from those that can be seen in traditional classrooms (Cory & Martin, 2018; Gerasimova et al., 2017). There is a notable concern about the training of instructors to implement these methods effectively (Reinholz, 2017).

This concern, however, is beyond the scope of this study. Additionally, two recent studies have noted the gender inequities in performance through active-learning strategies (Johnson et al., 2020; Maries et al., 2020). Johnson and colleagues (2020) examined student performance across gender in abstract algebra with Inquiry-Oriented Instruction (IOI). Their results indicated that the IOI setting was beneficial only for men's performance in the course. Because I focused on the first-year course barriers for women majoring in STEM, however, the Johnson et al. (2020) study goes beyond the scope of my current study, as abstract algebra is generally an upper-level mathematics course and thus beyond calculus. More closely related to my study, Maries et al. (2020) examined active learning methods in a calculus-based physics course in comparison to lecture-based courses. Findings from this study indicated that active learning methods may contribute to a larger gender gap than lecture-based courses as the men outperformed the women by a larger percentage in the active learning classroom (10%) than in the lecture classroom (6%; Maries et al., 2020). Maries and colleagues (2020) also found that female students who felt as if they were not included in the learning environment due to stereotype threat performed significantly worse compared to the male students. The authors suggested that to empower all students an instructor should attend to providing an equitable and inclusive learning environment (Maries et al., 2020). A feeling of not being included can be equated to a lack of sense of belonging, in which my study is centered. Therefore, I intend to extend the results of the Maries et al. (2020) study by investigating both performance and the feelings of the students in an active learning environment.

Active learning environments such as IBL are centered on student engagement and contributions to group and whole class conversations. Getting students to contribute to these classroom discussions can be difficult (Walshaw & Anthony, 2008), especially for women who have felt invisible or excluded during group work (Joseph et al., 2017; Solomon, 2007). It is important to understand the way students interact in the mathematics classroom because “the ways people talk and interact are powerful influences on who they are, and can become, with respect to mathematics” (Bishop, 2012, p. 34). Additionally, by understanding how students’ interactions can influence one another during collaborative group work, educators can begin to create equitable working groups for all students in active learning classrooms. In the following section I will describe the dynamic of influence framework (Langer-Osuna, 2016) that was used to help understand the social interactions among students during the group test.

Dynamics of Influence

When grouping students together in a classroom, instructors may want to consider how “the power of groups is reflected in the impact they have on students’ actions” (Johnson & Johnson, 2004, p.7). That is, when people are in distinct groups, the people in these groups define what is right and wrong within that group, what is acceptable behavior (Johnson & Johnson, 2004), and influence what individuals think and how they act in that group (Bishop, 2012; Johnson & Johnson, 2004). To this end, Langer-Osuna (2016) examined students’ actions in a fifth-grade collaborative mathematics classroom in the Southeastern United States. The students in this class regularly engaged in collaborative group work and her analysis centered on the students’ authority relations while working in pairs. Previous work has noted the development of different authority positioning of students between one another that emerges when engaged in group work in student-centered mathematics classrooms (Esmonde & Langer-Osuna, 2013; Langer-

Osuna, 2011; Wood, 2013). In her 2016 study, Langer-Osuna built on these "efforts to understand relationships of power among students by examining a case of collaborative problem solving from the perspective of how students position themselves and one another with different forms of authority" (p. 109). Specifically, she discussed two forms of authority: intellectual and social[4].

Langer-Osuna (2016) analyzed a 90-minute video in which two students collaborated to answer a problem about designing a fruit and vegetable garden. During the collaboration, Langer-Osuna (2016) analyzed the speech and actions of the students throughout the video coding for changes in influence, merit of the ideas each student gave, merit of the behavior of the students, the intellectual authority of the student speaking, and the social authority of the student speaking. The five components (see Table 1) together makeup the dynamics of influence framework that was used as a guide to understand the student interactions in this study (see Chapter 3 for more details).

---

[4] Langer-Osuna uses the term directive authority in place of social authority. The term directive authority relates to a student giving directions to another. I will continue using the term as social authority since "social" is commonly used throughout this proposal.

Table 1.

Influence Framework Components from Langer-Osuna (2016).

Framework Component	Definition
Influence	The student’s idea was either accepted or rejected as part of the solution path.
Intellectual Authority	The student is treated as a credible source or as lacking credibility of their information.
Social Authority	The student is treated as having the right to give directions or not to the group.
Intellectual Merit	The student’s mathematical argument was either accepted or rejected as part of the solution path.
Social Merit	The student’s words or actions are seen as on task or off task by the group.

The pair of students in the study by Langer-Osuna (2016) consisted of one girl (Ana) and one boy (Jerome). The bulk of her data analysis consisted of 89 social interactions in which one student offered an idea or issued a direction to the other, labeled as proposal negotiation units (PNUs). A PNU ended with the other student either accepting or not accepting the idea or directive. Each unit was coded either positively or negatively with one of the five components of the influence framework. For example, if a student suggested a mathematical idea to their partner and the partner agreed with it, then the student who suggested the idea would be coded with a positive intellectual merit. The results of the coding showed that Ana had considerably more influence than Jerome in this collaborative investigation (Langer-Osuna, 2016). Ana had more instances of each component in the framework with social authority ranking the highest (i.e., 20 occurrences of Ana issuing directives were taken by Jerome). Langer-Osuna (2016) noted how both the social and intellectual authority in this study affected which mathematical ideas were accepted and followed when solving the problem. Additionally, Langer-Osuna (2016) noted how the accepted social authority “can become linked to interactions that position the same student as a credible source of mathematical knowledge (intellectual authority)” (p. 120). That is, both the social and intellectual influencers geared the mathematical thinking of the pair of students and ultimately the answer the students generated for the given mathematical problem.

Since student-to-student interactions can influence students' learning, the mathematical ideas that are explored in the group, and the positioning of students in that group (Langer-Osuna, 2016), I took into consideration the arrangement of students into their collaborative working groups for this study. Although the authority was the female of the pair of students in Langer-Osuna’s (2016) study, recall these students were in fifth grade in which the influence of gender roles may have not been prevalent (Shapiro et al., 2015). The influence of men and women working in collaborative groups may be significantly different than these fifth-graders (see Women in Mathematics). In Chapter 3 I describe the methods used to purposefully place students into groups in which they would learn best by taking into consideration the input of the SPTs, instructor, and students. The arrangements are purposeful to ensure students are placed in groups where they feel their voices would be heard by their fellow group members and consider the sense of belonging of each student (i.e., feel as if they belong to that particular group). In the following section, I specifically focused on the literature regarding students working in groups during a group testing environment. The literature review below includes areas outside of only mathematics, as group testing is not currently a common student-centered assessment practice in mathematics classrooms.

Group Testing

The use of group testing is a student-centered assessment method during which students need to communicate and collaborate to solve problems. This type of assessment method is most commonly seen in the medical field and has been successfully implemented with nursing students (Duane & Satre, 2014; Eastridge, 2018; Lusk & Conklin, 2003; Sandahl, 2010), psychology students (Curless, 2012; Pandey & Kapitanoff, 2011; Vogler & Robinson, 2016), physiology students (Cortright et al., 2003; Giuliodori et al., 2008; Rao et al., 2002), and biotechnology students (Srougi et al., 2013). A few other courses have begun exploring group assessments such as a general science education course for undergraduate students (Gilley & Clarkston, 2014), non-mathematics majors' statistics courses (Kapitanoff & Pandey, 2018; Revere et al., 2008), and engineering students in a mechanics course (Jang et al., 2017). However, this assessment strategy is scarcely seen in lower-level undergraduate mathematics classrooms for STEM majors. This could be due to the perceived applied nature of the fields above and the view of mathematics teaching and learning as simply procedural instead of a creative problem-solving process.

Most of these studies using group tests in other fields conceptualized the group test similarly, in which they allowed students to work in groups on a test in class immediately after the same or similar test taken individually (Cortrigh et al., 2003; Duane & Satre, 2014; Eastridge, 2018; Gilley & Clarkston 2014; Giuliodori et al., 2008; Rao et al., 2002). There were slight differences in all of the studies such as the size of groups or the time between the individual and group test, but essentially the procedure of implementation of the group test was the same (i.e., students first took an individual test, then answered the same or similar questions in a group after the individual test). The remaining studies conceptualized group testing differently, I will describe the differences in these last few studies.

The group testing used in Jang et al. (2017) was similar to the studies mentioned previously in that the students first answered questions individually then as a group. The difference, however, was the students used an online team-based assessment system to complete the tests. Each of the remaining studies conceptualized the use of group testing differently. For example, Srougi et al. (2013) gave students a take home portion of an exam, which the students could work through together. Revere et al. (2008) engaged students in an in-class Jeopardy game where teams of students answered questions from several categories. Finally, Lusk and Conklin (2003) compared two nursing courses, one of which used three individual unit tests and a final exam and the other used the three collaborative unit exams and the same final exam.

Although each study was different, each noted some sense of increase in either student communication, social skills, collaboration, accountability, and preparedness (Duane & Satre, 2014; Jang et al., 2017; Lusk & Conklin, 2003; Revere et al., 2008; Sandahl, 2010), problem solving skills (Giuliodori et al., 2008), individual student learning measured by performance on individual exams (Eastridge, 2018; Gilley &

Clarkston, 2014; Giuliodori et al., 2008; Srougi et al., 2013), understanding of concepts (Cortight et al., 2003; Rao et al., 2002), long-term retention of concepts (Cortright et al., 2003; Vogler & Robinson, 2016), students' perceived learning gains (Revere et al., 2008), or lowered testing anxiety (Kapitanoff & Pandey, 2018; Lusk & Conklin, 2003; Pandey & Kapitanoff, 2011). Additionally, multiple studies noted the increase was beneficial to both low- and high-performing students (Eastridge, 2018; Giuliodori et al., 2008; Jang et al., 2017; Kapitanoff & Pandey, 2018; Revere et al., 2008). Jang et al. (2017) examined the possibility that gains for lower performing students could be due to the propagation of answers from the higher performing students. They found, however, that groups that did not start with a higher performing student (i.e., one that knew the answer to the problem before the group test) generally were able to get to a correct answer within three tries on a problem.

Even with the noted success of group assessment within other courses, and the common use of group testing with medical students, this assessment method is not as common in mathematics classrooms. Mathematics educators are starting to consider group testing currently at the high school level. For example, Russell (2019), in a post on A MiddleWeb Blog, described her experimentation with the use of group tests in her high school mathematics course. She decided to try group testing after a comment from Hunt (2018), another high school mathematics teacher, who recommended the idea on a previous blog. Russell (2018) questioned whether her classroom practices helped or hurt the learning of the students in her class. Hunt (2018) explained his use of group testing in lieu of extra credit with students in Algebra II through Calculus. Although this is not empirical evidence, it is important to discuss the advantages and disadvantages Russell

(2019) discussed regarding the use of group testing and to note that teachers at the K-12 level are informally discussing forms of group assessments. Russell (2019) noted the use of the group test allowed her to get a sense of the group's understanding but not the individual students' understanding. Additionally, she noted how some students did more work than others in their group but received the same grade as their group members. This last disadvantage, however, was not brought up by the students in the class as her students stated how they wanted to contribute, and one felt that they should have helped the group more. The fact that her students felt they wanted to contribute more to the group coincides with the finding from Duane and Satre (2014), Jang et al. (2017), and Revere et al. (2008) where group testing can increase student accountability and preparedness. The advantages Russell (2019) observed were the increase in argumentation as students had to defend their answers and explain their thinking to their peers as well as decreased texting anxiety for students who regularly experience this anxiety. The decrease in testing and mathematics anxiety was also noted in one empirically based study which is consistent with that of the nursing students Lusk and Conklin (2003) and psychology students Pandey and Kapitanoff (2011).

In the remainder of this section, I provide a literature review of the four studies that have reported the use of group testing in mathematics classrooms. To find these studies using collaborative testing methods I searched Google Scholar, ERIC - Education Resources Information Center, Journal for Research in Mathematics Education, and Research in Undergraduate Mathematics Education Conference Proceedings for the past ten years. The initial search yielded only two results, and to accommodate for the lack of results I widened the search to the past 15 years. Each search was conducted with key words such as collaborative test, group test, paired test, team test, group assessment, partner testing, or collaborative assessment. The search resulted in a total of two articles and two conference proceeding papers where group testing was used in mathematics courses. I provide a detailed review of each study and explain how my study will extend the results or add to this literature base.

The first study conducted by Berry and Nyman (2002) consisted of students enrolled in a four-week intensive mathematical modelling course at Alma College[5]. The students were organized into teams of two students for the duration of the course. Assessment methods for the course included active participation in class discussion, oral presentations, one team test, one individual test, and a poster session. Similar to my proposed study and contrary to the studies outside of mathematics, the students took the team test before the individual test in this course. The team test centered on the students' understanding of the major concepts of the mathematical modeling process and was implemented to examine their problem-solving skills. Students took the test with their assigned partner in a separate room where they were free to discuss the questions and format their collective response. Student feedback regarding the team testing method was collected with a post-test survey which contained open-ended questions. In response to the questions that asked students how they felt about the formal team test, students' responses indicated they were generally positive about their experience. The students

---

[5] Information about the course topics are provided for each week along with details of each assessment method used. However, the authors do not give information about the students in the course. For instance, the number of students enrolled, gender of students, majors of students, and any other characteristics of the students in the course are not included in this paper.

commented on the balance of the weaknesses and strengths between them and their team members (Berry & Nyman, 2002).

The second article was a practitioner piece by Goetz (2005) in which he expressed a perspective similar to mine with regard to the balance needed between the learning environment and the assessment environment. Goetz (2005) explained the use of group cooperation time during the final exam with the rationale that "since group work has been such an important component of their semester's work, it is reasonable to expect students to do some form of group work on their final exam" (pp. 12-13) in his course with a high school precalculus course. Therefore, since students in his class were working in either cooperative or collaborative groups throughout the semester, he made a portion of the final exam a group exam where students worked collaboratively[6] to solve a problem. This group portion of the final exam consisted of one question that was worth 25% of their final exam grade. The groups were semi-randomly assigned with the intent to maintain as many heterogeneous groups as possible. Goetz (2005) did not state the characteristics he uses to make the group heterogeneous (i.e., gender, ethnicity, or achievement). However, he does state that the students did not know their group assignment until they arrived at the final exam. Pairing students who may have not worked together throughout the semester may be seen as a potential problem but the majority of the groups scored between 23-25 points for the group exam (maximum allotted points was 25 for the group portion). This could be explained by the nature of the course and the students as they

---

[6] Goetz uses the terms cooperative and collaborative interchangeably. However, with regards to these terms and how they are defined for this study, his use of group testing is collaborative as students must work together to solve a problem and only one answer per group is submitted for a grade.

were all in an honors precalculus course and destined for Advanced Placement calculus. Goetz (2005) noted the reduced pressure students felt during the final exam and increases in confidence. Specifically, he stated that “allowing students to work in groups, even when under the pressure of a final exam, can ease tension, buoy confidence, and get students off to a good start on the test” (Goetz, 2005, p. 17).

The third study was conducted in a New Zealand university where they adapted Team-Based Learning strategies from the medical field (Paterson et al., 2013). This strategy consisted of students in teams of five to seven working together throughout the semester in two different mathematics courses (i.e., Combinatorial Computing and Dynamical Systems) and one mathematics education methods course. In all courses the students took both individual and team tests similar to Cortrigh et al., (2003), Duane and Satre (2014), Eastridge (2018), Gilley and Clarkston (2014), Giuliodori et al. (2008), and Rao et al. (2002) where students first took an individual test and then a team test. Unfortunately, the testing method was not the focus of their disseminated results from Paterson et al. (2013) and the paper only provides a small amount of student feedback. However, Paterson et al. (2013) did note the increase in student communication and that students focused more on understanding and thinking instead of their individual grade.

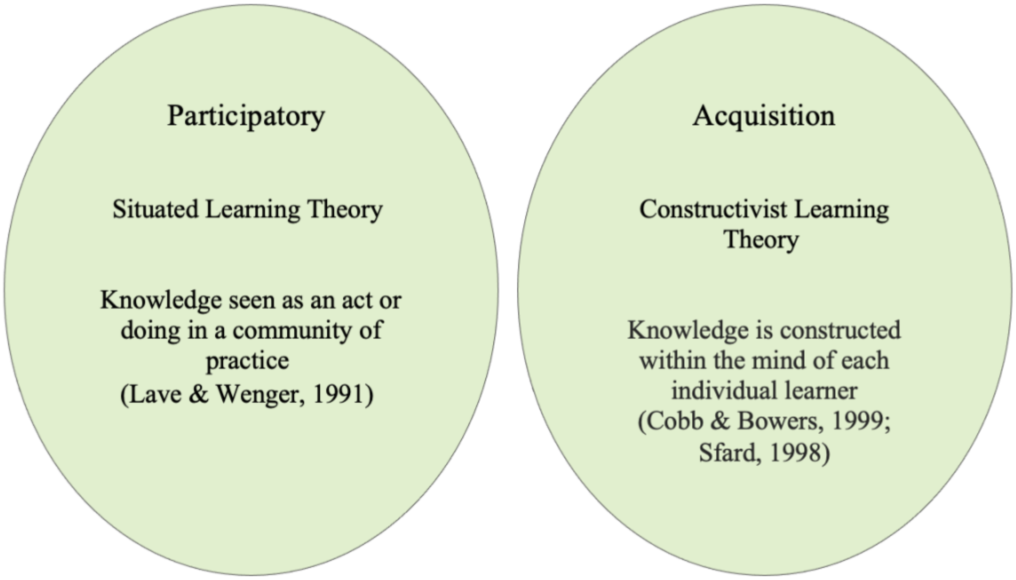
The final study was recently presented at the Research in Undergraduate Mathematics Education Conference in 2019 by Kelly MacArthur, a doctoral candidate at the University of Utah. During the presentation, she discussed her effort to rehumanize mathematics (Gutiérrez, 2018) for her Calculus II students by restructuring her assessment methods. The goal of her study is similar to mine in the aspect of having assessments that mimic how STEM students will work when in the STEM workforce (i.e., as scientists, engineers, and mathematicians). She used cooperative structure for the group exams as opposed to the structure of the tests for this study (see Chapter 3 for more details). That is, students were allotted 15 minutes to work on the problems for the group test first and then were given 40 minutes to discuss the problems and their solution methods in their groups (MacArthur, 2019). The study was conducted with a large number of calculus students (N =174) during the Spring 2018 semester at the University of Utah. MacArthur (2019) analyzed the test scores throughout the semester, which consisted of three midterm exams and one final exam; she also conducted both focus group and individual interviews of 18 students involved in the course. The analysis of the test scores indicated similar results as those from Eastridge (2018), Gilley and Clarkston (2014), Giuliodori et al., (2008) Srougi et al. (2013) and showed a gradual increase in student performance across the test scores from the beginning to end of the semester as well as the impact on both low- and high-performing students (MacArthur, 2019). One additional result her study added to the field was that statistically there was no difference across genders (MacArthur, 2019). Since both men and women benefited from the use of group tests in Calculus II, my study attempts to extend this result and the results from the other studies by noting the influence of this assessment method of group testing on women’s sense of belonging.

Theoretical and Conceptual Frameworks

Two theories of learning guided each aspect of this dissertation study (i.e., setup, course design, data collection, and data analysis) and provided the theoretical foundation for the study: situated learning theory and constructivist learning theory. The course in which the study was enacted was designed to use a combination of the two learning theories in a way that would meet the 10 Needs of the Learner enumerated by Sfard (2003): meaning, structure, difficulty, repetitive action, significance and relevance, social interaction, verbal and symbolic communication, well-defined discourse, a sense of belonging, and balance. Learning occurs under both models, though the theory surrounding the construction of knowledge differs with respect to where the knowledge is being constructed under each learning theory. Sfard (1998) metaphorically described students' development of knowledge under situated learning theory metaphorically as participatory and under constructivist learning theory as acquisition. In other words, under situated learning theory, knowledge is constructed socially through participation, outside the learner, and under constructivist learning theory, knowledge is constructed within the learner’s mind by acquisition. Consider Figure 2, which displays the metaphors, learning theories, and where knowledge is considered for each model.

Figure 2.

Theoretical Framing.



While learning occurs whether an instructor designs from a perspective of situated learning or constructivist learning, some of the Needs of the Learner are better met through one learning theory over the other. The course in which this dissertation study was enacted was designed to combine Situated and Constructivist Learning Theories in a way that was hypothesized to best meet the Needs of the Learner. The remainder of this section provides a brief description of the two learning theories, followed by a description of the Needs of the Learner, including information on the Learning Theory best suited to meet that need. Finally, this section ends with my conceptual framework in which I provide an arrangement of the 10 needs of the learner with respect to the two learning theories.

The acquisition perspective is influenced largely by Piaget’s (1953) and Vygotsky’s (1962) work regarding learning. Knowledge in the acquisition metaphor can be thought of as something learners *acquire,* and learning is considered the act of acquiring and accumulating knowledge. It is important to note, in the acquisition metaphor, knowledge is constructed *within the mind of each individual learner* (Cobb & Bowers, 1999; Sfard, 1998). In the acquisition metaphors, the individual student is the basic unit of knowing and learning (Paavola & Hakkarainen, 2005). That is, learning and knowledge can be considered very personal (i.e., different from one learner to another) and is contained internally (i.e., in one’s own mind) when we consider the acquisition perspective. Sfard (1998) warned that subscribing to this metaphor for learning alone can lead to the ideas of the student as a *passive learner* and of knowledge as something that is transmitted from the instructor to the student.

The participatory metaphor Sfard (1998) described as “learning a subject is now conceived of as a process of becoming a member of a certain community” (p. 6). The participatory perspective is influenced by the social learning theories in which knowledge is not viewed as an acquired commodity, but instead knowing is seen as an action or doing within a community of practice (Lave & Wenger, 1991). In this perspective, knowledge can be co-created among the members of the community of practice and refined within that community during specific interactions within a specific context (Lave & Wenger, 1991). In the participation metaphor, the community (group or class of students) is the unit of knowing and learning (Paavola & Hakkarainen, 2005). That is, learning is not equated to knowledge as something someone has (i.e., an acquired internalized commodity) but instead focuses on the act of learning through activities in a given context (Sfard, 1998; Lave & Wenger, 1991).

For this study, I consider both metaphors for learning, and consider knowledge as something that is both acquired in the individual learner's mind and co-created from participation within a community of practice. The largest difference between the two perspectives is regarding the emphasis and need of social interaction and where knowledge is constructed (i.e., in the mind of the individual learner or in the community). I consider learning as a whole, or holistically, as both an individual acquiring and constructing knowledge (cognitive-acquisition perspective; Cobb & Bowers, 1999, Sfard, 1998) and their participation within a community of practice (situated-participatory perspective; Cobb & Bowers, 1999, Sfard, 1998; Lave & Wenger, 1991). Therefore, I situate the study within the amalgamation of the participatory and acquisition perspectives for learning and each associated theory.

Expanding upon Sfard's (2003) presentation of the 10 needs of learners, I have constructed an organized model to represent these needs within the situated (participatory) and constructivist (acquisition) learning theories. This organization is by no means an indication of how these needs *must* be organized, but rather my interpretation of how the needs can be organized to enhance teaching and learning for undergraduate mathematics courses with a focus on equity. When arranging these needs, I first considered how each one aligned the ideas of learning and knowledge to either an individual unit (i.e., acquisition) or a group unit (i.e., participatory): each need may be internally driven (i.e., for internal knowledge construction) or socially driven (i.e., for participation within a community), or a combination of both. I argue, however, the majority of the needs as defined by Sfard (2003) are *primarily* driven either only by one or the other. I classify the needs into three main categories: internal needs, social needs, combination needs. In the following subsections, I provided a description of each of the 10 needs of the learner as defined by Sfard (2003) organized by the three categories (i.e., internal, social, then combination). Within each description, I present my argument based on the theory of learning behind each need and classify them as either an internal need (i.e., based on constructivist theories of learning) or a social need (i.e., based on social views of learning or psychology) that must be met for each student for learning to occur. I then explain how these needs are arranged with respect to the participatory and acquisition metaphors for learning (i.e., the conceptual framework) and how this arrangement is used throughout the study.

Internal Needs

The internal needs of the learner correspond more to the acquisitionist view of learning by which knowledge is adapted, processed, and constructed inside a learner's mind (Sfard, 1998). Constructivist views of learning attend to the internal needs of the learner (i.e., meaning making, structure, repetitive action, and difficulty). Either Piagetian or Vygotskian, or both, perspectives of constructivism influence each of these needs. The four internal needs each student has in order to construct mathematical knowledge are a need for (1) meaning, (2) structure, (3) repetitive action, and (4) difficulty. These internal needs may be met by different means for each learner, and each of these needs are described further below.

Meaning

Meaning making is an *internal process* used by all learners to understand and make sense of the world around them (Piaget, 1953; Sfard, 2003). Piaget described this as a natural need for learners to make sense of the world around them. As learners construct their own conceptions about the world, however, they may develop alternative conceptions that can be difficult to change (Piaget, 1953; Sfard, 2003). This leads to the need for meaning making with communication, as Vygotsky further connects the need for meaning making with communicating of one's experiences to others (Bruner, 1997). All learners have this need to make meaning of the world around them and communicate their meaning of the world to others (Bruner, 1997; Sfard, 2003; Vygotsky, 1962). Often, in education, the communication of the *meaning* someone has created has been translated into individual testing. For example, students in mathematics can communicate the meaning they have constructed internally through answering problems on a test. Teachers can then formally assess each individual student's constructed meaning of mathematical concepts for correctness and possible alternative conceptions (through student's wrong answers). This view of meaning as internally constructing knowledge about the world around them and assessing one's internally constructed knowledge aligns with the acquisition view of learning.

Structure

The need for structure incorporates both the Piagetian notion of reorganizing mental schemes when acquiring new information and Vygotsky's hierarchical organization of new concepts (Bruner, 1997; Sfard, 2003). As mathematics itself can be described as a well-structured discipline, the learning of mathematics can be accomplished through connections of previously learned concepts. Sfard (2003) stated how this particular need "in mathematics it may be the very essence of learning" (p. 360) as the need for structure incorporates what students already know (i.e., previous or prerequisite mathematical knowledge) and connects this knowledge with new ideas and concepts. This view of structure is an extension of the need for meaning and builds on one's internally constructed knowledge, and therefore also aligns with the acquisition view of learning.

Repetitive Action

Again, both Piaget and Vygotsky saw the importance of *action* for learning and understanding (Sfard, 2003). Piaget suggested that knowledge is constructed through internalization of one's *actions* on objects (Bruner, 1997; Piaget, 1953; Vygotsky, 1962). Additionally, Vygotsky suggested the internalization, or mental functions, are created through *activity* (Kozulin, 1990; Vygotsky, 1962). In mathematics these objects can be tangible, such as manipulatives or intangible, such as numbers, functions, or abstract sets (Sfard, 2003). Through the process of repetitive action with mathematical ideas learners can reorganize their internalized knowledge (i.e., structure) to add the new mathematical object into their current knowledge schema (Piaget, 1953). This view of repetitive action is an extension of the need for structure, as a way of organizing new meaning, and rectifying one's understanding (Sfard, 2003) which considers knowledge to be something that is internally constructed. Therefore, repetitive action aligns with the acquisition view of learning.

Difficulty

The process of moving an idea into an object of mathematical knowledge requires learners to struggle (Sfard, 2003). Sfard (2003) stated that "true learning implies difficulties" (p. 366), what these difficulties are, however, is not the same for all learners. Therefore, learning requires a level of difficulty unique to each learner depending on previous constructed mental schema (Piaget, 1953). This corresponds to Vygotsky's Zone of Proximal Development (ZPD) in which each student has already built a certain level of ability and learning occurs inside of their ZPD with assistance (Vygotsky, 1978). If students do not exhibit any difficulty, then they are not learning new knowledge, but simply applying knowledge they have already obtained to the current situation. If the problem is too difficult for the students, however, the ideas are outside their ZPD and they cannot use their current knowledge or understanding to build new knowledge (Vygotsky, 1962). This view of difficulty incorporates one's internally constructed knowledge and cognitive abilities, which aligns with the acquisition view of learning.

Social Needs

The social needs of the learner correspond to the participatory view of learning according to which knowledge is not a tangible object obtained and held within a learner, but learning is a process of participation and obtaining membership in a certain community (Sfard, 1998; Lave & Wenger, 1991). Multiple learning theories that consider participation in a community for the learning of mathematics (i.e., sociocultural, social constructivism, situated learning) attend to the social needs of the learner (i.e., social interaction, verbal and symbolic interaction, well-defined discourse, and sense of belonging). For example, social interaction is an important component for learning in both social constructivist and sociocultural learning perspectives (Sfard, 2003; Lave & Wenger, 1991). Although social constructivism and sociocultural theory do influence many of these needs, separately they do not explain the importance of all of the social needs of the learner. The four social needs that each student has in order to construct mathematical knowledge are a need for (1) social interaction, (2) verbal and symbolic use, (3) well-defined discourse, and (4) sense of belonging. As with the internal needs, the social needs may be met with different means for each learner, and each are further described below.

Social Interaction

Social interaction has not been disputed as an important aspect of learning in mathematics education for over twenty years (Cobb, 1995; Cobb & Bowers,1999; Forman et al., 1998; Lampert, 1990; O'Connor & Michaels, 1996; Schoenfeld, 1996; Sfard, 2003). Sfard (2003) broadly defined a social interaction as an exchange of ideas between an individual and another person via written or verbal communication. Being able to communicate one's knowledge was one of the important aspects for meaning making as discussed by Vygotsky (Sfard, 2003; Vygotsky, 1962). Even Piaget noted a benefit of social interaction to a learner when constructing internalized knowledge, although he did not claim it was essential for learning (Sfard, 2003; Piaget, 1953). Vygotsky (1962), however, did suggest that social interaction is an essential part of learning especially for one to obtain a conceptual understanding of objects or ideas.

In the field of cognitive psychology, which is grounded in the work of Piaget and Vygotsky, gain of conceptual understanding is explained through cognitive conflict (Springer et al., 1999). When students learn with and from one another they must explain their reasoning to a peer. If the reasoning presented does not align with their internal meaning constructed previously, this could cause a cognitive conflict for the student. Springer et al. (1999) explained that a gain in conceptual understanding is due to the cognitive conflict created in the student's mind, and when this occurs, the student must correct their reasoning. Therefore, social interactions are a necessary requirement for learning because these interactions provide the space for students to challenge their preexisting notions about mathematical ideas. Additionally, Lave and Wenger (1991) suggested that social interaction is essential for one to *show knowing* by *doing* within a community (i.e., situated learning theory). Therefore, the need for social interaction aligns with the participatory view of learning.

Verbal and Symbolic

Social interaction and being able to communicate one's ideas and knowledge require learners to use both verbal (i.e., talking and listening) and symbolic (i.e., reading and writing) representations of knowledge and thoughts (Sfard, 2003). This idea that knowing comes from having a word or symbol is grounded in discursive psychology (Edwards, 1993; Edwards & Potter, 1992; Harre & Gillet, 1995). Sfard (2003) described the importance of either a verbal or symbolic use to give meaning as “one just cannot construct the meaning of a concept before introducing a word or a symbol with which one can think about that concept. The sense of *understanding* [emphasis added] then develops through the use of the word or symbol” (p. 374). Therefore, verbal and symbolic representations are a need of learning mathematics because these representations provide the tools for social interactions to occur, and aligns with the participatory view of learning.

Well-defined Discourse

The need for well-defined discourse follows from the needs for social, verbal, and symbolic interactionism. The term discourse itself is defined in the dictionary as a verbal interchange of ideas (Merriam Webster, 2021). Discourse in mathematics becomes well defined when sociomathematical norms centered around what is considered mathematical argumentation are established in the classroom (Yackel & Cobb, 1996). Sociomathematical norms are beyond regular classroom norms, in which the teacher must also establish norms for both the quality of mathematics and mathematical discourse, and are not the only requirement for creating a learning environment that includes mathematical argumentation (Yackel & Cobb, 1996). A classroom where the norms are established, and students actively participate in mathematical argumentation is essential for high-level thinking of mathematics (Forman et al., 1998). Engagement in mathematical argumentation requires students to have intellectual autonomy (Yackel & Cobb, 1996). Specifically, students would need to understand their own “mathematical capabilities” (Yackel & Cobb, 1996, p. 473) as they judge and decide what constitutes a mathematical argument of another student. This notion of discourse in the classroom among students includes both social interaction and the use of verbal or symbolic representations of knowledge, and aligns with the participatory view of learning.

Sense of Belonging

A sense of belonging is a human need to feel valued as a member of a community. A student’s sense of belonging can be defined as their “sense of being accepted, valued, included, and encouraged by others (teachers and peers) in the academic classroom setting and of feeling oneself to be an important part of the life and activity of the class” (Goodenow, 1993, p. 25). It is important to note that a sense of belonging is context specific, a student may feel they belong in one group in a certain class and not in another (Strayhorn, 2019), and can affect both men and women in an academic setting (Good et al., 2012; Strayhorn, 2019). Ultimately, if students do not feel they belong in a field of study, classroom, or group they are less likely to continue or engage with that group (Cheryan et al., 2009; Strayhorn, 2019), and thus makes this is an important need that must be addressed for all mathematics learners (Sfard, 2003). This need aligns with the participatory view of learning because it includes a community or members outside of oneself.

Combination (Internal and/or Social) Needs

Recall the harm of picking just one metaphor for learning (Sfard, 1998), and consider that learning requires both acquisition of knowledge and participation in a community of learners (i.e., to acquire the knowledge). The last two needs of the learner can either be aligned with one, the other, or both metaphors depending on the student (i.e., significance and relevance), or must consider both the participatory and acquisition metaphors for learning (i.e., balance).

Significance and Relevance

Significance and relevance are combined as one need, and is the need that drives the motivation for the learner (Sfard, 2003). If we consider the ideas from Piaget where new knowledge can only grow out of existing knowledge, then learners must internalize the significance of what they are learning as they reorganize their mental schemes (Sfard, 2003). If we consider instead the importance of social interaction in learning, the significance and relevance of new knowledge can come through social interactions and participation within a community (Lave & Wenger, 1991). Because the significance and relevance can be formed/made either internally by one's own drives and thoughts, or influenced by their peers in their learning community, this need is placed in the middle of the internal/social needs.

Balance

The last need may seem obvious now and yet may be the hardest one to meet: the need for balance. Sfard (2003) suggested "to meet learners' multifarious needs, the pedagogy itself must be variegated and rich in possibilities. The learning individual is a complex creature with many needs that must all be satisfied if the learning is to be successful" (p. 384). Educators must help create learning environments that allow students the opportunity to meet each need through equitable instruction (Sfard, 2003). Therefore, equitable instruction refers to the instructional techniques used to support the development of the ten needs for *all* students and promote a balance among them. For instance, collaborative group work is an equitable instructional practice that if implemented with fidelity, students are given opportunities to explain their thinking and justify their mathematics using both verbal and written methods (Perry, 2018).

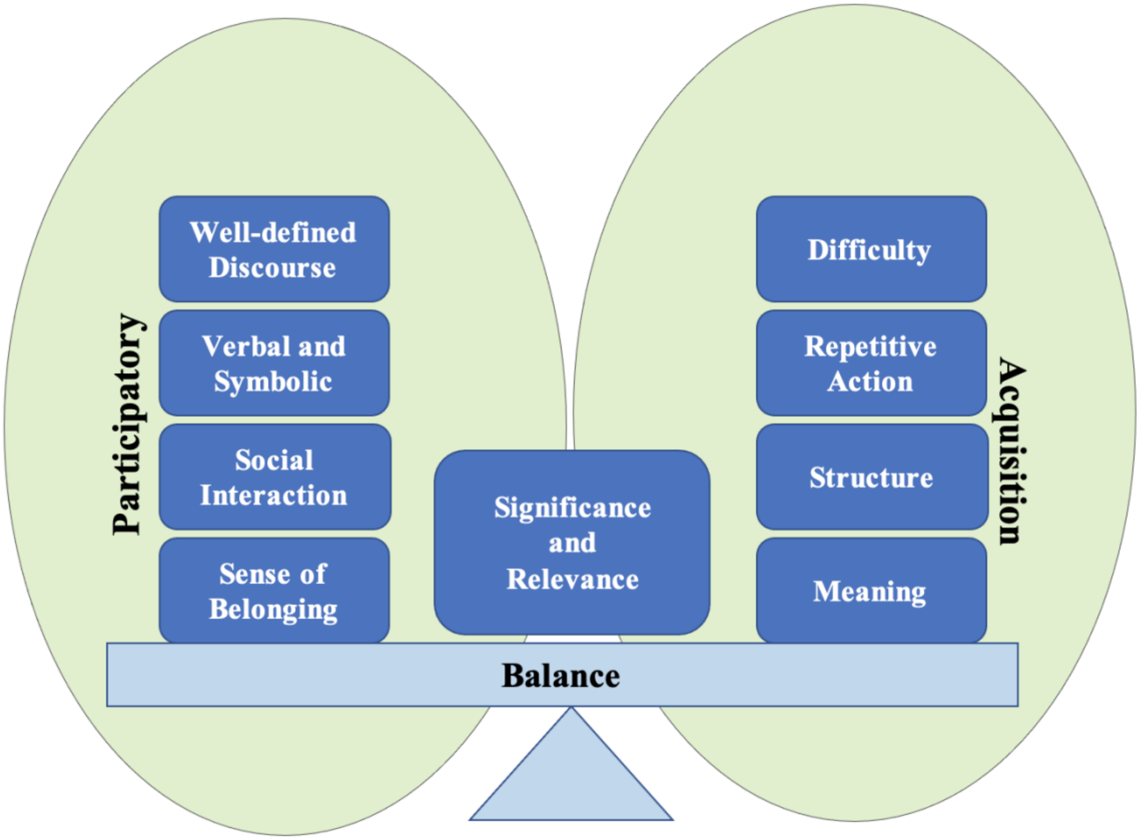
Balanced Learning Needs Framework

The conceptual framework (Figure 3) that guided this dissertation is my conception of the arrangement of the 10 learning needs described above how each of these needs align with either the participatory or acquisition metaphors for learning. The arrangement of the needs is placed either on a participatory side (i.e., left) or acquisition side meaningfully (i.e., right); the vertical placement, however, is not meant to imply any order among those needs placed on each side. For example, well-defined discourse was not meaningfully placed above verbal and symbolic communication. The placement on the left (or right) only implies these constructs lean toward the participatory metaphor (or acquisition metaphor) for learning.

Sfard (2003) herself indicated that not all of the needs attend to both the participatory or acquisition metaphors for learning. For example, when she described the need for well-defined discourse and the rules (i.e., sociomathematical norms) of such discourse Sfard (2003) stated "the need to know the rules of the discourse to be mastered is a participationist *counterpart* [emphasis added] of the acquisitionist need for structure" (p. 377). While she explicitly stated these two can be seen as "counterparts," it is important to note that the counter balance is at the participatory and acquisition levels and not at the level of each individual need. In addition to the counterbalance of the learning metaphors, the nine learning needs and balance (i.e., the tenth learning need) can be considered a social need, individual need, or both. Social needs of the learner include an aspect of socializing with others (i.e., specifically concerning knowledge as the participation in a community), these needs are: well-defined discourse, verbal and symbolic, social interaction, and sense of belonging. Internal needs of the learner include a cognitive aspect of learning (i.e., specifically concerning knowledge construction in one's mind), these needs are: meaning, difficulty, structure, and repetitive action.

Figure 3.

Balanced Learning Needs Framework.



While each of these needs described above hold equal importance for learning, my study focused on the sense of belonging and social interaction needs specifically for the women participants. Women generally have a lower sense of belonging to mathematics; therefore, one can hypothesize that if women were to strengthen the ties between their sense of belonging and communication, they would feel a stronger connection to their mathematics community.

Chapter Summary

In conclusion, in this chapter I provided an overview of the relevant literature concerning women in mathematics, undergraduate calculus reform, equitable education techniques, and the use of group assessments in mathematics classrooms. I also described how the literature has been used to inform my study and where my study is situated within this literature. Additionally, by investigating an assessment method that mirrors student-centered learning and seeing the effects on women's sense of belonging, this study may offer an additional method for retention of women in the STEM pipeline during and past Calculus I. Moreover, this study provides a connection of the four reverent areas reviewed in this chapter: women in mathematics, calculus reform, equitable education techniques, and group testing. Finally, I concluded this chapter with a description of the conceptual framework which guided my study. The framework was built around my view of what it means to learn mathematics and guided the design methods, course implementation, and data analysis presented in the next chapter.

CHAPTER THREE: METHODOLOGY

Introduction

The purpose of this dissertation was to investigate women's sense of belonging to a mathematics community and their social interactions in a reformed Calculus I course while engaged in group testing. Recall, sense of belonging refers to the feeling of being a contributing and respected member of a certain group in a given context. To investigate the feelings of students I collected more than simple quantifiable measures. That is, I needed to understand how the students felt during certain interactions by asking the students about their feelings. When McLeod (1994) summarized 25-years' worth of research published in the *Journal for Research in Mathematics Education*, he discussed not only the results of the summarized studies but also their methodological choices and concluded:

> If we are to study affective issues in the context of the reform movement in mathematics education, we must choose methods that will help us capture the complexity of the issues. Therefore, we must build our research studies on methodological and theoretical foundations that are broad and open enough to allow for this kind of complexity. (p. 644)

Keeping McLeod's advice in mind, I used an exploratory descriptive mixed method design. Therefore, I collected both qualitative and quantitative data at multiple time points throughout the study to answer my research questions. In the following section, I provide an overview of the research methodology I used in the study.

Research Overview

I used an exploratory descriptive mixed methods design to investigate female students' sense of belonging and their social interactions with group members in a Calculus I reform class. The social interactions with their group members were examined during group testing environments. The sense of belonging was examined through the female students’ responses to questions about their experience during the group test in the form of a post-test reflection and quantitatively through their sense of belonging measured by a validated and reliable survey described further in this chapter. The students took a group test and responded to post-test reflection questions twice throughout the course. Their quantitative measure of their sense of belonging was obtained on the first and last day of class. The exploratory descriptive mixed method design allowed me to explore the viewpoints of multiple participants involved in the same experience (i.e., a group test) while collecting both quantitative and qualitative data sources for their sense of belonging measure (Patton, 2015), which will be explained in detail in this chapter.

Research questions include the following:

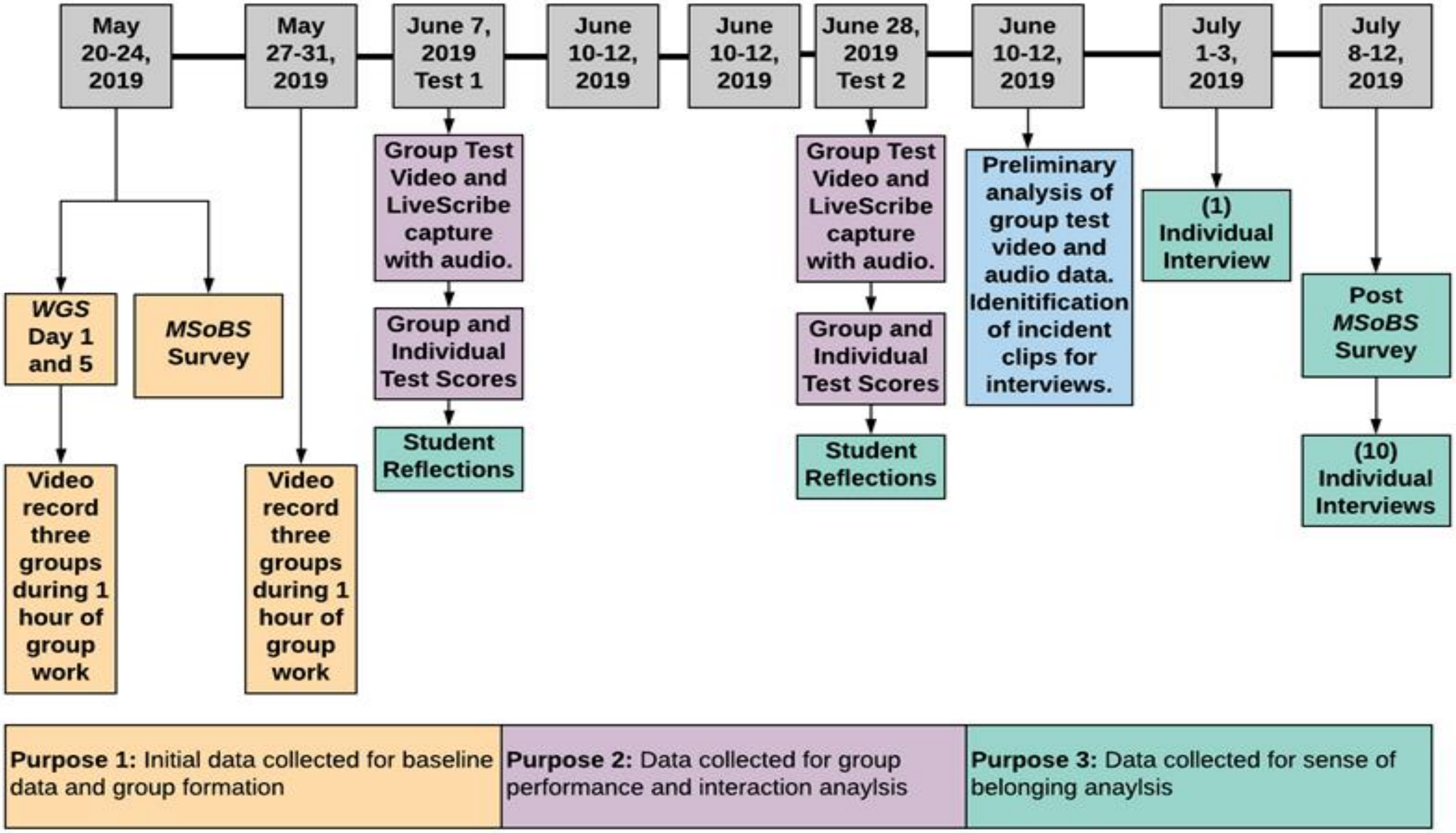
1. In a reform Calculus I course, how do students’ performance on group tests compare with their performance on individual tests; and how do gender, and the intersection of race and gender contribute to the differences, if at all?
2. How does the sense of belonging to a mathematics community change for female students after engaging in group testing in a reform Calculus I course, if at all?
3. How are social interactions within the group composition different based on the composition of the group, if at all, and can any differences be used to explain differences in the female students' sense of belonging?

Regarding student performance, I focus on 30 students who were arranged into eight permanent groups throughout the course. These groups contained both male and female students in different arrangements and are discussed later in this chapter. To examine students' social interactions during the group test environment, I focused on three groups in the course to which I will refer as the research groups. Regarding the sense of belonging, I focused my study on the 15 female students in the course.

Both qualitative and quantitative data were collected multiple times throughout the study. Not all of the collected data, however, were used to answer the research questions. Altogether, I collected data for three main purposes: The first purpose was to place students in their ideal permanent working groups; The second purpose was to explore the performance and interactions of students engaged in a group test; The third purpose was to understand the possible effects of group composition and engagement in the group test had on the sense of belonging for the female students in the course. See Figure 4 for the timeline of my data collection.

Figure 4.

Data Collection Timeline.



The quantitative data included students' individual test scores and group test scores, and students' pre and post sense of belonging scores. The research design allowed me to also collect qualitative data to better understand the quantitative measures of both the groups' test scores and the students' sense of belonging measure. The qualitative data for this study include video recordings, audio recordings, and a Livescribe captures (described further in this chapter) of students engaged in the group test for three of the research groups, and individual interviews with the eight female students in the three research groups. In this chapter, I first provide my personal motivation for this study in the form of my story as a woman in mathematics. I provided my story as I myself and my experiences informed the motivation, design, and data collection of the study. Additionally, my story is provided as a means for the reader to understand the lens through which I interpreted the results, and to be as transparent as possible in effort to build trustworthiness. Next in this chapter, I explain my data collection and analysis for each of the three purposes (i.e., formation of groups, data collection for group test analysis, data collection for sense of belonging analysis) outlined in Figure 4.

The remainder of this chapter is organized as follows, I first provide my motivation, then details on the context of the research including the university and the Calculus I course the study took place. Next, I provide details about the survey instruments used in the study. I then detail the three purposes of the data collection. Within my description of the first purpose, I provide a description of the three research groups and each of the participants. Finally, I conclude this chapter by addressing trustworthiness, limitations, and delimitations of the study.

Motivation

My story begins before I entered college. I went to an all-girls college preparatory high school where I excelled at mathematics but little else. I took five honors mathematics courses in my four years of high school by doubling up in my sophomore year. Due to my underperformance in English and history, however, and one sad attempt at art, my graduating GPA was under 3.0. At the end of my senior year, I went into the college counselor's office to ask about next steps. I don't remember much about this moment: I can't recall the specific person I talked to, the room I was in, or even what day it was. What I do remember is this person telling me "College isn't meant for everyone." It took me eight years after high school to matriculate officially into a four-year university. I recognize now that I entered with an already lowered sense of belonging as the comment from my past still haunted me.

Entering college, I was originally a business degree seeking student, as I was working at a finance firm at the time. I felt good about my mathematics skills and even placed into Calculus I through the placement exam, though I enrolled in the two-semester business calculus sequence as it was the only mathematics required for my degree. During the second semester course, my professor commented on my mathematical knowledge and skills. He also noted how students in the class came to me with their questions. At the time, I was even involved with a study group that met on the weekends. In this study group, I helped many of my fellow classmates with the course material. The professor of the course encouraged me to take a few more mathematics classes, even though he knew they were not required for my degree. I obliged as I enjoyed the material and enrolled in a precalculus course the next semester. I have often thanked this particular professor as his noticing and encouragement of me is what initially sent me on my path to become a mathematics major.

I finished the first precalculus course and again was encouraged to continue down the mathematics road by this professor. After the second precalculus course I took Calculus I without encouragement, but now I had an internal motivation to keep learning. During this calculus course I did not have the same experience as the past few mathematics courses I had taken. Instead, I did not like the classroom environment nor the method of instruction and often skipped class. Everything felt different. The instructor just delivered material to the class, assigned homework from the book, but rarely handed back any graded work, and never gave feedback on why things were wrong. Even though I was rarely in class, I did show up to tests and to hand in homework: that was about it! Although my behavior and presence in the course was different from the previous few mathematics courses I had taken, this instructor recommended me to the department to take Honors Calculus II the next semester. This was now the third instructor who had taken notice of my mathematical skills. After this, I heeded the call of mathematics and switched my major to mathematics.

I have many more encouraging stories about my professors like these above. The tone shifts, however, when I reflect on my interactions with my peers in my courses after I officially became a mathematics major. There are three main incidents that always come to mind. First, while in Complex Analysis, a junior level undergraduate course, I had one male acquaintance in the class. Like many of my classmates, he was ten years younger than I, and was very talented in mathematics. One day I was sitting in a coffee shop on campus reading the textbook before class and he happened upon me. He gave me a curious look and said “Wow, you actually read the textbook?” I wasn’t sure how to take this comment and it made me feel uncomfortable, so I discussed it with him later that same week. I told him the way it came across was that he was surprised and repulsed at the idea that I read the book. He responded that he thought I was smart enough to not have to read the book. This comment felt like an attack on my knowledge and skills, and it felt like he disregarded the fact that I was a top performer in the class because of the implication that I had to work harder to understand the material.

The next incident occurred during my first semester of graduate work in mathematics. I was taking a particularly challenging course, Algebraic Coding Theory, in which all the students were having a hard time understanding the material. I walked into the classroom one day and three of my male peers were discussing one of the homework problems at the whiteboard in the front of the classroom. I recall walking in and thinking, “oh good! I wasn’t the only one confused by this question.” I said something to that effect and the three turned around, looked at me, didn’t say a word, and turned back to the whiteboard talking to each other about the problem. I felt as if I wasn’t allowed to be a part of their conversation. Reflecting on my graduate experience, I realize this was the moment I started to pull back more in class and was not as open about sharing ideas or answering questions during whole class discussions. Although there were also personal difficulties in my life at the time, that moment also contributed to my lack of classroom interaction: the exclusion from their community hit me hard. The professor did notice my lack of involvement and expressed his concerns. One time he even specifically addressed the entire class and said to us if we ever need someone to talk to know his door is always open. Years later, I was working with one of the men who was at the white board and this story came up in conversation. At this time, we had both graduated and were now working at the university. I explained a bit of the missing information about my life at the time and he explained “Oh, you’re why he said that!” He explained how he and his friends in the class thought the professor’s comment seemed out of place at the time.

The last incident occurred during my final semester of graduate work. In a Real Analysis class, there was a problem on the board that contained a difference of cubes. I recall saying “we can factor this” and quickly stated the factored form of the polynomial. The class got quiet, and the professor looked at the board as if to say “okay” and then wrote the factored form. He then addressed the class, as if to alleviate their worries that they did not know how to factor these, and said “she tutors precalculus students, that’s why she knows that.” This was the same professor I had had for both Honors Calculus II and Differential Geometry – an upper-level course I had gotten permission to take in my sophomore year without the prerequisite courses. He was well acquainted with how I thought about mathematics, at least I thought he was at the time. I was left unsure about and discounted my own mathematical knowledge after this incident. I had been working in mathematics for eight years at this point and felt as if tutoring lower-level mathematics students was privileged over the work I had done to master mathematics.

During my undergraduate experience, I felt very connected to the mathematics and had internal motivation to learn more because I enjoyed the subject. However, I never felt a part of the mathematics community in the classroom. Although I would ask questions in class directly to the instructor, I preferred not to talk to my peers. Throughout the beginning years of my undergraduate education, I remember having a distaste for group work, especially in my calculus courses. I was one of the few women in the class, older than all of the other students, and a top performer on assessments. I remember thinking, “why do I have to turn and talk to my neighbor?” I knew the mathematics, I knew the best way to get an answer quickly, why did I have to talk to them? I also did not socialize with my peers outside of class. They would form study groups and hang out in the tutoring center and I would isolate myself in the library or at home. They did not make an effort to include me in their studies, but I did not make the effort either. While in class, I would become irritated during group work as my partners would often convince me their answer or method was correct. I thought I simply did not have the confidence to defend my answer.

It was not until my experience in upper-level mathematics courses that I understood why engaging with my peers and communicating ideas about the mathematics we were learning was so important. At this point, the mathematical content was very intense, and problems would take hours, if not days, to solve. The instructors were not able to have the traditional formal assessment of a timed test. Instead, we had to work in groups to investigate problems, some of which we did not have the tools to solve and present our investigation methods and findings to the class. This presentation, often accompanied by an individually written paper, was our new formal assessment. All of a sudden, my grade was no longer based only on my individual performance, and I had to learn how to convey my ideas to my peers. This was when I had to learn how to communicate my mathematical ideas. At first, I found myself trying to bypass the conversation with a peer and go straight to the instructor. I quickly learned, however, I needed to talk through my ideas in this collaborative investigation with my partner. At this point I made my first (male) friend in a mathematics course.

Fast forward to the end of my graduate school experience, I was working with some of my peers in the computer lab and we began talking about our past experiences. I was telling one of the men how I was not even sure of everyone's name in class during those early years, and he said to me, "yeah, we just thought you were a loner." I explained to him that I wanted to be a part of the group but felt left out. It was at that moment I felt as if I created my own isolation due to how I felt and behaved in the classroom during those early courses. Then I realized, it was during calculus that I did not feel a part of the community, and that feeling of isolation lasted for years. It did not just make me a "loner" but made others perceive that I did not want to be part of their community.

Research Context

The study focused on students enrolled in a reform Calculus I course at a public Midwestern university during the 2019 summer semester for which I was the instructor in the classroom. I taught the course for two purposes: (1) to serve as a role model for women in mathematics to help reduce the stereotype threat on the women in the classroom (Marx et al., 2005; Shapiro & Williams, 2012), and (2) to ensure the student-centered activities were implemented as intended. To help reduce researcher bias, the Associate Dean of the College of Science and Health Professionals served as the instructor of record. Along with assigning final pass/fail grades for each student in the course, the dean graded all of the students' tests (i.e., individual, group, and final exam). There were two individual tests, two group tests, and one comprehensive final exam. The tests were de-identified before they were sent to the dean to grade in order to preserve the anonymity of the students because most of the women in the course were participants in the study. The following two sections provide more detail about the university and the Calculus I course.

University

The university where the study took place is a public university located in the Midwestern United States. This university has a total population of 17,730 undergraduate and graduate students with a diverse student body of 27% non-white or URM groups and 55% female.

Calculus I Course

The study took place in a reform Calculus I course during the summer of 2019. This course was created specifically for minoritized students majoring in STEM who entered college below the calculus level. The program, Operation STEM (OpSTEM) funded through National Science Foundation Grants No. DUE-1161152 and HRD-1304371, started in 2013 and the program contained five of the seven retention methods outlined for successful calculus programs by Bressoud and Rasmussen (2015): examining pass rates and retention rates to guide future reform, coordination of precalculus and calculus instruction, maintaining academically challenging and engaging courses; student support services for at-risk students, and use of student-centered pedagogies and active learning strategies (Carver et al., 2017). The course was held for students who were either OpSTEM Scholars and completed Precalculus I and II with a B or better, or scholarship students who applied to take the special course. Additionally, because the design of the OpSTEM program was created with the goal of diversifying the STEM workforce, the eighth characteristic diversity, equity, and inclusion practices (Hagman, 2019) were inherently included in both the program and design of this specific Calculus I course.

The Calculus I course ran for eight weeks with five three-hour classes each week (i.e., a total of 120 contact hours). During the course, students spent 50% of the time working individually and 50% of the time working in groups in order to maintain a balance between a participatory learning and acquisition learning focused classroom. The breakdown of class time spent with students working individually to working in groups aligns with the conceptual framing of the study (i.e., the same amount of time aimed at meeting the students' individual and social needs for learning). Table 2 shows the percentage of class time spent in the following categories: lecture and whole class discussion, group work, project-based learning sessions, individual work time, review, testing, breaks, and time in a computer lab to work on online homework.

Table 2.

Course Time Breakdown.

	Individual Setting	Group Setting	Other
Lecture/Discussion	25.2%		
Group Work		29.1%	
Group Project		4.4%	
Individual Practice	9.4%		
Testing	3%	3.7%	
Computer Laboratory*			16.2%
Breaks/Transitions			9%
Total Percentage of Time	37.6%	37.2%	25.2%

*Note: During the computer laboratory time students were allowed to work alone or in groups, data was not collected for the students and their work preference during this time.

The context of this specific Calculus I course is important for this study, and consistent with the conceptual framework as there are both the individual and social settings for students to learn within a participatory and acquisition metaphor (see

Theoretical and Conceptual Framework in Chapter 2). Each individual class period (i.e., three hours) consisted of some lecture, student-centered activities including group work, time in a computer laboratory to work on the online homework system, or group tests and individual tests. Each component of every class period attempted to meet one or more of the needs of the learner. See Table 3 for a description of how each of the needs were met by one or more component of the course. This table offers brief descriptions of each course component contented with respect to the learning needs. After these descriptions, I supply the reader with detailed descriptions of each component of the course.

Table 3.

Learning Needs Matched to Course Component(s).

Learning Need	Course Component	Description of how the component of the course was designed in an attempt to meet each learning need
Meaning (Internal)	Lecture and Computer time	Allows students to work through problems and build their own internal meaning.
Structure (Internal)	Lecture and Computer time	Attempt to build structure in lecture by identifying connecting concepts from student's previous knowledge. Homework allows students to extend and organize new information to their previous knowledge.
Repetitive Action (Internal)	Lecture and Computer time	Individual practice time in lecture and the homework time on the computer allow the students multiple problems to practice.
Difficulty (Internal)	Lecture and Computer time	Problems in lecture and homework were designed with a certain level of difficulty to allow the students to productively struggle when building their knowledge.
Social Interaction (Social)	Group Setting and Computer time	Student-centered activities during group work required students to talk through problems and social interact to complete the learning activity. During the computer time students could choose to either work alone or with a peer allowing for more possible time for social interaction.
Verbal and Symbolic (Social)	Lecture and Group Setting	During lecture, after students initially explored a concept during group work, I would provide the students with the formal terminology and symbolic representations of the concepts. During group work, students needed to communicate with one another using both verbal and symbolic representations for the mathematics.
Well-defined Discourse (Social)	Lecture and Group Setting	During lecture, the instructor sets the sociomathematical norms and models mathematical discourse that is needed for the students to engage in mathematical argumentation. During group work, the students establish the group's norms and engage in discourse.
Sense of Belonging (Social)	Lecture and Group Setting	During the lectures, as the instructor, I encouraged all students to participate in whole class discussions and feel that they could share any idea, thought, or question. During group work, groups were created and monitored to ensure each student felt comfortable: (1) with their group members, (2) to voice their ideas, and (3) encouraged to share their ideas with their group.
Significance and Relevance (Combination)	Lecture, Group Setting, and Computer time,	Because this need is driven by the learner and can be met either while working alone or in a group setting, this need can be met for any learner within any one or all of the course components.
Balance (Combination)	The entire course	The instructor attempted to maintain balance by providing equal time for each need to be met in their designed course component.

SPT Selection and Training

Supplemental peer instructors have been shown to help students succeed in mathematics courses and can be considered student support services (Carter et al., 2016; Carver et al., 2017). In the Calculus I classroom for this study, I had four SPTs along with myself as the instructor full time in the classroom. Prior to the first day of class there was a pre-course meeting of the supplemental instructors (i.e., SPTs as defined in Chapter 1) and the director of OpSTEM, with me as the instructor of the class. The pre-course meeting included training for the SPTs on how to take classroom observations and key characteristics of student dynamics to identify while they were engaged in group work. Both my observations and the SPT observations during the first week of class time were used to form the permanent working groups. Because the SPT observations were a key part for the beginning of the study, in this next section I explain the training the SPT’s had to establish trustworthiness.

The SPTs for the course were selected from the current pool of undergraduate and graduate students who were (1) employed by OpSTEM, (2) had previously completed the OpSTEM precalculus sequence and Calculus I course, and (3) were chosen by the director of OpSTEM. The SPTs were in the classroom both to provide additional learning support of the students and to serve as research assistants for the study. During each class period, the SPTs were asked to keep an observation log for the day. Because observational skills are not necessary for supplemental instructors, prior to the beginning of class, I trained the SPTs on observation methods using a video of a middle school mathematics classroom.

Lecture Setting

Although active learning techniques are important for student learning, educators cannot discount the importance of lecture (Bransford et al., 2000). Using my conceptual framework as a guide, I attempted to maintain a balance between the participatory and acquisition learning metaphors in the class. Additionally, I attempted to meet each of the learning needs for the students within different aspects of the course. Therefore, the lectures were used to help students create meaning and structure by connecting the calculus concepts to previous concepts students had learned. The lectures often included previous material (i.e., algebra or precalculus concepts) needed for the given day's calculus content. Additionally, after students explored a concept, the class returned to lecture where I would explain the mathematical notation generally used. This part of the lecture setting was used to provide the students with the words and symbols used (i.e., to meet the verbal and symbolic learning need) and to establish the sociomathematical norms in the classroom (i.e., to meet the well-defined discourse learning need).

Student-centered Activities and Group Work

The majority of the calculus concepts in this course were first explored using student-centered activities. Students were arranged in groups of three or four and worked through activities designed by me and investigations designed by a mathematician. The group work setting was used to provide students with a place for social interaction (i.e., to meet the social interaction learning need). Additionally, the groups were created and monitored by myself and the SPTs to ensure students felt comfortable in their groups to voice their ideas (i.e., to meet the sense of belonging learning need). During the first week of the course, the sociomathematical norms were created as students were encouraged to engage with their peers during group work. These norms were created by myself and the SPTs as we monitored each group while they worked on the initial activities. Students were told to read each problem out loud and to discuss their possible approaches. When students felt their group was ‘stuck’, they were allowed to ask for help from either myself or an SPT. If groups requested help, in order to reinforce the norms of the classroom, we would ask each member of the group their ideas and indicate if one or more of their original ideas was correct. If there was no correct idea among the group members, we would pose purposeful questions to the group to help guide their thinking and collaboration. Finally, we would not spend more than five minutes with on group, as the majority of the time it took one or two questions before the group members would be able to continue working on the investigation or activity.

Students also engaged in project-based learning in which they explored the use of various calculus concepts through a project that spanned the 8-weeks of the course (Quinn, 2018). The project was used to help students see the connection of the course material to a real-life problem (i.e., to meet the significance and relevance learning need). During the time the SPTs and I monitored these activities, if the groups had questions, we answered their questions by first inquiring whether each individual group member had contributed their ideas to the group. The assurance that each member had contributed to the conversation was to ensure and encourage student-to-student communication while exploring the mathematics. If all ideas from each group member were explored and if the group still felt they were “stuck” on a problem, then we provided guidance and answered questions.

Computer Laboratory Homework

Students were required to spend a minimum of 30-minutes a day in a computer laboratory, located directly next to the classroom, to work on homework. We utilized the online homework system, MyOpenMath, which automatically grades the students' homework. This time was dedicated for students to work on the online homework problems specifically to practice the procedures of the concepts learned in class and allowed students to build meaning of the mathematics learned through the repetition of practicing multiple problems around one concept (Sfard, 2003). That is, the homework was set up in an attempt to meet the need of repetitive action for each student. During the computer laboratory time students were permitted to either work together or individually.

Testing

During the course there were three days that consisted of only testing for the student: on the third and sixth Fridays, the students took both a group and individual test. On the last day of the course students took an individual, comprehensive two-and-a-half-hour final exam. Time allotment on test days is displayed in Table 4 below.

Table 4.

Sequence and Time Allotment of Testing Days.

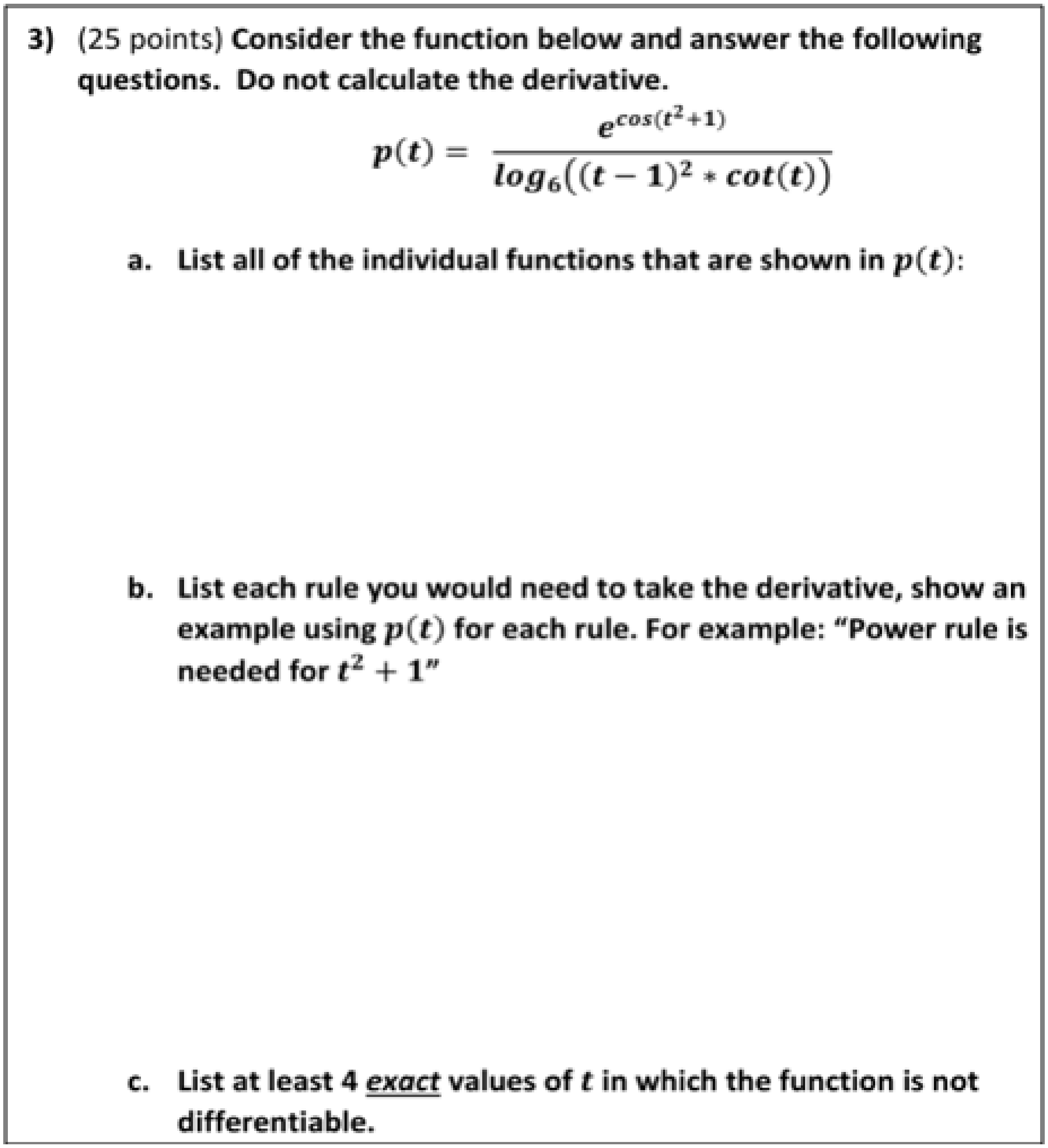
Sequence	Test/Data Source	Time Limit
1	Group Test	1 hour 15 minutes
2	Individual Test[a]	1 hour
3	Post-test Reflection	30 minutes

Note. Students will be given a 15-minute break between the group test and individual test.
[a] The individual tests will not be analyzed for this project.

Each group test consisted of four conceptual calculus problems and a bonus problem for up to 10 additional points. The individual tests covered the same concepts as their corresponding group test but were designed for evaluating the individual student's procedural understanding. For example, a group test included a question in which students identified the number of individual functions in a composite function problem and what they would need to know to take the derivative without actually taking the derivative of the function (see Figure 5). After the group test, the students took an individual test where they were asked to take the derivative of composite functions (see Figure 6). All tests were scored on a 100-point scale. Additionally, the first group test is provided for the reader in Appendix C: Group Test 1.

Figure 5.

Example Problem from Group Test 2.



3) (25 points) **Consider the function below and answer the following questions. Do not calculate the derivative.**

$$p(t) = \frac{e^{\cos(t^2+1)}}{\log_6\left((t-1)^2 * \cot(t)\right)}$$

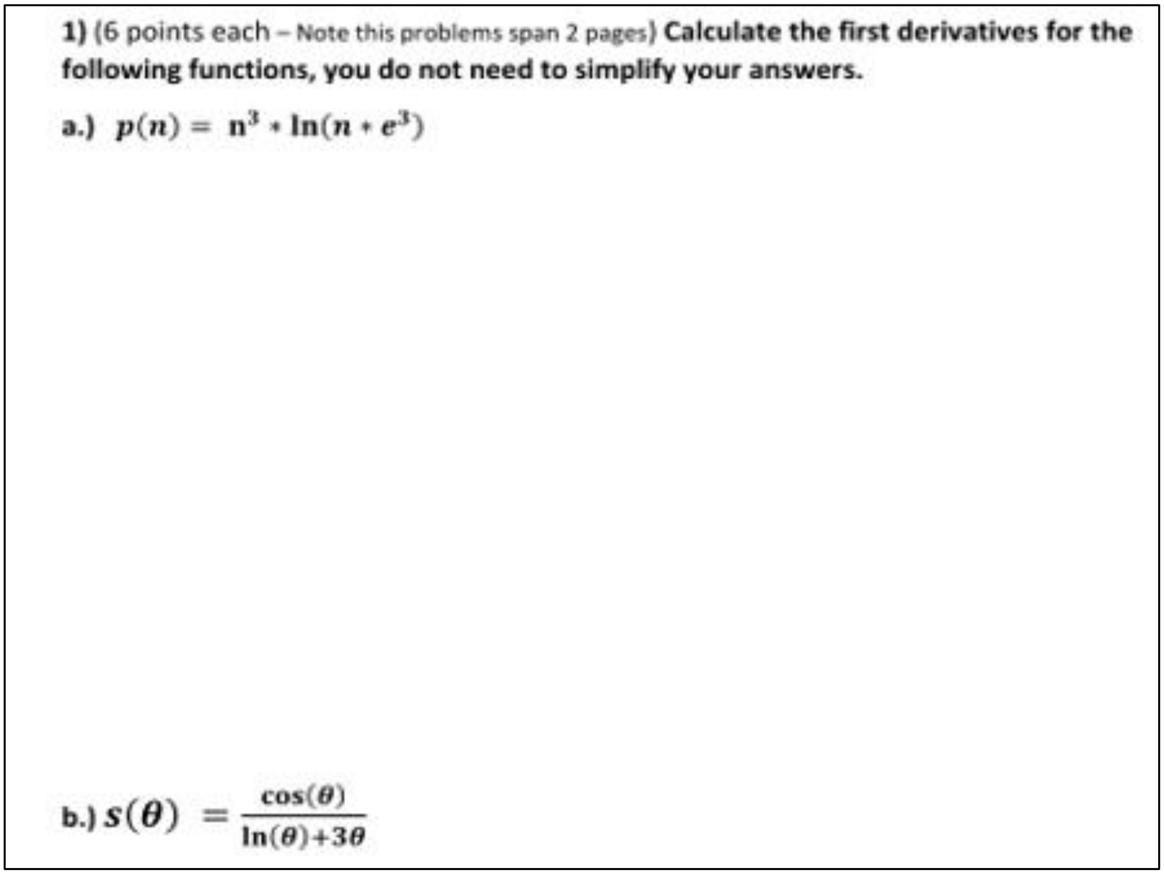
a. **List all of the individual functions that are shown in** $p(t)$:

b. **List each rule you would need to take the derivative, show an example using** $p(t)$ **for each rule. For example: "Power rule is needed for** $t^2 + 1$**"**

c. **List at least 4 *exact* values of** t **in which the function is not differentiable.**

Figure 6.

Example of Corresponding Problem on Individual Test 2.



1) (6 points each – Note this problems span 2 pages) **Calculate the first derivatives for the following functions, you do not need to simplify your answers.**

a.) $p(n) = n^3 + \ln(n + e^3)$

b.) $s(\theta) = \frac{\cos(\theta)}{\ln(\theta)+3\theta}$

Group Tests. The students took two group tests during the course. Each group was given one test, four calculators, and one pen. Each group member had to be the writer for one of the questions and only the writer of the problem could hold the pen during that problem. The choice of pen, and not pencil, was due to the design of the study and explained later in this chapter within the data collection description. Students were required to complete the group tests in their assigned groups and the group test occurred prior to an individual test during the same class period. On these tests, groups were required to justify their reasoning, understanding, and solutions in the space provided. To encourage students to voice their knowledge in conversation, no scratch paper was allowed and only one student at time could only write on the given test paper. If the group could not agree on a single answer, as may happen in a group engaged in mathematical argumentation, the member (or members) who disagreed with the final solution or justification were able to write their own explanation and reasoning on the back side of the response paper. Each group member who did agree with the group's represented solution had to initial in the agreement box for each problem when they agreed with the group’s solution/justification. If there was a disagreement in the group, students were also required to explain why the group did not come to a consensus. Most of the time students came to a consensus on their answers, there was only one disagreement that occurred during the course.

Individual Tests. Twice during the course students took an individual test directly after completing their group test. The individual tests occurred the same day as the group test and focused on procedural fluency, see Table 4 for the time allotted on the testing days. The individual test was designed to measure students' procedural fluency for topics discussed in the course. The last page of the individual exam contained post-test reflection questions regarding their experiences during the group test. The student post-test reflections are discussed in detail further in the next section. These data were used for both Purpose 2 (i.e., group test) and Purpose 3 (i.e., sense of belonging). In the following section, I provide a detailed description of my data collection, methods, measures, and analytic frameworks used to answer the three research questions.

Measures and Data Collection

Two survey instruments were used in this study: the existing *Math Sense of Belonging Scale* survey (*MSoBS;* Good et al., 2012) and the *Working Group Survey* (*WGS*) that I created in 2018 for my pilot study. All students in the course took the two surveys electronically via Qualtrics on the first day of class. All students completed the *WGS* again on the fifth day of class and the *MSoBS* survey again during the last week of the course. In addition to the two surveys, the data corpus included scores on group and individual tests, the transcripts of the three research during the two groups tests, and individual student post-test reflections. In the following subsections, I describe the elements of the data in more detail.

Math Sense of Belonging Scale Survey

Good et al. (2012) conceptualized a sense of belonging to an academic domain as "one's personal belief that one is an accepted member of an academic community whose presence and contributions are valued" (p. 701). Their research centered around the stereotype threats women in mathematics are impacted by, as well as the entity or incremental theories of intelligence of their instructors. I acknowledge that I have an incremental view of intelligence (i.e., I truly believe anyone can learn calculus), therefore this topic was not explored in this study.

In 2012 Good and colleagues conducted two studies to design and validate a scale to measure one's sense of belonging to mathematics. In order to measure a student's sense of belonging to mathematics, Good et al. (2012) created a 30 question 8-point Likert Scale survey aimed at five components of belongingness: membership, affect or feelings, acceptance, trust, and desire to remain in that domain (Good et al., 2012). See

Table 5 for the survey questions for each of the five components. Each question is preceded with the statement “When I am in a math setting” and asks participants to choose their degree of agreeance with each statement from strongly disagree (1) to strongly agree (8) (see Appendix A: Math Sense of Belonging Scale). One’s sense of belonging score is calculated by taking the average of the 30 responses. Higher averages indicate more responses on the strongly agree side (6-8) and lower scores indicate more responses on the strongly disagree side (1-3).

The validity and reliability evidence for the *MSoBS* survey was provided through two studies at a university in the Northeast United States with students who were currently enrolled in a calculus course (Good et al., 2012). The goal of the first study was to confirm the five-component structure of the *Math Sense of Belonging* measure. This was done using a total of 997 participants (465 men and 532 women), 90 of whom were removed from the study due to not completing the survey. Using exploratory factor analysis and confirmatory factor analysis resulted in substantial alpha (Cronbach’s $\boldsymbol{\alpha}$=.81 and Cronbach’s $\boldsymbol{\alpha}$=.78 respectively) on the five-components. The second study tested the reliability of the instrument and predictive power of the scale related to one’s intent to further pursue mathematics. This study consisted of 73 participants (30 men and 43 women) who returned to a laboratory twice to complete multiple questionnaires. The test-retest reliability was calculated using the average across the five components and had a correlation of .87 for the composite score. The predictive validity was confirmed after controlling for stereotype expectations and gender-based rejection for females.

The questions on the survey are broken down as follows: there are four questions related to membership, four questions related to trust, eight questions related to affect with four questions reverse coded, ten questions related to acceptance with five questions reverse coded, and four questions related to desire to fade with one reverse coded. For example, one of the positive coded affect questions asks the students “When I am in a math setting, I feel valued” and a negatively coded affect question is “When I am in a math setting, I feel disregarded.” The questions that are reverse coded are done in order to maintain the high and low averages for the compiled score. That is, if a student chose two on a negatively worded question (i.e., that they strongly disagreed with the statement) then it would be coded as a seven (i.e., indicating a higher sense of belonging). The questions were created not only to measure one’s sense of belonging to a field, but to also understand why female students may stop studying mathematics or leave their desired STEM major (Good et al., 2012).

Table 5.

Math Sense of Belonging Scale Components and Questions.

Component	Question
Membership	I feel that I belong to the math community. I consider myself a member of the math community. I feel like I am part of the math community. I feel a connection with the math community.
Affect (positive)	I feel like I fit in. I feel at ease. I feel comfortable. I feel content.
Affect (negative)	I feel anxious. I feel inadequate. I feel tense. I feel nervous.
Acceptance (positive)	I feel valued. I feel accepted. I feel respected. I feel appreciated. I feel calm.
Acceptance (negative)	I feel like an outsider. I feel disregarded. I feel neglected. I feel excluded. I feel insignificant.
Trust	I trust the testing materials to be unbiased. I have trust that I do not have to constantly prove myself. I trust my instructors to be committed to helping me learn. I trust my instructors to have faith in my potential.
Desire to Fade (positive)	I enjoy being an active participant.
Desire to Fade (negative)	I wish I could fade into the background and not be noticed. I try to say as little as possible. I wish I were invisible.

I used the survey as both the quantitative pre- and post-*MSoBS* measure for the students in the study, and as codes for the students' qualitative post-test reflection data described later in this chapter. Due to the 8-point Liker scale range, and for the purpose of this study, I classified the scores as follows: (1) a student who responded with a 1, 2, or 3 on majority of the questions were classified as having a low sense of belonging score; (2) a student who responded with 4 or 5 on majority of the questions were classified as having a medium sense of belonging score; and (3) a student who responded with a 6, 7, or 8 on majority of the questions were classified as having a high sense of belonging score. The class had a mean pre-*MSoBS* score of 5.89 and ranged from 3.67 to 7.70. Because I wanted to investigate the changes in *MSoBS* for students with different ranges of belonging, I chose a range of students for the study based on their composite pre-*MSoBS* score. Specifically, I wanted both male and female students who either had a high, medium, or low sense of belonging scores. The students chosen for the study had a slightly lower mean pre-*MSoBS* score of 5.42 and ranged from 3.67 to 7.07. More details regarding data analysis and use for qualitative data analysis are provided in the Analysis of Sense of Belonging Data section of this chapter.

Working Group Survey

The *Working Group Survey* (*WGS*) was developed during my pilot study to assist in making permanent working groups and consists of 12 questions using a 6-point Likert-scale and five open-response questions, for a total of 17 survey items (see Appendix B: Working Group Survey). Five of the Likert-scale questions were adapted from the Attitude Survey created by Brookstein et al. (2011). The Likert-scale questions ask students about their preferences and feelings about working in groups, perceptions toward mathematics, their mathematical skills, and their confidence with mathematics. The five open-ended questions ask the student to elaborate on their understanding of and role during group work activities, who they learn best with while engaged in group work, why they believe they work well with said person, and why they prefer to work alone or why they prefer to work in group[7]. All students in the course completed the working group survey as I used the survey results to create the permanent groups for all of the students in the course. Details about how the results of this survey were used to create groups are described in the Formation of Groups section in this chapter.

Other Data Sources

The remaining data sources include: test scores for both group and individual tests, transcripts from three groups for both Group Test 1 and Group Test 2, and post-test reflections from the 15 women in the course from both the first and second test, and interviews from the eight women recording during the group tests. In the following subsections I give a detailed description of each data source, and the analysis of each is provided further in this chapter.

Test Data

The group tests scores were collected for both group tests from eight groups (i.e., 30 students). The individual test scores were collected for 30 students for both the first and second test. The group and individual test data were only collected for students who were: (1) in groups that all participants signed consent for the study and (2) students in those groups completed all tests in the course.

---

[7] The open-ended question: “If you would prefer to work in a group, please describe the strategies you use, how to learn mathematics, and how working alone would prevent you from using these strategies. Additionally, describe how completing the work in a group will benefit you in your future studies and career?” was added to balance the survey and confirm answers with the similar question that asks students if they like to work on their own instead of in groups.

Transcripts

The transcripts for the three groups during both group tests were created using video, audio, and Livescribe data captures. The Livescribe pen is a technology enabled ballpoint pen with an embedded computer and audio recorder. The pen must be used with Livescribe digital paper to record what is written and said simultaneously. When the data from the Livescribe pen is uploaded to a computer, it synchronizes the written notes with the recorded audio. Audio of the group was captured primarily through the Livescribe pen. Since the presence of a video camera can influence student interaction, the cameras were placed too far away to capture the group's written work on the test. The written work was captured through the use of the Livescribe, as this device allows for real time capture of student work along with their audio (Blikstad-Balas, 2017; Linenberge & Bretz, 2012). This technological tool allowed me to see exactly who was speaking while answers were written on the test. Two video cameras were placed on different sides of each research group to capture student gestures, facial expressions, and spatial discrepancies that could not be captured from the audio of the Livescribe pen. An audio recorder was placed on the desk for each research group and used as the secondary method for recording what the students said during the test.

Post-test Reflections

At the end of each test day (see Table 4) students were given prompts to reflect on their experience during the group test. Through reflection, students can become aware of their feelings and positions in relation to their group as a community (Farabaugh, 2007). The post-test reflection questions, displayed in Table 6, required the students to reflect on both their actions and feelings during the test. Students answered these questions on a piece of paper that was attached to their individual exam.

Table 6.

Post-test Reflection Questions.

Do you feel you communicated your mathematical ideas during the group test? Please explain:
How do you feel your contributions were received in the group during the test?
Please provide one positive comment about your experience in the group test.
Please provide one negative comment about your experience in the group test.
Please share any comments you may have about your experience with the group test.

Individual Interviews

A total of eight female students were interviewed at the end of the course. These eight female students were participants in the three research groups that were recorded during the group tests. A full description of the groups and the participants are provided in the next section. All interviews occurred after the second group test, in order to not influence the students' behavior during the testing environment. During the interview students were given their responses to their post-test reflections for the two tests and asked to expand on and explain their responses. Because the interviews were further away from the event in question (i.e., the group test), the students did not recall exactly how they felt during the test. Therefore, the main purpose of the interviews was to gain clarification on their responses to the post-test reflections

Groups and Participants

I used the responses from the *WGS* and *MSoBS* survey to place the students in groups. In this section I will describe how I used their responses to form the groups, and detail the three research groups and participants of the study (i.e., members of the three research groups).

Formation of Groups

Considering learning as a process of participation and obtaining membership in a certain community (Sfard, 1998; Lave & Wenger, 1991), I arranged students into groups in which they could feel accepted and valued as a member of that community to best enhance their learning experience. If we consider the group as the community in which the student is a member, then the importance of group composition arises, as the particular students who make up the group (i.e., group composition) matters regarding student-centered learning activities in mathematics (Wiedmann et al., 2012). Therefore, I used the students' responses from the *WGS* when placing students in their permanent working groups. Additionally, I attempted to maximize the pre-*MSoBS* mean score for each group when placing students. This was done in an attempt to ensure equity among the groups for the sense of belonging need for learning. In addition to student responses surveys, I assigned students to groups according to the following criteria:

1. No group will consist of men and just one woman, because groups perform better on problem solving tasks when more women are present in the group (Woolley et al., 2010).
2. Women will also be paired with at least one other women in a group to reduce the possibility of stereotype threat (Spencer et al., 1999).
3. No group will consist of more than one person who selected strongly agree on the question "I prefer working alone rather than in groups when doing math" or strongly disagree on the question "I feel

comfortable speaking in front of my peers" on the WGS to ensure these students are included in group discussions.

To create the groups, I first looked at students' responses from Question 9 "I prefer working alone rather than in groups when doing math" and Question 10 "I feel comfortable speaking in front of my peers." Students who strongly agreed to Question 9 and disagreed with Question 10 were placed in separate groups. Next, I looked at the responses for Question 8 "I enjoy hearing the thoughts and ideas of my fellow students in class." I intentionally paired with students who agreed with this statement and placed them with the isolated students from the previous question responses. This placement was to help encourage students who do not like working in nor speaking in groups to voice their ideas more. Ideally, students who strongly agreed with Question 8 helped encourage their peers to talk during the group setting.

After the initial separation of students with these three questions, the open response answers to Question 13 and 14 were examined. Anecdotally, when students hold a more egocentric view for this question, they work in a more cooperative way rather than collaborative way. For example, students would work on their own and check each other's answers, if there was a disagreement in answers then students would redo the problem by themselves or check over each other's work. Therefore, to encourage group communication students who answered Question 13 from the *WGS* (i.e., Suppose you got one answer to a problem and your partner(s) got a different answer, how do you determine who has the correct answer) in opposite ways were placed in a group together. Opposite responses would be if one student spoke in a more egocentric way such as "I would look over each member's work" versus group-centric speak such as "We try to explain how we got that answer […] we each try to replicate the other's work" (see Table 7 for one example group placement with student responses).

Table 7.

Sample Group Responses to Question 13 "Suppose you got one answer to a problem and your partner(s) got a different answer, how do you determine who has the correct answer."

Group	Egocentric	Group-centric
Group 2: with two males and two females	The way I would determine who has the correct answer is to go over the problem step by step-by-step to see where any of the common mistakes were made, and also explain why that mistake was made.	We talk out how we got the answer and determine if any mistakes were made in either of our calculations. If no mistakes were made, I will do the problem again and have my group partner check my work. They will do the same and I will check their work.
	To determine the correct answer, I would look at both individuals' answers and compare work to see of a discrepancy occurs. If no discrepancy is apparent, I would then work the problem from scratch again.	We all explain how we got the answer we came up with and then re-did the problem together. If we are still indecisive, we call an SPT or the instructor over for help.

Finally, as student input is also important in helping the students feel comfortable in their groups, I considered their responses to Question 15 "Have you been working with someone already that you think you learn well with and would like to keep working with? If so, please list their name(s) and explain why you work well together." If the students provided a compelling reason why they learned well with a classmate for this question, these students were placed together in a group. Example responses that would be considered or not considered for this question are displayed in Table 8.

Table 8.

Exemplary Student Responses for Consideration in Group Placement on the Working Group Survey.

Question 15: Have you been working with someone already that you think you learn well with and would like to keep working with? If so, please list their name(s) and explain why you work well together.
Considered Responses
Person X and Y, because they both take what I say and actually make it into what I meant to say. Plus they are both amazing characters and don't fault me even when I'm wrong.
I work well with my current group. We all listen to each other and I don't feel like I'm drowned out by other members. We're all friendly and I don't feel anxious at all to state my opinion or ask for help.
Non-considered Responses
Person X because we have had some of the same math classes in the past.
Person X because they explain well and Person Y because they know how to solve problems.

Students were arranged into their groups on the third day of the course and took the *WGS* again on the fifth day of class. Because I was working with people and not just numbers, the groups were observed during the third and fourth day of class and changes were made to the groups if needed. There were two students who were switched due to personal issues observed in one group. The two students who were switched had similar responses to Questions 8, 9, and 10 on the *WGS* and were both male. A total of 10 groups were formed for the class. See Table 9 for the number of male and female students in each group along with their URM status.

Table 9.

Permanent Student Groups Descriptions and Demographics.

Group Number	Group Type	Male Students	Female Students
1	4 students information removed from study due to non-consent		
2[ab]	All Female		3 non-URM 1 URM
3	2F-2M	1 non-URM 1 URM	1 non-URM 1 URM
4	2F-2M	1 non-URM 1 URM	2 non-URM
5[b]	2F-2M	2 non-URM	1 non-URM 1 URM
7	3F-1M	1 non-URM	1 non-URM 2 URM
6	4 students information removed from study due to non-consent		
8	All Male	2 non-URM 1 URM	
9[b]	3F-1M	1 URM	3 URM
10	All Male	3 non-URM 1 URM	

[a] This group lost one non-URM female student before the second group test. All other groups remained the same.

[b] Research groups described further in detail in this chapter.

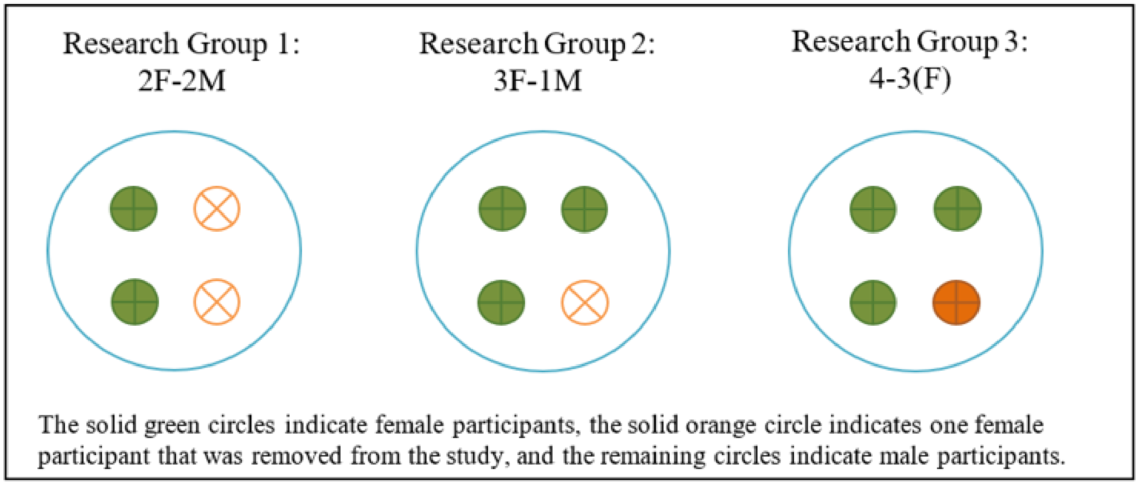
Participants

Participants for the study were chosen after the groups were formed. As I wanted to explore the different group setting for female students I had only one choice for the all-female group, but three choices for the 2F-2M group, and two choices for the 3F-1M group. Please note, there were not enough women enrolled in the course to make two all-female groups and multiples of the two other types of groups. In the following subsections I provide a brief introduction to each group and each participant in the group including their major, why they chose that major, and their past experiences with mathematics. Note, all names in this dissertation are pseudonyms.

The three research groups contained eleven individual students as depicted in Figure 7. Each research group consisted of four students who were purposefully placed into one of the three groups at the beginning of the course. The groups are labeled with the number of female students to male students in each group (i.e., 2F-2M). The first research group, 2F-2M, consisted of two female students and two male students. The second research group, 3F-1M, consisted of three female students and one male student. The third research group, 4(3)F, consisted of four female students in which one was removed from the study before the second group test. Detailed information about the students in the research groups are described in the Participants section in this chapter.

Figure 7.

Depiction of Research Groups.



The solid green circles indicate female participants, the solid orange circle indicates one female participant that was removed from the study, and the remaining circles indicate male participants.

2F-2M Research Group

This group contained two non-URM males (Isaac and Carl), one non-URM female (Sophie), and one URM female (Annie). All students were freshmen at the time of the course and required Calculus I for their intended degree. Each member in the group knew of at least one of the other members from a previous class. In their past precalculus classes, Carl had worked with both Annie and Isaac. During the classroom setting, the group would often discuss the mathematics very passionately and all four members spoke very highly of each of their peers in their group. The members had become friends and joked that in their group that Isaac was the dad and Sophie was the mom of the group.

Isaac. Isaac is a non-URM male who was a first-year student and placed into the 2F-2M research group. He had a pre-*MSoBS* score of 7.07 (i.e., a high sense of belonging). When entering college, he was originally a computer science major. He decided to switch his major to mathematics because he “spends too much time using technology" in his everyday life, so he chose mathematics as his new major, preferably minoring in statistics. He said he was good at math up until his eighth-grade algebra class. However, during this time his family had moved and he said changing schools could have been the reason he had a hard time in algebra. In high school he was not a top performer in his mathematics classes, but he said that was due to him not doing his homework. He stated he was good but lazy. Isaac explained that since he felt he was good at doing mathematics, that when he first learned a topic, he would not worry about practicing problems on homework. He had previous experience with group examination methods in high school. These groups switched three times each quarter and he said he drifted between good performing groups and the lower performing groups because he was "seen as an average student." Isaac felt that the group tests in high school were easier than the individual tests, and said that he could have completed them by himself and was not sure he needed the collaboration it required. However, he did end up taking many of the tests himself in high school due to his poor attendance record.

Carl. Carl is a non-URM male who was a first-year student and placed into the 2F-2M research group. He had a pre-*MSoBS* score of 3.90 (i.e., a low sense of belonging). He said he was leaning toward bio-chemistry as a major since he liked mathematics and science in school. But then said he did not want to "do science" after taking a chemistry class. He likes mathematics more than science, so he chose electrical engineering as his major. Carl's choice in major was also influenced by his father being an engineer. During middle and high school, he worked well by himself, it was not until college he started working in groups in mathematics classes. Carl claimed he had a "shaky start" when he got to college. In the past he liked being by himself and felt he learned well individually, but as he got older, he started to understand more about group dynamics. As he became more comfortable with his group, he felt he was learning more. He felt he could gather more knowledge by talking with his peers than he could by himself. His first experience with group work in a mathematics class was with the precalculus courses at this university, and his first experience with group examination methods was with this study. Carl was shy in the beginning of the course and did not speak that much during class or in the interview. He enjoyed the group work in the course and said that he hopes this will help him to become more outspoken in his future classes.

Sophie. Sophie is a non-URM female who was a first-year student and placed into the 2F-2M research group. She had a pre-*MSoBS* score of 7.07 (i.e., a high sense of belonging). She is a biology and chemistry major on the pre-medicine track. She is contemplating on dropping the chemistry to a minor in order to lower her stress levels during her undergraduate education. Her plan is to eventually become a pediatric cardiologist because she loves children and has a heart condition herself. Sophie’s choice of major is very personal to her, she wants to help others, and heart conditions run in her family. She had a choice between taking calculus or statistics for her major, and chose calculus due to its prestige. She does not feel that statistics is as “impressive” as taking calculus because of its difficulty. Sophie did well in algebra, but not well in her high school precalculus courses. She said she had a strong understanding of algebra concepts, but when it came to precalculus her high school teacher went over the content very fast since it was an honors class. She would stay after class to ask questions and said she felt her teacher was mad at her when she failed the class. Sophie enjoyed group work in both high school and college precalculus because she could talk to her peers. If she did not understand something, she would feel comfortable and often “get it” after a peer explained a problem. She also feels she learns well when she practices concepts. Sophie said that having someone explain it is not enough for her and she needs to practice more problems to see different ways she can solve something. Sophie particularly likes solving trigonometric proofs because after practicing them over and over she could find the “best” way to solve them.

Annie. Annie is a URM female who was a first-year student and placed into the 2F-2M research group. She had a pre-*MSoBS* score of 6.10 (i.e., a high sense of belonging). Annie is a civil engineering major planning to also obtain a mathematics minor. She chose her major because she likes structure and architecture, but does not want to become and architect. She felt civil engineering would be a good major for her current interests. Prior to this class, Annie felt mathematics to be easy. She said she just “understood everything” and would get close to perfect scores on her tests and quizzes. Annie did do practice work in her classes and completed her homework, but she never had to do extra work outside of class to understand the mathematics. She said that all changed when it came to calculus. This was the first time she had to put in more work to study to understand what we were doing in class. This class was a big shift for her and how she studied mathematics. Annie said she had to go back and practice algebra before the calculus concepts since every topic was building on each other. She spent each night studying during this course, and said she was not used to studying mathematics so much. Annie thought it might have been the pace of the class itself and that was why this class was the first time she was having difficulty learning mathematics.

3F-1M Research Group

This group was composed of one URM male (Kurt) and three URM females (Kathy, Mary, and Dorothy) in which two of the females already obtained their bachelor's degree (Kathy and Dorothy) and were talking Calculus as a prerequisite for graduate school. The group worked well together during the classroom setting; however, Kurt would often ask for the help from an SPT before discussing with his groupmates. Kathy noted this in her interview that Kurt often "hogged the SPTs" in class. None of the members in the group had previously worked with each other in past classes.

Mary. Mary is a URM female who was a first-year student and placed into the 3F-1M research group. She had a pre-*MSoBS* score of 5.07 (i.e., a medium sense of belonging). She is a chemical engineering major, a choice that was largely influenced by her father. She stated that her father is an engineer and feels that engineering is a "respectable" career. Mary first thought she was not smart enough to be an engineer until she enrolled in college. When she started meeting other engineering majors in her classes, she realized she was as intelligent as the other students. When Mary was asked to explain her past experience in mathematics classes she simply replied "tears." Both her parents were college educated in mathematically intensive fields (here mother had a masters of finance) and would lose patience with her when she was in elementary and middle school. Mary said even in the classroom she always felt like she need "two more minutes" to understand the mathematics. She would often look for shortcuts to doing her work and just thought she was not good at mathematics. Even though she lacked the confidence in her skills, she was enrolled in calculus in high school. When she talked about her teacher in her high school calculus course, she noted how he was nice to the boys in class but did not appear the same to the girls. She went to ask him for help because she was afraid of disappointing him and failing his course and he told her "you're not going to disappoint me a lot of people are bad at math."

Kathy. Kathy is a URM female who was a post-baccalaureate student and placed into the 3F-1M research group. She had a pre-*MSoBS* score of 3.83 (i.e., a low sense of belonging). She was older than most of the other students in class and already had a degree in biology. Her undergraduate degree did not require calculus and she enrolled in the course because she originally wanted to go to pharmacy school. However, Kathy was very interested in medical research and switched her graduate interests to medical school. She is preparing to take the Medical College Admission Test (MCAT) this summer. Kathy seemed to have a roller coaster relationship with mathematics. She never really liked mathematics in school and said she always had a "distain" for the subject. Although her mother was really good at mathematics, she just found herself frustrated when learning. Until fifth grade, when her teacher helped her relax and gain confidence with mathematics. This changed her view on learning mathematics until high school. She took precalculus and trigonometry in high school, but as a Black woman in the south she felt left out in the class. She said her teacher "was very particular" to the students she helped. Kathy explained that this teacher treated the white students in class differently than the black students. Particularly, she felt her teacher gave more attention to the white athletes in their mathematics class. She recalls working in a group with one black male and two white females in this class, and when the teacher came over to help, she would only talk to the two white girls. The occasion of racial disparities in the classroom effected Kathy greatly. She noted, even in this class which was created to help URM students, she was a minority and then named the few black students there were in the class.

Dorothy. Dorothy is a URM female who was a graduate student and placed into the 3F-1M research group. She had a pre-*MSoBS* score of 3.67 (i.e., a low sense of belonging). She was a few years older than many of the other students in class and already had a degree in biology and minored in chemistry. Although she took many upper-level chemistry classes, and did research in chemistry, her undergraduate degree did not require calculus. She enrolled in the course because she was accepted into pharmacy school and calculus was a pre-requisite. Dorothy was a McNair scholar and had already been introduced to doing research. She talked about the pharmacist her grandmother has and how helpful he was to both of them. Although she did not want to enter the retail side of the pharmacy, she enjoys the research and chose pharmacy school because she wants to help others with their medication. It had been many years since she was in a mathematics classroom and could not recall many experiences. She summed her past experiences by saying she knows she is “bad at math” and has to practice the homework and study to pass tests. She could not recall anything about her feelings in past mathematics classes.

Kurt. Kurt is a URM male who was a first-year student and placed into the 3F-1M research group. He had a pre-*MSoBS* score of 4.90 (i.e., a medium sense of belonging). He has not declared a major yet but is debating between either mathematics or finance. He enjoys working with numbers and talking with people about mathematics, but he has not thought about his future plans. He describes his past experiences in mathematics class as typical. The teacher would lecture then the students would work on example problems. In his high school, the classes would take group tests and quizzes. Kurt said he would sit with his friends and they would divide up the problems to complete the test or quiz. He explains that with his experience with group work, he and a partner work on problems on their own and then come together to confirm answers. He believed mathematics was learned by first seeing problems from a teacher, practicing similar problems with others in class, then on homework, then being tested on those problems. Kurt said this course was nothing like his past experiences and made him anxious (although he also said he has anxiety issues). He was used to cooperative group work instead of collaborative group work. Additionally, he stated how he was not used to exploring the concepts in mathematics without the teacher first doing problems or telling the students how to do the mathematics. Kurt came from an all-male college preparatory high school and was currently in a fraternity on campus. He understood what the study was about and agreed to participate, but he said he was afraid to come across as sexist.

4(3)F Research Group

This group was composed of three non-URM females (Maryam, Pratibha, and Rin) in which one was an ESL student (Pratibha) and one URM female (Georgia). The non-URM student Rin was only present for the first group test and was removed from the second group test. None of the women had previously worked with each other in past classes. Maryam and Georgia enjoyed working together but both felt that Pratibha thought “faster” than them.

Georgia. Georgia is a URM female who was a first-year student and placed into the 4(3)F research group. She had a pre-*MSoBS* score of 7.07 (i.e., a high sense of belonging). She is a computer engineering major and chose this due to the prevalence of technology in the world. She said you just cannot escape technology and she wanted to do something that was "already out there." She did not find her past mathematics classes to be difficult, and said mathematics used to be her favorite subject. Mathematics was more interesting to her than any other subject and she enjoyed the rush she would get when she would find the correct answer to a problem. Georgia was enrolled in honors mathematics in high school which had smaller class sizes than her other classes. She could not recall details about her mathematics classes in high school, she remembered doing little homework and just said the work was not that hard. She enjoyed learning mathematics until she took Precalculus II in college (this course focuses on trigonometry). This was the first mathematics class she ever failed. She understood that mathematics got harder as she took higher level courses, but said she did not expect it to be as hard as it was for her to learn.

Maryam. Maryam is a non-URM female who was a first-year student and placed into the 4(3)F research group. She had a pre-*MSoBS* score of 4.73 (i.e., a medium sense of belonging). She plans on become a mathematics teacher in the future and is majoring in middle childhood education with a focus in mathematics and science. Maryam always wanted to be a teacher but started college as a nursing major because of her mother's influence. Her mother said she could "do better" than teaching, but Maryam changed her mind her second semester of college and decided to follow her passion. She was advanced in mathematics at a young age and started pre-algebra in seventh grade and algebra in eighth grade. By her junior year in high school, she was able to enroll at the local community college and took college algebra. This calculus course was her first mathematics class at a four-year university. Her experience in mathematics class from eight grade through high school all contained group work and group testing. Her school district used the College Preparatory Mathematics (CPM) texts books and many of the teachers followed the book. Maryam did not like when the teacher did not “actually teach” the mathematics and said some teachers in her past experience would get upset if you asked them a question instead of asking your group. She was never able to pick her teammates as each teacher in her experience assigned students to their groups. She claimed no one in her high school liked the teamed tests. Maryam also felt the team tests were harder than the few individual tests they took in high school.

Pratibha. Pratibha is a non-URM female who was a third-year student and placed into the 4(3)F research group. She had a pre-*MSoBS* score of 6.23 (i.e., a high sense of belonging). She was older than many of the other students in class, an ESL (English as a second language) student, and already had an associate degree in networking technology. Pratibha was majoring in computer science because she wanted to be more involved in the programming side of computers. She said that the programming side of computers is a “better payoff” than the hardware side in which her current degree was situated. She said for all of her mathematics classes in the four-year university and previous college she always needed a tutor to help her with class. When she talked about her experience she focused only on the grades she received in class. To Pratibha getting a good grade meant she was good at mathematics. She enjoys the computing part of mathematics but has trouble with concepts. Pratibha explained that she can easily calculate a limit or derivative but to find a limit from a graph was hard.

After the students were arranged into their permanent working groups, they worked together for half of the class time in those groups. At the end of the third and sixth week of class, the students were tested as a group and individually. In the next section I provide a description of the data collected for the participants described above during their group tests (i.e., group test data).

Analysis of Group Test Data

Group Test Student Interaction Quantitative Analysis

The individual and group test scores were analyzed to answer the following research question:

1. In a reform Calculus I course, how do students' performance on group tests compare with their performance on individual tests; and how do gender, and the intersection of race and gender contribute to the differences, if at all?

To examine student performance on the tests, I first considered the test scores for overall student performance in the class (*N* =30), next I compared test scores by gender (Males [*N* =15] and Females [*N* =15]), third I compared test scores across student gender and race demographics (non-URM Males [*N* =10], URM Males [*N* =5], non-URM Females [*N* =7], and URM Females [*N* =8]). Due to the relatively small sample sizes and observational nature of the study, all comparisons were done using descriptive statistics techniques. The results of this analysis are presented in Chapter 4.

Group Test Student Interaction Qualitative Analysis

Because group interactions during a group test is a central topic for this study, the group test data was analyzed to answer the following research question:

3. How are social interactions within the group composition different based on the composition of the group, if at all, and can any differences be used to explain differences in the female students' sense of belonging?

To answer this question, the group test data was fully transcribed and I created narratives of the groups' interactions during the test for each student utterance. The narratives include the student talk turns (i.e., each utterance for every student) for the entirety of the test. That is, when the talk (or utterance) went from one student to another, that is considered a "talk turn". For example, consider the following:

Kurt: Alright, so who's doing number one?

Kathy: Number one, it's the one we did yesterday.

Mary: I don't feel comfortable with this one.

Kurt: Okay.

Mary: Like, like I could riddle it all our right now, anybody?

This exchange consists of four talk turns. The first one is counted with Kurt's initial question and then Kathy's response. The moment the speaker went from Kurt to Kathy the talk turn was completed. The next talk turn is when Mary spoke (turning the talk from Kathy to Mary). Some talk turns were also completed from one speaker to themselves (i.e., Mary to Mary). This would be self-responsive talk turns and determined by reviewing the audio files. That is, when the audio files were reviewed, if students paused during the middle of one talk turn and started speaking again but not but in a different tense this would be considered two talk turns. For example, if a student started their talk turn by asking a question, but then switched their tone and answered their own question; the question was considered a talk turn and the statement (i.e., answer) was considered another talk turn. After the narrative was created, it was uploaded to an excel sheet and each line of talk was coded separately by myself and a fellow researcher. After the initial round of coding by the two researchers, we met to discuss any discrepancies in the coded talk turns until we reached 100% agreement.

To understand and explain how the groups interacted, students' individual talk turns were coded line by line using the Group Problem Solving Framework defined by Chiu (2000). When solving the problems during the test each talk turn (i.e., each time a student spoke) was coded along the three dimensions of the Group Problem Solving Framework: Evaluation of Previous Action [EPA], Knowledge Content [KC], and Invitational Form [IF] (Chiu, 2000). As social interaction is a key learning need for participatory learning (i.e., See **Error! Reference source not found.** in Chapter 2), the Group Problem Solving Framework helped to understand how the students interacted, and allowed for classification of their interactions during the group test. That is, this framework provided me with a way to see how the students talked about the mathematics and which type of talk was equitable or inequitable for all group members.

To examine the students' social interactions, each problem (i.e., problem one, two, three, four, and the bonus) was coded separately by two researchers. The reading of the directions was not coded as they did not contain any talk around mathematics. Two researchers coded the problems for each group separately then met to discuss and reconcile any codes that were not in agreement. During this process the code book (Appendix E: Group Problem Solving Code Book) was refined and the final two tests (i.e., Group Test 2 for 2F-2M group and 4(3)F group) were coded by one researcher and agreed upon or discussed by the second researcher. In this section I provide a description of the coding framework by Chiu (2000), how it was refined, and examples for how the tests were coded.

Group Problem Solving Framework

When analyzing talk turns during group tests, I utilized a framework that includes three dimensions for how students verbally respond to one another: Evaluation of Previous Action (EPA), Knowledge of Content (KC), and Invitation Form (IF). The first dimension, EPA, describes how a student evaluated the previous speaker and is coded as either supportive, critical, or unresponsive to the person who spoke prior to them. The second dimension, KC, describes the mathematical content offered in the talk turn as either a new contribution, a repetition of the previous mathematical idea, or not containing any mathematics. The third dimension, IF, was coded for each talk turn as either a statement, command, or question. Every talk turn was coded with each dimension and thus a single talk turn contains three codes (i.e., one for EPA, one for KC, and one for IF). It will benefit the reader to understand how each dimension is defined and coded before seeing how the code was applied. These descriptions are in the following three paragraphs, and example responses and codes from the first test are provided in Table 10, after the for descriptions of each dimension and their codes.

The first dimension, EPA, was coded for the type of talk of the student. A student’s response to their peer was either supportive (coded as +), critical (coded as -), unresponsive (coded as 0), or unresponsive-neutral (coded as 0*). Supportive responses included speech that reinforced the direction of a group’s current approach to solving the problem (Chiu, 2000). Responses that were considered supportive included encouragement such as “yeah”, “okay”, or “mm-hmm”, or responses that encouraged and added to the current mathematical ideas. Critical responses included speech that altered the current problem-solving approach (Chiu, 2000). Responses that were considered critical could be a simple “no”, “I don’t think so” or alterations to the current approach such as “yes, but” with an intention of changing the way the group was answering the question. Unresponsive responses included speech that did not acknowledge or evaluate the previous speaker (Chiu, 2000). Responses of this nature could include off-task speech such as “what time is it”, logistical speech such as “did you sign your name”, or speech where the previous person was not listened to (see Table 10). Unresponsive neutral speech was added to our coded book for responses in which the previous person's talk was heard but not evaluated (Chiu, 2000). This code was suggested as a possibility by Chiu (2000) if such utterances could be distinguished. Responses that were coded as unresponsive neutral include the first time a person offered a mathematical idea since the person offering the idea heard the person reading the question. Other responses that were considered unresponsive neutral include speech such as “hmm?”, “let me think about that”, “I don’t know”, and “wait, what?”.

The second dimension, KC, was coded for the mathematics in each student utterance. Their response could either include a contribution (coded as C), a repetition (coded as R), or was null (coded as N) and contained no mathematics. Contributions were coded when a student offered a new mathematical idea or introduced new mathematical ideas into the conversation (Chiu, 2000). Responses coded as contributions included the first mathematical idea that was offered to solve a problem, fundamental mathematics that extended the first idea, or alterations such as a new idea as a critique of the first. Repetitions were coded when a student repeated the knowledge of the previous contribution (Chiu, 2000), finished the idea or sentence of the previous speaker, or kept the current problem-solving approach in support of the first speaker. The difference between responses coded as a contribution or repetition may seem subtle. However, a distinction was made when considering the rubric for the test and diverge slightly from the original framework. If the students' conversation was supportive and kept the group with the same problem-solving approach these interactions were coded as repetitions. If a student added an idea to the mathematical conversation that added points to their score (i.e., a fundamental part of the answer that if they had not included it, their score would have been lower) these utterances were coded as contributions. This distinction was made to determine who offered the original idea for each problem as I wanted to track the progression of one mathematical idea. If we had followed the coding scheme as defined by Chiu, there would have been more contribution (C) codes than repetition (R) codes in the final analysis. Because we were trying to track the progress and conversations around initial ideas for each group, having only one contribution allowed for us to easily identify and track each new idea. Finally, null statements were coded for any statement that did not include mathematical content (Chiu, 2000). Responses that were coded as null include "mm-hmm", "yeah", "so", "wait, why" and "no" for a few examples.

The third dimension, IF, was coded for the form of talk of each student. Their response was either a statement (coded as _.), question (coded as ?), or command (coded as !!) (Chiu, 2000). Questions were coded both for when a student made a why, what, how type of comment as well as when the inflection of their voice had notable concern. Talk turns were coded as a statement when a student made a declarative comment that did not question the previous speaker or command an action such as writing on the test.

Commands were coded when a directive was issued to the group or writer. Commanding statements were either given by the intellectual or social authority at the time. Recall Langer-Osuna (2016) defined intellectual authority as a student who is treated as a credible source of information, and a social authority as a student who is treated as having the right of giving directions. Intellectual authority commands would include comments that direct the writer of the problem to write down a specific answer. Social authority commands would include comments that would direct the group to move on to another problem. An example of an intellectual command is given in Table 10.

When combining the codes along the three dimensions, there are a total of 36 different combinations that can represent one student talk turn. For example, a supportive repetition statement would be the combination of the codes +R_., where a critical null command would be the combination of -N!!. By looking at the combining of the three dimensions (EPA, KC, and IF) I could better understand how the students talked about the mathematics and created their answers for each question. However, to understand how the groups created their answers, the coding scheme only applied to on-task talk as this talk centered around the mathematics and both off-task and logistical talk were coded with a “0” for the EPA level and an “N” for the KC level. The off-task and logistical conversations were removed from the overall analysis as for many groups the frequencies of off-task and logistical were overwhelming when examining how the students interacted around the mathematics. All of the information from the last four paragraphs is displayed in Table 10.

Table 10.

Example Coding of Talk Turns.

EPA	KC	IF	Speaker	Talk turn
				On-task Talk Turns
+	R	!!	Isaac:	So, for this one write false
+	R	_.	Carl:	I would definitely say that's false. That's, that's true.
0	R	!!	Isaac:	And say, um, um, I don't know, like uh, hm. Depending on the function, um, you could have inputs that are both, uh, the negative and the positive version of that integer. I guess. I don't know. That sounded really weird way to say it.
+	N	_.	Carl:	mm-hmm
+	R	!!	Isaac:	Um. Just say that, it depending on the function you can have a negative or positive integer equal to the same output.
				Off-task Talk Turns
0	N	_.	Maryam:	Good thing this class is pass or fail.
0	N	_.	Georgia:	mm-hmm
0	N	_.	Maryam:	We just need a C.
0	N	_.	Rin:	C's get degrees.
0	N	_.	Georgia:	I need a B.
0	N	?	Maryam:	For what?
0	N	_.	Georgia:	To pass, not to pass but, to get like a good GPA and stuff.
0	N	_.	Maryam:	Well this class doesn't factor GPA.
				Logistical Talk Turns
0	N	_.	Kurt:	Let's see this next one
0	N	_.	Mary:	ooo (noise)
0	N	_.	Kathy:	Oh that looks fine.
0	N	_.	Kurt:	This one's nice one this one should take like ten minutes to be honest (looking at problem 4) [\ alright \]
0	N	?	Kathy:	[\ wait are \] we finished with the other side?
0	N	_.	Mary:	Not yet
0	N	_.	Kurt:	Not yet but we can //
0	N	_.	Dorothy:	\\ Yeah
0	N	_.	Kurt:	Come back to that honestly

The table above shows three different conversations, where only the on-task talk was centered around the mathematics. To investigate all of the on-task talk turns, the codes were examined in bundles around a conversation (i.e., each coded student talk turn throughout the conversation). If students were building on each other’s mathematical ideas then the conversation would contain many +R_. type of statements. If students were agreeing with one another but not adding to the mathematical ideas, then the conversation would contain many +N_. type of statements. If one code could not satisfy the nature of a single talk turn, then this talk turn was split into multiple talk turns for each part of the student’s utterance that required a different code. In Table 11 on the following page I provide the reader with examples of coded statements for three different types of on-task conversations that occurred during the first test. Note in the transcript, the // code corresponds to when one student is talking and their talk is cut off by the symbol \\ when another student interrupts their speaking. Additionally, talk that is in brackets [\ … \] corresponds to when two or more students are talking at the same time. When students talk at the same time their overlapping talk is given the same codes for the three dimensions. This is why in the first example of the Repetition Pathway Kurt, Kathy, and Mary are all coded as + R _. as they finished Mary’s initial statement at the same time. Ideally, if students are discussing and building their mathematical answers as a group, this would be the most common code seen in the transcripts of the tests.

Table 11.

Coding of Student Interactions.

Conversation for Repetition of Mathematical Idea

EPA	KC	IF	Speaker	Talk turn
0*	C	_.	Mary:	And then the definition. Okay, a function is something that you can put input values and get
+	R	_.	Kurt:	[\ Output values \]
+	R	_.	Kathy:	[\ Output values \]
+	R	_.	Mary:	[\ Output values \]
+	C	_.	Dorothy:	Oh, a pattern, a pattern input value which produce output
				Conversation of a Null Agreement
+	C	_.	Georgia:	Well I do remember the three different ways numerical, algebraic, and [\ graph \] the tables.
+	R	_.	Rin:	[\ Graph \]
+	N	_.	Rin:	Yeah
+	N	!!	Georgia:	So we can write those down. Make sure like you do a complete sentence.
0	N	!!	Maryam:	First let's answer the first part then do the examples.
+	N	_.	Georgia:	Okay.
				Split Speech and Commanding Statements
0*	C	_.	Isaac:	the derivative is the uh limit of, or the limit as X or no as H approaches zero of the difference quotient.
0*	N	?	Isaac:	You know what I mean?
0*	R	!!	Isaac:	So wait, let me say that again. The derivative, oh wait, (whispers) *the derivative*. (*the group laughs*)
0	N	_.	Isaac:	I felt like I was too loud, everybody stopped talking. (whispers) *No more talking* Okay, never mind.
0*	R	_.	Isaac:	Um, a derivative is the limit of the difference quotient as H approaches zero. That's math, that's mathematical [\ definition \]
+	N	_.	Sophie:	[\ yeah, yeah, yeah \]
+	N	_.	Annie:	[\ yeah \]
+	N	_.	Carl:	[\ that's definitely how it is \] I was thinking about [\\ that in my head too \\]
0	R	!!	Sophie:	[\\ So wait it's the \\] limit //
+	R	_.	Isaac:	\\ So a derivative,
+	N	_.	Sophie:	Going to have to write it out.
+	R	_.	Isaac:	a derivative is um the limit of the difference quotient as H approaches 0. Um. Which is //
0	N	?	Sophie:	\\ Should we, should we write an example,
0	N	!!	Sophie:	I'm just going to do it anyway.

To gain an overall sense of how the three groups talked during the test, the coded talk turns were combined into seven main themes: (1) supportive talk with mathematical content, (2) supportive talk without mathematical content, (3) critiques with mathematical content, (4) critiques without mathematical content, (5) unresponsive statements, (6) commands, and (7) questions. See Table 12 below for the combination of codes that are present for each of the seven themes.

Table 12.

Combined Codes for Seven Categories of Student Talk.

Category of Student Talk	Codes from Group Problem Solving Framework
Supportive talk with mathematical content	+C_. +R_. 0*C_. 0*R_.
Supportive talk *without* mathematical content	+N_. 0*N_.
Critiques with mathematical content	-C_. -R_.
Critiques *without* mathematical content	-N_.
Unresponsive statements	0C_. 0R_. 0N_.
Commands	+C!! +R!! +N!! -C!! -R!! -N!! 0C!! 0R!! 0N!! 0*C!! 0*R!! 0*N!!
Questions	+C? +R? +N? -C? -R? -N? 0C? 0R? 0N? 0*C? 0*R? 0*N?

The results of the data analysis above are provided to give an overview of the three groups' interactions. To understand how each student in each group spoke in response to one another, however, more detailed analysis was completed. Details on how the student-to-student data analysis was completed is provided in the next section.

Sociograms of Student Interactions

When one person speaks during the test there are 576 possible responses that could happen each time. This accounts for the 36 different combinations of the three dimensions and the four members of the group because each student could also respond to themselves. When a student asked and answered their own question without giving the group a chance to respond, their talk turn was split into multiple talk turns and each line was coded for the three dimensions (i.e., responding to themselves). The talk was split if at least one of the dimensions had a different code. For example, if a student asked a question but then immediately answered their own question, the talk turn was split into two talk turns, with one coded as a question and the next as a statement. With this large number of possible combinations, I counted the most frequent occurrences for each group to determine how they constructed their answers for any talk that occurred at least five times (i.e., accounting for any combination that had a 1% chance of happening). For example, if person A responded to person B with a supportive repetitive statement at least five times this would be included in the overall analysis of the students' interactions during the group tests. However, if person A responded only four times to person B with a supportive repetitive statement during the test, these would not be included.

Using social network analysis, a sociogram was generated for each group along each dimension for all interactions that occurred a minimum of five times during the test. The sociograms were created to help the reader envision the interactions that occurred among the group members. Sociograms are graphs that plot the structure of interpersonal relations of a group. These graphs contain a node (i.e., circle) for each participant in the group and different colored lines that connect one node to another. The nodes are color coded by gender (i.e., male is teal, and female is orange) and each node has the first initial or first two letters of the participants name on the node to identify the person. The lines are generated by the frequency of talk turns between each person that occurred during the test. A line from one node to another node indicates a person’s response to the original speaker. For example, if speaker A was represented by a teal node and speaker B was represented by an orange node, then a line going from node A to node B indicates speaker A was responding to speaker B. The thickness of the lines refers to the relative frequency of the certain talk turn that occurred. That is, lines that are thicker happened more often than the thinner lines in the sociograms. For each test, I created three sociograms (i.e., one for each dimension) for each of the research groups, and chose colors for the lines to display the different items along the three dimensions. For example, for the Action Dimension, supportive comments are green, critical comments are red, neutral comments are gold, and unresponsive comments are blue. Therefore, a green line from one teal node to another teal node indicates the frequency of supportive talk between two male students. For the Content Dimension, contributions are colored red, repetitions are bright green, and null mathematical comments are purple. For the Form Dimensions, commands are colored red, statements are bright blue, and questions are olive green. The colors are provided on a legend for all sociograms in this chapter for each research group. An ideal set of sociograms that would display students collaborating to build their mathematical answers would have a majority of green lines in the first image, green lines in the second image and blue lines and yellow in the third image (i.e., yellow and blue make green).

I analyzed the interactions of the three research groups during both the first and second group test. The first group test contained conceptual questions about functions, limits, the formal definition of derivative and its applications. The only rule of engagement during the test was that each person had to be the writer for one of the questions and only the writer of the problem could hold the pen during that problem (i.e., encouraging students to discuss the mathematics and explain their ideas). The second group test contained conceptual questions about derivatives, derivative rules, implicit differentiation, and derivatives as a function. For the second test, I did encourage the students to look over and discuss each problem before assigning the writer for each problem. Again, each group member had to be the writer for one of the questions and only the writer of the problem could hold the pen during that problem.

By examining the connections in each sociogram for each group, I was able to describe how the students spoke to each other and built their answers for the group tests. The descriptions of their interactions for each group are then discussed in conjunction with the female students' post-test reflections who were members of the group in order to answer the third research question. The results of the group test interaction analysis for each research group are presented Chapter 4.

Analysis of Sense of Belonging Data

The sense of belonging data included the student responses to the *MSoBS* survey, post-test reflections, and individual interviews. The quantitative data consists of the *MSoBS* survey results, and the qualitative data is the post-test reflections. The individual interviews were used to gain more insight and clarification from eight of the females regarding their post-test reflection responses. Data sources were analyzed for the possible shifts in the female students' sense of belonging due to the social interactions during the group tests. In this section, I provide details of the data collected and analysis methods of the sense of belonging data.

Sense of Belonging Data Collection

At the end of the third week and sixth week of the class, after students engaged in the group and individual tests, each student was asked to reflect on their experiences by answering five open-ended questions in a post-test reflection. Additionally, at the end of the first day and during the last week of class each student took the *MSoBS* survey for a quantitative measure of their sense of belonging to mathematics. Details about the data collection can be found previously in this chapter under the Measures and Data Collection section. In the following subsections, I provide details about the analysis methods of the sense of belonging data.

Sense of Belonging Data Quantitative Analysis

The pre- and post-MSoBS survey results were used to determine if there was any quantitative change in the female students' sense of belonging. I examined the pre- and post-*MSoBS* survey scores separately for the URM and non-URM females for the: (1) overall *MSoBS* score, (2) Membership composite score, (3) Trust composite score, (4) Desire to Fade composite score, (5) Affect composite score, and (6) Acceptance composite score. Recall the sense of belonging scores were classified for this study as:

1. high *MSoBS* scores for students with majority of item responses at a 6, 7, or 8 on the Likert-scale,
2. medium *MSoBS* scores for students with majority of item responses at a 4 or 5 on the Likert-scale,

3. and low MSoBS scores for students with majority of item responses at a 1, 2, or 3 on the Likert-scale.

I chose these classifications as the survey uses an 8-point Likert scale does not contain a neutral choice, therefore scores under 4 leans towards the students disagreeing with statements (i.e., a low sense of belonging), and scores over 5.80 leans towards students strongly agreeing with statements (i.e., a high sense of belonging). Due to the relatively small sample sizes and observational nature of the study, all comparisons were done using descriptive statistics techniques. In Chapter 4, I provide the results of the analysis with descriptions of the following:

1. Scores that increased (i.e., pre score was in the low range and post score was in either the medium or high range, or pre score was in the medium range and post score was in the high range).
2. Scores that stayed in the same range (i.e., pre and post scores were both within the high range, or pre and post scores were both within the medium range, or pre and post scores were both within the low range)
3. Scores that decreased (i.e., pre score was in the high range and post score was in either the medium or low range, or pre score was in the medium range and post score was in the low range).

The results that presented in Chapter 4, will be described and compared across race for all female students in the study. Additionally, any changes in the quantitative scores are further discussed with the qualitative sense of belonging to mathematics data analysis results.

Sense of Belonging Data Qualitative Analysis

The second main purpose of the study is female students sense of belonging to mathematics, to qualitatively answer the question:

2. How does the sense of belonging to a mathematics community change for female students after engaging in group testing in a reform Calculus I course, if at all?

The student post-test reflections were analyzed for the female students. I first deidentified the post-test reflections and interview data and assigned the same pseudonym for the females in the research groups. The post-test reflections were then uploaded to an excel sheet and then first deductively coded for positive and negative occurrences of membership, acceptance, affect, trust, and desire to fade (see Table 5 for the description of the five sense of belonging components). See Table 13 for examples of comments from the reflections that were coded for the positive and negative sense of belonging components. The positive would indicate an increase whereas the negative would indicate a decrease in feeling as if one belonged. I performed two iterations of deductive coding for each of the sense of belonging components. During the second iteration, I simultaneously used open coding for interesting comments that fell outside of the sense of belonging component codes (Patton, 2015).

Table 13.

Sense of Belonging Post-test Reflection Codes.

Code	Definition of Code	Examples from Post-Test Reflections
Membership (+)	Comments regarding "team" or "team work" in which the student felt a part of their group community.	*"I truly enjoy my group. I think this is the best possible group I could be in, they help [me] understand material."*
Membership (-)	Comments in which the student did not feel as part of the group, or left out of the group discussion. Not feeling as a member of the group community.	*"My ideas and thoughts were not listened to and disregarded by all of my group members. Even if my answers were wrong, I would have liked it if someone could have told me that I was wrong and explained why."*
Affect (+)	Comments in which the student felt listened to and comfortable in their group to speak their ideas.	*"I felt that I was able to share anything I needed to say and got great feedback."*
Affect (-)	Comments in which the student feels stressed, inadequate, and not listened too by their group.	*"I don't think that I was listened to at all."*
Acceptance (+)	Comments in which the student felt as a valued and respected contributing member of their group.	*"I think my contributions were received well, when I added to the conversation, I felt like my comment was being considered."*
Acceptance (-)	Comments in which the student feels disregarded or excluded by the other group members.	*"[…] even when I did speak or try to contribute my ideas to the group, I felt like they just assumed I was wrong and disregarded what I said."*
Trust (+)	Positive comments about the testing materials or instructor's commitment to the students' learning.	*"I think it was a nice way to test."* *"I could express more knowledge of things I learned on the group test rather than the individual one."*
Trust (-)	Negative comments about the testing materials or instructor's commitment to the students' learning.	*"I thought it was much more difficult than everything we did so far (classwork, homework, practice test/game)"*
Desire to Fade (+)	Comments in which the student describes their active participation in the group.	*"We bounced off more ideas and felt more open about voicing opinions"*
Desire to Fade (-)	Comments in which the student describes being less of an active participant in the group.	*"I was more quiet during this group test and I didn't want to speak in fear of confusing people or holding the group back."*

Note that multiple codes could apply to a single comment, for example consider the response from Annie from the first question on post-test reflection one:

> *My ideas and thoughts were not listened to and disregarded by all of my group members. Even if my answers were wrong, I would have liked it if someone could have told me that I was wrong and explained why.*

This comment was also coded as Affect (-) for the feeling of not being listened to, and coded as Acceptance (-) for feeling disregarded by her group members (i.e., wishing someone would have explained why she was wrong).

After the first round of coding, I open-coded the comments that were tagged as "interesting" and fell outside of the sense of belonging codes inductively. These tagged comments were categorized to describe emergent themes in the data that were not represented in the MSoBS subcomponents (Patton, 2015). I identified five new codes from the open coding and did ten rounds of coding in order to ensure each new code was applied to the data when appropriate. These new codes were: increased confidence, decreased confidence, testing environment, time management, and frustration. After the last round of coding, I consulted another researcher to discuss and confirm each assigned code. Additionally, I condensed all sense of belonging with a positive affect into MSoBS(+), and all sense of belonging with a negative affect into MSoBS(-), and compiled the remain codes into the following seven:

1. MSoBS(+) positive affect (includes membership(+), affect(+), acceptance(+), trust(+), and desire to fade(+))
2. MSoBS(-) negative affect (includes membership(-), affect(-), acceptance(-), trust(-), and desire to fade(-))
3. Working together (includes collaboration, discussion, feedback)
4. Increase in self-confidence
5. Decrease in self-confidence

6. Specific comments about the test or testing environment

7. Time management

8. Frustration including general testing complaints, problems with group members, not being able to hold the pen, the group disagreeing.

Finally, I identified themes among the codes from both the deductive and inductive coding processes. The description of how the themes were derived from the data is provided in Chapter 4 along with the distribution of the codes above for the URM female and non-URM female students. Additionally, the frequency of themes for the URM female and non-URM female students, and connections between this qualitative and quantitative *MSoBS* results are discussed in Chapter 4.

Trustworthiness, Limitations, and Delimitations

As with any qualitative study there are limitations, delimitations, and trustworthiness issues. As the analysis of the data relied on my interpretation of it, in order to establish trustworthiness, I was explicit on my theoretical perspective and its influence on this study. Additionally, I worked with another researcher to code the transcripts of the group tests and to account for my possible bias. To support reliability, both I and the other research separately coded the same excerpt from the data with the defined framework. We met once a week and discussed our coding results. We continued this process throughout the entirety of the coding process and discussed all discrepancies until we came to a consensus.

In continuing to establish trustworthiness as a researcher and for the purposes of this study, I will elaborate on the limitations and delimitations of the study. Limitations of the study include a possible bias as the researcher since I was the instructor of the course.

Precautions were taken to avoid this bias: I did not grade any of the student work and did not assign final grades for the course. Another limitation is in the Math Sense of Belonging Scale survey. This survey does not contain questions that consider the classroom as one community and the small groups as another community that is a subset of the larger community. Precautions were taken into consideration for the limitation in the construct and used to guide questions during the individual interviews.

One significant delimitation of the study is the student's previous relationships with one another. Since the study investigated OpSTEM students, many of the students had taken their Precalculus courses together, during their Calculus I course many students had already established social relationships. The choice to examine women in Calculus I in this program, however, was intentional because the program contains many of the characteristics of successful Calculus I programs (Bressoud & Rasmussen, 2015) and the use of group testing is intended as an additional reform method aimed at increasing women's sense of belonging in previously established successful Calculus I programs. Another delimitation of the study was the choice to interview the students after the event in question (i.e., the group tests). The choice to interview the students after the second group test meant students had difficulty recalling their feelings during the first test due to the time lag and intervening activities between the first group test and the interview. This choice, however, was made in order to not influence the students and possibly change their group dynamics between the two group tests. Finally, the results are not meant to be generalized but instead are meant to describe the experiences of men and women in a group testing environment and can still offer insight into women's mathematical sense of belonging to the mathematics education community (Flyvbjerg, 2006).

Chapter Summary

In this chapter, I outlined the methodology I used in this study. The details I included regarding the proposed exploratory mixed method design study are intended to explain the research methodology in detail and support the use of this methodology to answer the study’s research questions. An exploratory mixed method design was appropriate for this study because I analyzed both quantitative and qualitative data sources at different phases in the study to gain an understanding of the social interactions in differently composed groups during a group test. Additionally, I attempted to understand if any of these interactions or differences in the interactions could explain the change of a woman's sense of belonging to mathematics in a Calculus I course.

CHAPTER FOUR: RESULTS

The results from this study will be presented in three parts. I first share the quantitative results of the student performance on the group and individual tests to answer the first research question:

1. In a reform Calculus I course, how do students' performance on group tests compare with their performance on individual tests; and how do gender, and the intersection of race and gender contribute to the differences, if at all?

In order to answer the first research question, quantitative data from 30 students' (i.e., eight groups) individual and group tests were compared. The test scores are first compared by gender, then by the intersection of race and gender, then by group composition. Information on how the test scores were examined and the results are in the section titled **Test Performance**.

To understand the impact of group testing on female students' sense of belonging, I present the analysis of the quantitative and qualitative data for the 15 female students in the course to answer the second research question:

2. How does the sense of belonging to a mathematics community change for female students after engaging in group testing in a reform Calculus I course, if at all?

The quantitative analysis consists of the female students' pre- and post-*MSoBS* scores and the qualitative analysis consists of the female students' post-test reflections from the first and second group test. I obtained additional clarification of the post-test reflections from interviews with the eight female students in the research groups. Information on how the data was examined and the results are in the section titled **Sense of Belonging**.

Finally, the qualitative analysis of the student interactions during the group tests for the three research groups along with the female students' post-test reflections are presented to answer the third research question:

3. How are social interactions within the group composition different based on the composition of the group, if at all, and can any differences be used to explain differences in the female students' sense of belonging?

The qualitative data from the three research groups' interactions during the two group tests were analyzed along with the post-test reflections from the women in each research group. For more information on how the interactions and post-test reflections were examined and on the process of coding the data see Chapter 3. These results are presented at the end of this chapter in the section titled **Social Interactions within Group**. I conclude this chapter with a synthesis of the commonalities and differences among the results from the three research questions.

Test Performance

To examine student performance, I analyzed the scores for the group and individual tests for 30 students. I collected data for each of the students who were placed in groups for which every group member consented to the study and completed all the tests in the course including the final exam (see Chapter 3 for more details). Students took a total of five tests during the course. Of the five tests, on average students scored higher on Group Test 1 and Group Test 2 (*M* =80.8%, *SD* =8.29% and *M* =78.9%, *SD* =12.7% respectively) than on Individual Test 1 and Individual Test 2 (*M* =68.4%, *SD*

=13.1% and *M* =69.6%, *SD* =14.2% respectively), and the students scored the lowest on the individual comprehensive final exam, with the class average of 67% (*SD* =15.3%). Overall, of the 30 students investigated, 25 completed the course successfully (i.e., 83% of students passed). To answer the first research question, I compared students' scores across gender and the intersection of race and gender. In the following subsections, I provide details about student performance and any differences from the two group tests and two individual tests.

Gender Comparison

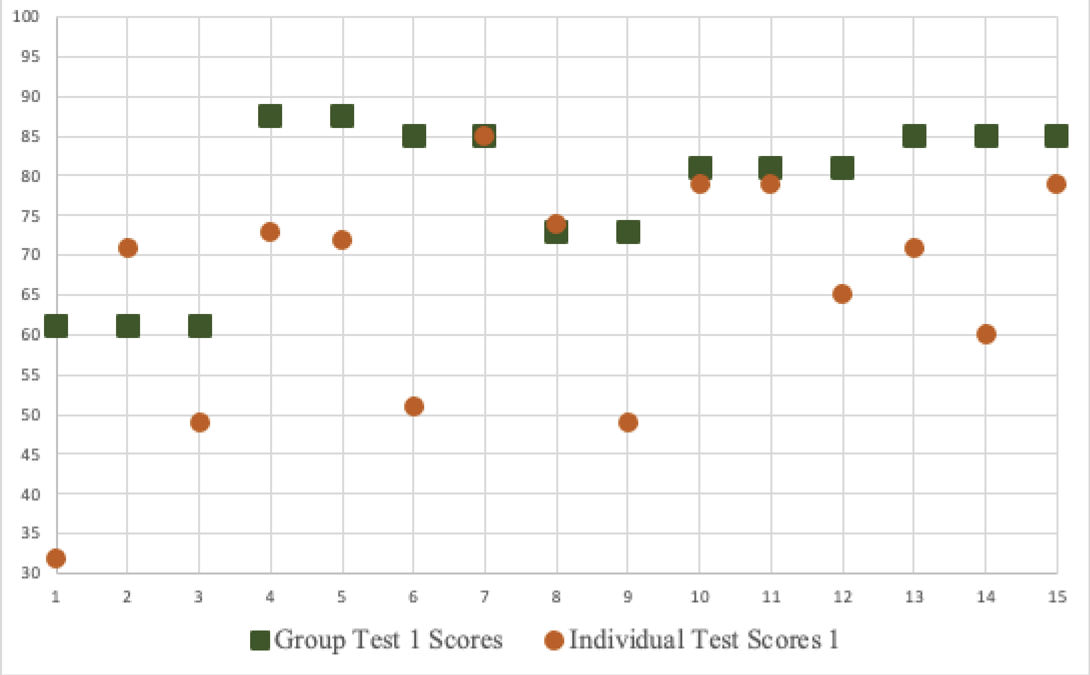
The central focus of this study is on women in mathematics so I first examined the test scores across gender for all male (*N* =15) and female (*N* =15) students to determine whether there was a difference in student performance based on gender. The male students had a higher mean score on the Group Test 1 (*M* =83.4%, *SD* =5.3%) than the female students (*M* =78.1%, *SD* =9.6%) by 5.3% on average. Additionally, the male students had a higher mean score on the Individual Test 1 (*M* =70.8%, *SD* =10.7%) than the female students (*M* =65.9%, *SD* =14.3%) by 4.9% on average. The difference in the group test to individual test scores (i.e., Group Test minus Individual test), however, was nearly the same for the male ($M_{difference}$=12.6%, *SD* =10.6%) and the female ($M_{difference}$=12.2%, *SD*=12.0%) students with only a 0.4% difference in mean scores. That is, both female and male students on average scored approximately 12% higher on the first group test than on the first individual test. So, although the male students scored higher on average on both the first group and individual tests (i.e., 5.3% and 4.9% respectively), the difference in performance from group to individual tests was nearly zero for male and female students for the first pair of tests (i.e., difference of 0.4%). See Figure 8 below for the male students test score for the first pair of tests and Figure 9 below the female students test scores for the first pair of tests. The test scores in both of these figures are arranged by group number. That is, each student that was in the same group will be next to each other on the figures below. This can be seen in each figure by the green dots, indicating the group test score, being the same for a few students in a row.

Figure 8.

Male Students' Test 1 Scores (N=15).

Figure 9.

Female Students' Test 1 Scores (N=15).



As seen in Figure 8 above, one male student scored the same on the individual and group test (i.e., student number five on the graph). This is indicated by the green square behind the orange dot on the graph for student number five. One male scored higher on Individual Test 1 than on Group Test 1 by seven percentage points (i.e., student number six on the graph). Note, these two students (student numbers five and six on the graph) were in the same group, as they were placed next to each other and had the same score for the group test. One male student scored nearly the same on the tests, scoring only one percentage point higher on the group exam than the individual test (i.e., student number 8). The remaining 12 male students all scored at least five percentage points higher on Group Test 1 than on Individual Test 1. The largest difference in scores for male students was 33% points (i.e., student number seven on the graph).

As seen in Figure 9 above, two female students scored higher on Individual Test 1 than on Group Test 1: one student by 1% point (i.e., student number eight on the graph) and one student by 10% points (i.e., student number two on the graph). One female student obtained the same scores on the group and individual test (i.e., student number seven on the graph). Additionally, two female students had very little differences between the test scores by scoring only 2% points higher on Group Test 1 than on Individual Test 1 (i.e., student number 10 and student number 11 on the graph). Note, these two females (student 10 and 11) were in the same group. The remaining 10 female students all scored at least 6% higher on Group Test 1 than on Individual Test 1. The largest difference in scores for female students was 34% (i.e., student number six on the graph).

For the second pair of tests (i.e., Group Test 2 and Individual Test 2), again I examined the test scores across gender for all male (N =15) and female (N =15) students. On Individual Test 2, the female students had a slightly higher mean score (M =70.3%, SD =16.3%) than the male students (M =69.0%, SD =11.1%) by 1.3% on average. On Group Test 2, the male students and female students had nearly the same mean scores for the second group test (M =79.4%, SD =10.4% and M =78.5%, SD =14.3% respectively) with the male students only scoring 0.9% points on average higher than females. Additionally, the difference in the group test to individual test scores (i.e., Group Test minus Individual Test) was higher for the male students ($M_{difference}$= 10.5%, SD =13.2%) than the female students ($M_{difference}$=8.2%, SD =22.0%) by 2.3%. Although the difference in test scores was higher for the male students on the second pair of tests, the range of scores (as indicated by the standard deviation) was higher for the female students. This is different from the first pair of tests in which male students had the higher average on both the first group test and individual test. That is, for the second pair of tests the female students on average scored higher on the individual test, where the male students on average scored higher on the group test. See Figure 10 below for the male students test score for the second pair of tests and Figure 11 below the female students test scores for the second pair of tests. The test scores in both of these figures are arranged by group number and are in the same order as presented in the previous figures for the first pair of tests.

Figure 10.

Male Students' Test 2 Scores (N=15).

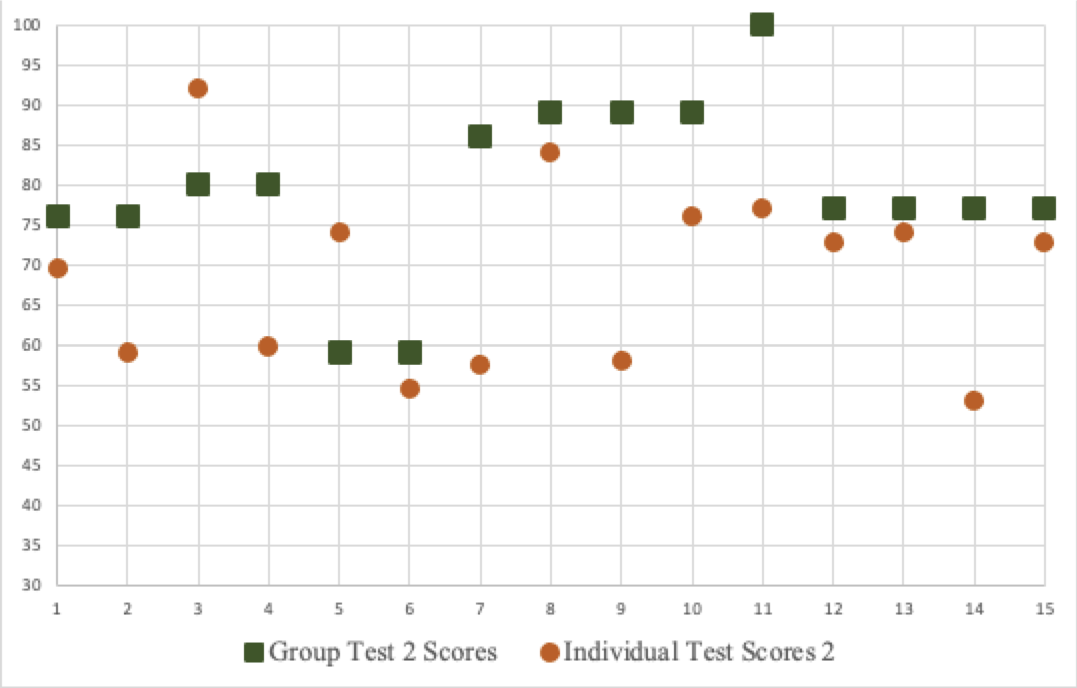
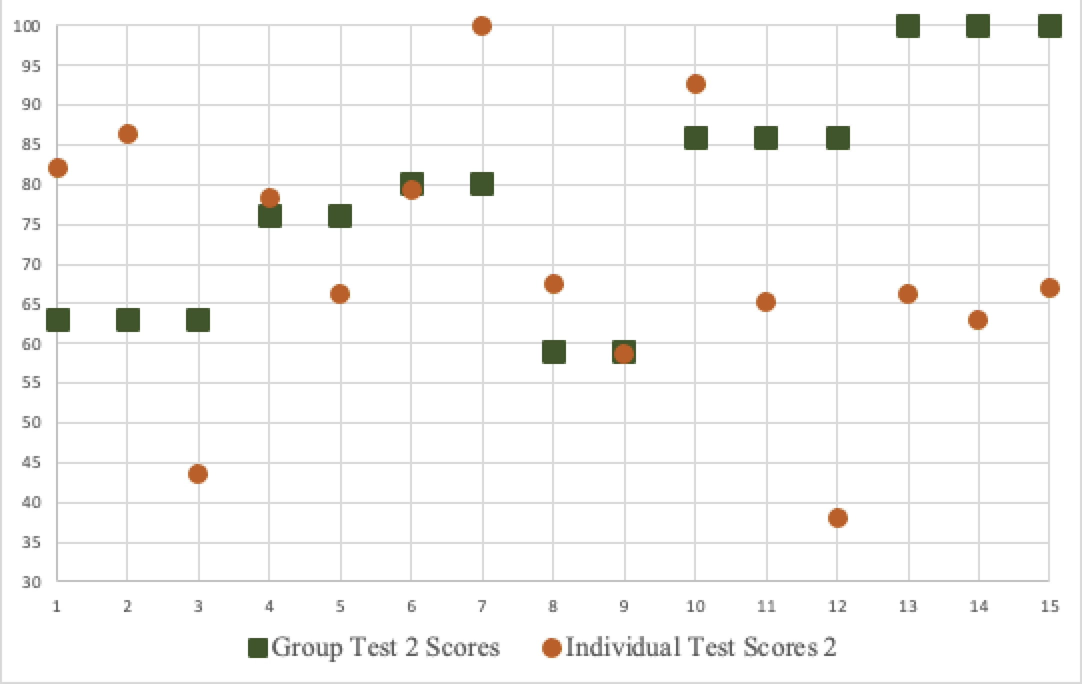


Figure 11.

Female Students' Test 2 Scores (N=15).



As seen in Figure 10, two male students scored higher on the Individual Test 2 than on the Group Test 2: one student by 12% points (i.e., student number three on the graph) and one student by 15% points (i.e., student number five on the graph). The two male students were not in the same group which is evident by the different group test scores on the graph. The remaining 13 male students all scored at least 3% higher on Group Test 2 than on Individual Test 2. The largest difference in scores for male students was 31% points (i.e., student number seven on the graph).

As seen in Figure 11 above, five female students scored higher on Individual Test 2 than on Group Test 2. Female student one scored 19% higher on the individual test and female student two scored 23.5% higher on the individual test. These two female students were in the same group, along with student number three who scored 19.5% lower on the individual test than on the group test. Female student number seven had a large difference in scores as well, scoring 20% higher on Individual Test 2 than on Group Test 2. Three other females scored higher on the individual test, but did not have as large of a difference in scores: 2.3% points for student number four, 8.4% points for student number eight, and 6.8% points for student number 10. Two of the female students scored nearly the same on both tests, but higher on the Group Test by less than 1%: student number six on the graph had a 0.7% points difference and student number nine on the graph had a 0.3% points difference in scores. The remaining seven female students all scored at least 9% points higher on Group Test 1 than on Individual Test 1. The largest difference in scores for female students was 48% (i.e., student number 12 on the graph).

Overall, the majority of the male students in this class performed better when tested in groups than when tested individually but the female students did not perform better in groups. Therefore, gender may be a factor that contributes to a difference in performance for students in group tests. To investigate whether the difference is only due to gender, I examined the score across both race and gender to explore the factor race may have in explaining these differences seen in the second pair of tests.

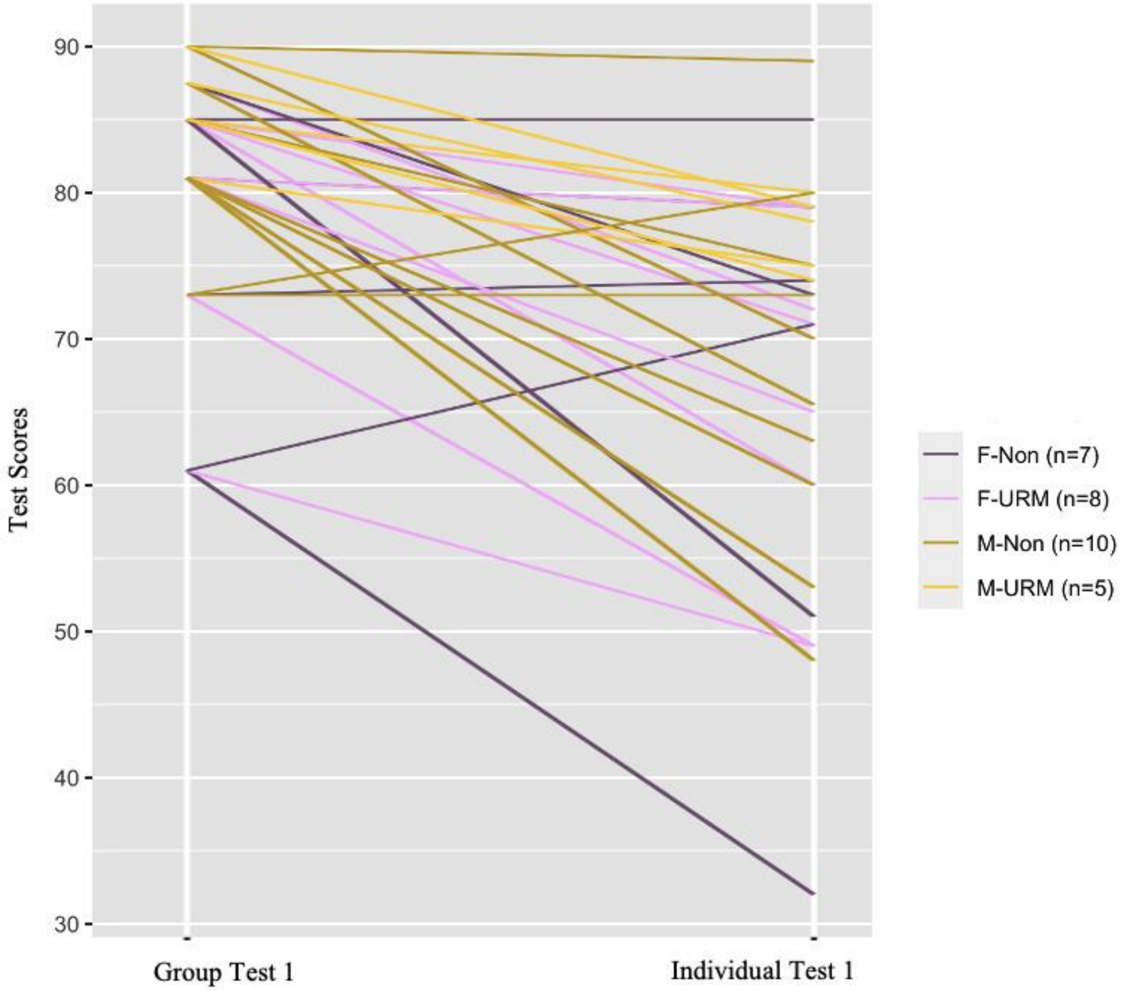
Gender and Race Comparison

I examined the groups of students by both gender and race to determine if there were differences for one specific subset of students. To do this, I took the four test scores (i.e., Group Test 1, Individual Test 1, Group Test 2, and Individual Test 2) and compared them across both race and gender which created four subsets of students. The subsets of students are: URM Male (N =5), non-URM Male (N =10), URM Female (N =8), and non-

URM Female (*N* =7). See Figure 12 for the comparison on Group Test 1 to Individual Test 1 scores for all students by subset.

Figure 12.

Comparison of Test 1 Scores by Race and Gender.

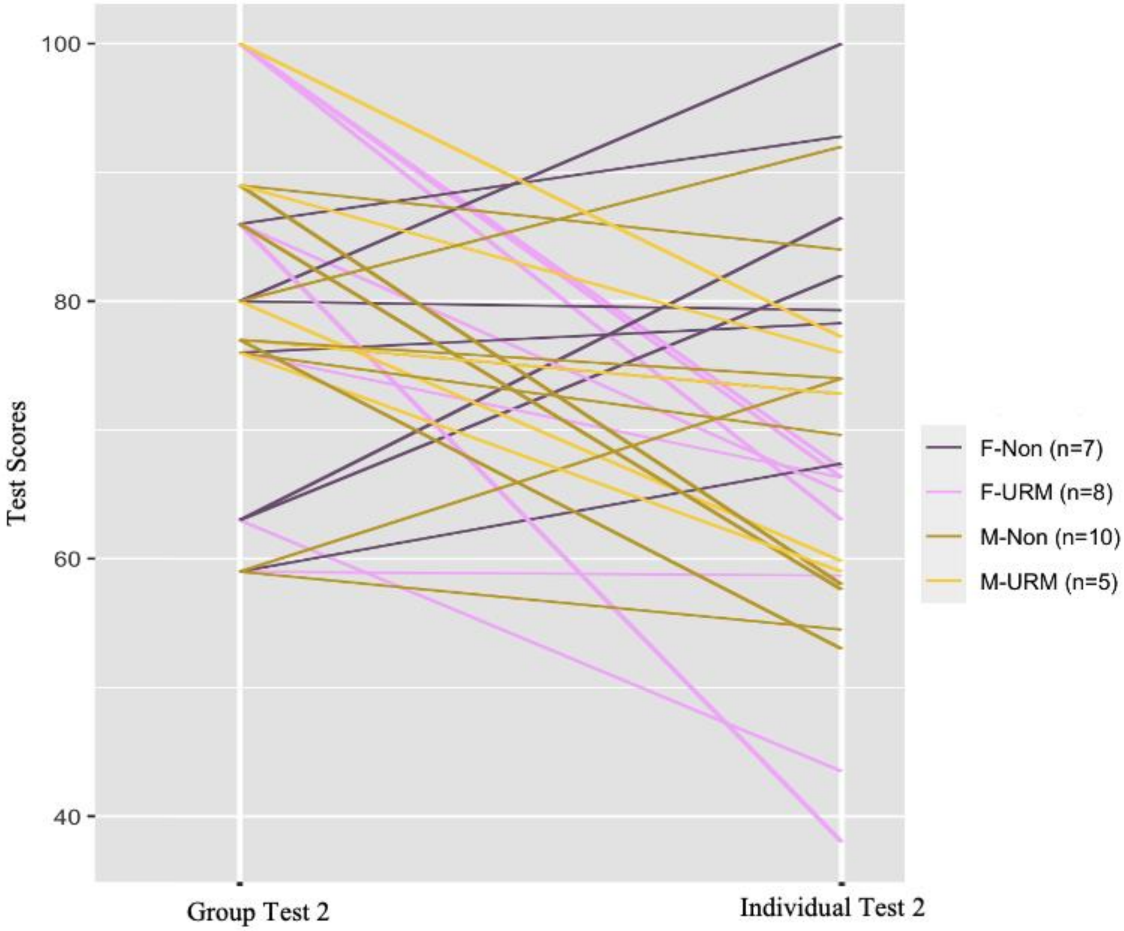


For the first pair of tests (i.e., Group Test 1 and Individual Test 1), as seen in Figure 12 and in the previous section, nearly all of the students performed better on the group test than on the individual test. Two non-URM women performed better on the individual test (by 1% and 10%) and one non-URM woman scored the same on the first group and individual test. One non-URM male scored higher on the individual test (by 7%), and one non-URM male scored the same on the first group and individual test. All URM students, regardless of gender, scored higher on Group Test 1 than on Individual Test 1.

For the second pair of tests (i.e., Group Test 2 and Individual Test 2), see Figure 13 below, all but one of the non-URM females scored higher on the individual test than on the group test. This is seen by the dark purple lines in the figure below increasing from left to right. The one non-URM female that did not score higher on the individual test had only a 0.7% difference between Group Test 2 and Individual Test 2. The two male students that scored higher on Individual Test 2 than Group Test 2 were also non-URM students as indicated by the two increasing dark gold lines in the figure. Similar to the first pair of tests, all URM students, regardless of gender, scored higher on Group Test 2 than on Individual Test 2. One URM female, however, only scored 0.3% higher on the group test.

Figure 13.

Comparison of Test 2 Scores by Race and Gender.



Overall, the majority of the students performed better when tested in groups than when tested individually except for the non-URM female students. These women (non-URM) had the highest mean score on the second individual test *and* the lowest mean score on the second group test indicating they did not perform better on the group test. Since the URM females performed better in groups and their performance in groups on average surpassed the non-URM male students, gender cannot be considered a factor alone for student performance within group testing. Therefore, it may be the intersection of race and gender that contributes to a difference in performance for the students taking the group tests.

Summary Test Performance

The test score data were investigated to see whether factors such as gender or the intersection of race and gender might contribute to differences of student performance on group tests. The majority of students performed better on the first group test than on the first individual test. When comparing the scores of the students' second group test to their individual tests, however, the non-URM women were the only subset of students to score lower on the second group test than on their second individual test. The differences seen in the second pair of tests indicates gender alone may not be a factor for performance, but instead the intersection of gender and race. To further understand the factors that contribute to the difference in the test performance for the non-URM women I examined their sense of belonging to a mathematics community. As the sense of belonging is an indicator of performance and persistence (see Chapter 2), I explore the changes of the female students' sense of belonging before and after engaging in group testing.

Sense of Belonging

In this section, I present the quantitative and qualitative results for the 15 female students regarding their sense of belonging and answer my second research question:

2. How does the sense of belonging to a mathematics community change for female students after engaging in group testing in a reform Calculus I course, if at all?

I first present the quantitative data which consists of the female students' pre- and post-*MSoBS* scores for both URM and non-URM females. Secondly, I present the qualitative data that consists of the female students' post-test reflections from the first and second group test. Finally, I explore the changes in the student pre- and post- *MSoBS* scores through the findings from the qualitative analysis.

Quantitative *MSoBS* Results

I examined the pre- and post-*MSoBS* survey scores separately for the URM and non-URM females for the: (1) overall *MSoBS* score, (2) Membership composite score, (3) Trust composite score, (4) Desire to Fade composite score, (5) Affect composite score, and (6) Acceptance composite score. Recall from Chapter 3, the sense of belonging scores were classified as high, medium, or low. In this section, I describe the change in *MSoBS* for the following:

1. Scores that increased (i.e., pre score was in the low range and post score was in either the medium or high range, or pre score was in the medium range and post score was in the high range).
2. Scores that stayed in the same range (i.e., pre and post scores were both within the high range, or pre and post scores were both within the medium range, or pre and post scores were both within the low range.)
3. Scores that decreased (i.e., pre score was in the high range and post score was in either the medium or low range, or pre score was in the medium range and post score was in the low range).

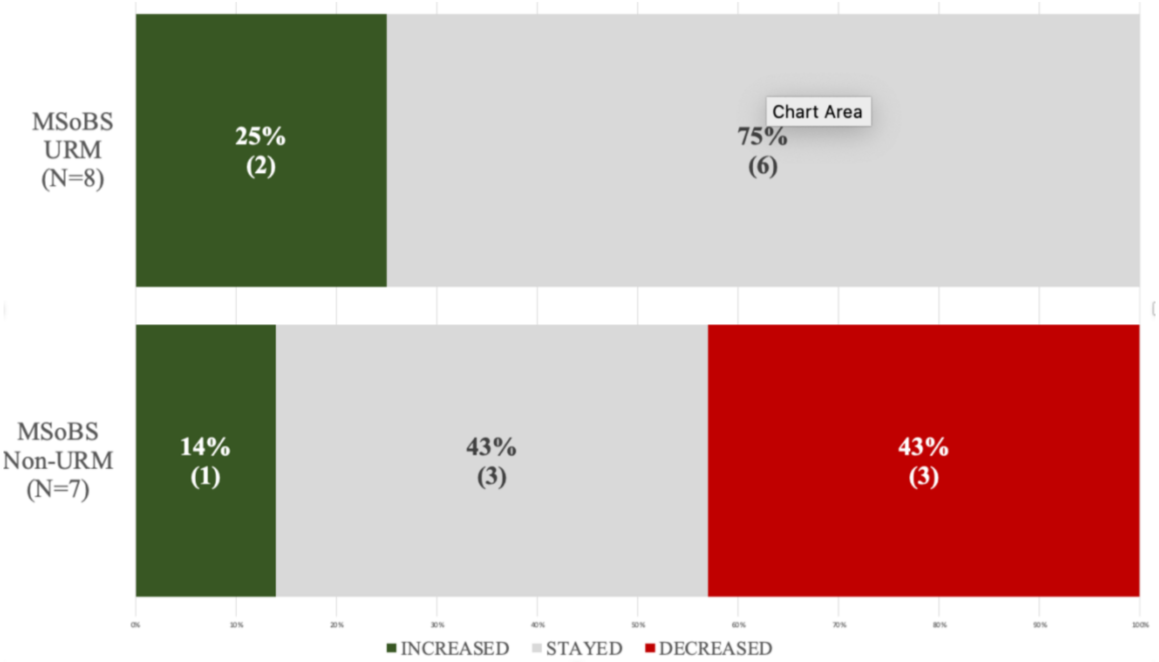
For the URM females, one student had an increase in their overall *MSoBS* composite score from a low *MSoBS* score of 3.83 to a medium score of 4.47. One URM female increased from a medium *MSoBS* score of 5.17 to a high score of 6.57. Four URM female pre- and post-*MSoBS* scores stayed in the high range, one stayed in the medium range, and one stayed in the low range. None of the URM female decreased in their composite *MSoBS* score.

For the non-URM females, one student had an increase in their overall *MSoBS* composite score from a medium *MSoBS* score of 5.47 to a high score of 6.40. Two non-URM female pre- and post-*MSoBS* scores stayed in the high range and one stayed in the medium range. One URM female decreased from a high *MSoBS* score of 6.13 to a low score of 3.30, and two decreased from high *MSoBS* scores of 6.23 and 6.70 to medium scores of 5.50 and of 4.30 respectively. See

Figure ***14*** for the relative frequency and count for the change in the pre- and post-*MSoBS* composite score by URM status of all female students. As seen in the figure and described above, more non-URM female students decreased in their overall sense of belonging score. Additionally, more URM females increased or remained the same in their overall sense of belonging score.

Figure 14.

Changes in MSoBS Composite Score by URM Status.



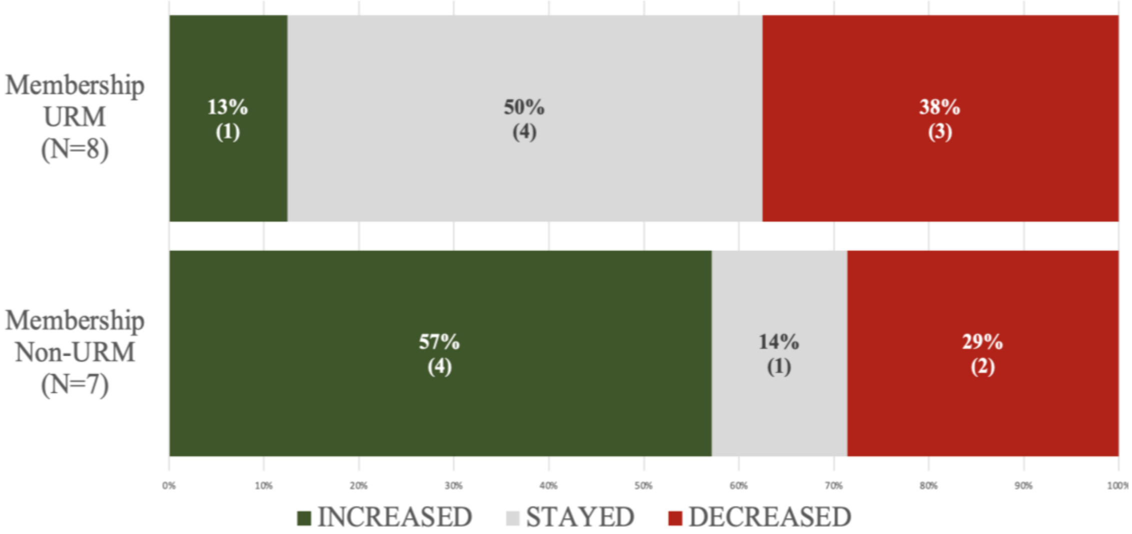
The overall *MSoBS* composite scores are calculated from the average of the student responses of the 30 questions. Each of the questions falls into one of the five components of belonging (i.e., membership, trust, desire to fade, affect, and acceptance). Each component has numerous questions in which I compiled the average response to obtain the composite score for the five components of belonging. Because this difference seen in the pre- and post *MSoBS* scores above may or may not be seen in each of the components for the URM and non-URM females, I examined each component separately.

For the Membership component (See Figure 15 below for the relative frequency and count for the change Membership composite scores by URM status of all female students) there is a contradictory pattern of increasing/decreasing than seen in the overall *MSoBS* scores. For the URM females, one student had an increase in their Membership composite score from a medium score of 5.00 to a high score of 8.00. Two URM female

Membership scores stayed in the high range, one stayed in the medium range, and one stayed in the low range. One URM female decreased from a high Membership score of 7.00 to a low score of 3.75, one decreased from a high score of 6.50 to a medium score of 5.00, and one decreased from a medium score of 4.00 to a low score of 3.00. For the non-URM females, four students had an increase in their Membership composite scores from medium scores of 4.00, 4.50, 5.00, and 5.75 to high scores of 8.00, 6.50, 6.75, and 6.75 respectively. One non-URM female Membership composite score stayed in the high range. One URM female decreased from a high Membership score of 6.00 to a medium score of 4.00, and one decreased from medium Membership scores of 4.00 to a low score of 2.25. Although more non-URM females overall *MSoBS* scores did not increase, it seems more of the non-URM females did feel as if they were a member of the mathematics community.

Figure 15.

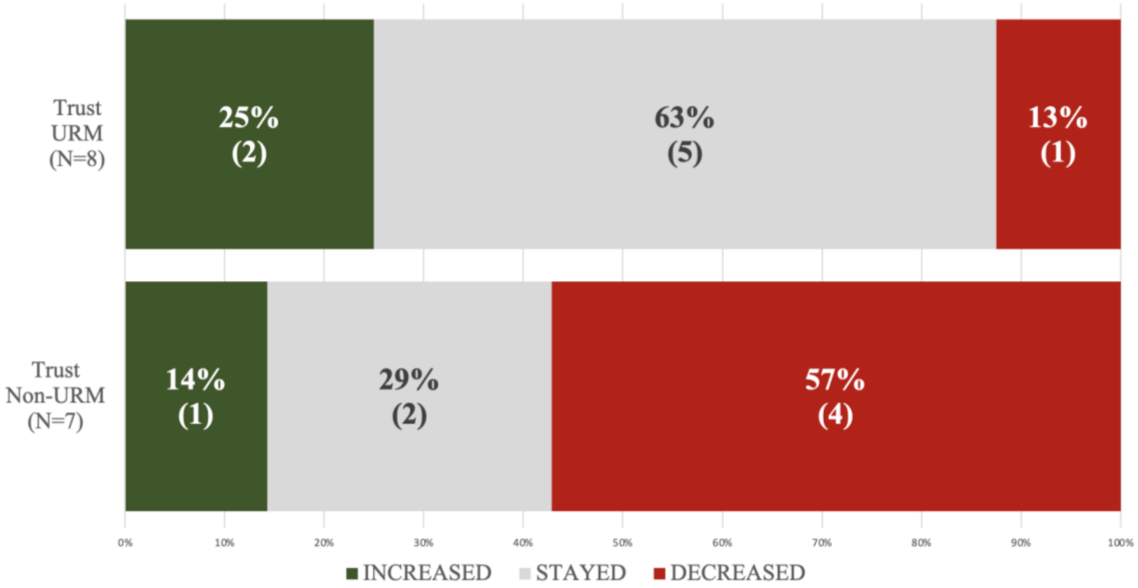
Changes in Membership Score by URM Status.



For the Trust component (see Figure 16), two URM females had an increase from medium scores of 4.50 and 5.00 to high scores of 6.00 and 6.25 respectively. Five URM female Trust composite scores stayed in the high range. One URM female decreased from a high Trust score of 6.50 to a medium score of 5.50. For the non-URM females, one student had an increase in their Trust composite score from a medium score of 5.00 to a high score of 6.25. Two Non-URM females stayed in the high range. One non-URM female decreased from a high Trust composite score of 6.25 to a low score of 3.50, and three decreased from high scores of 7.25, 7.00, and 6.75 to medium scores of 5.50, 5.25, and 5.50 respectively. Similar to the overall *MSoBS* scores, more non-URM Females *decreased* in the Trust component where the majority of the URM students already had a high measure and stayed in this range.

Figure 16.

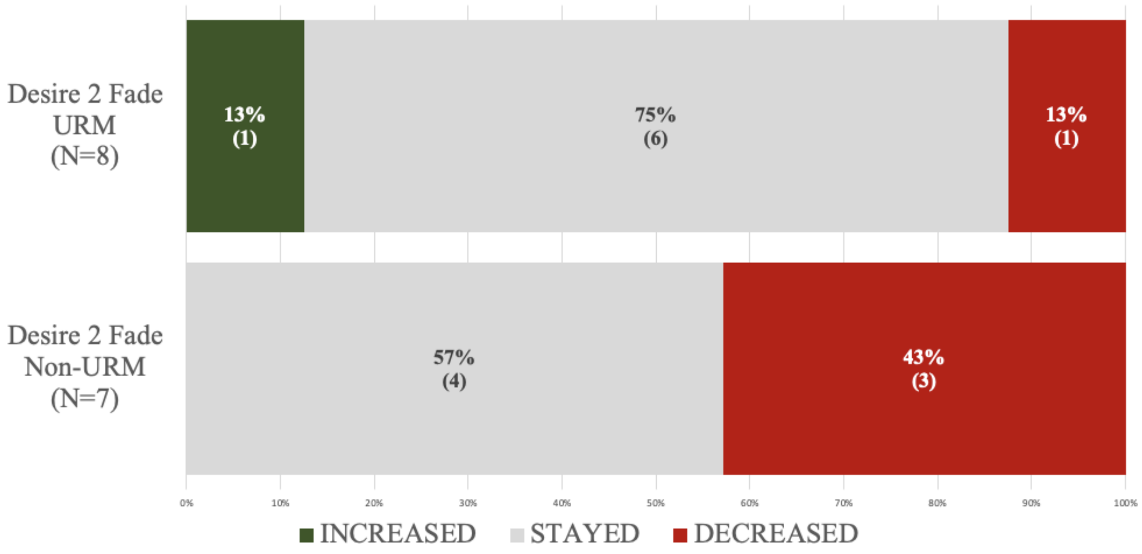
Changes in Trust Score by URM Status.



For the Desire to Fade component (see Figure 17), one URM female had an increase from a medium score of 5.25 to a high score of 7.25. Six URM female Desire to Fade composite scores stayed within the high range. One URM female decreased from a high score of 6.50 to a medium score of 5.00. For the non-URM females, none of the females had an increase in their pre- and post-Desire to Fade composite scores. Three stayed in the high range and one stayed within the low range. Three decreased from high Desire to Fade scores of 8.00, 7.00, and 6.50 to medium scores of 4.75, 5.75, and 5.25 respectively. As seen in the figure below, the majority of both URM and non-URM females did not have a change in scores for the Desire to Fade component. Additionally, more non-URM females decreased in this component than URM females.

Figure 17.

Changes in Desire to Fade Score by URM Status.

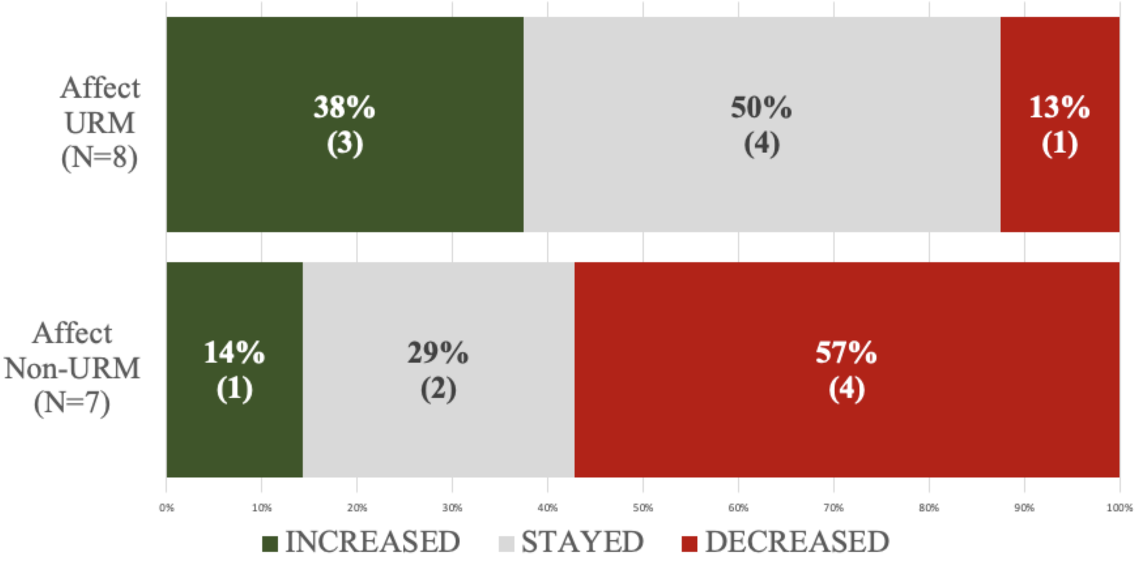


For the Affect component (see Figure 18), two URM females had an increase from medium scores of 5.88 and 5.88 to high scores of 6.00 and 6.38. One URM female had an increase from a low score of 3.63 to a medium score of 4.00. One URM female

Affect composite score stayed within the high range, one stayed in the medium range, and two stayed within the low range. One URM female decreased from a high score of 6.13 to a medium score of 4.38. For the non-URM females, one had an increase in their Affect from a low score of 3.75 to a medium score of 5.63. One non-URM female stayed in the high range and one stayed within the medium range. One decreased from a high Affect score of 6.63 to a low score of 3.50. Two decreased from high scores of 7.25 and 6.75 to medium scores of 5.88 and 5.13 respectively. One decreased from a medium Affect score of 5.88 to a low score of 2.88. As seen in the figure below, the change in Affect scores for URM and non-URM females were opposite of each other. More non-URM females had a decrease in their Affect score and more URM females either maintained within their original range or increased in their Affect score.

Figure 18.

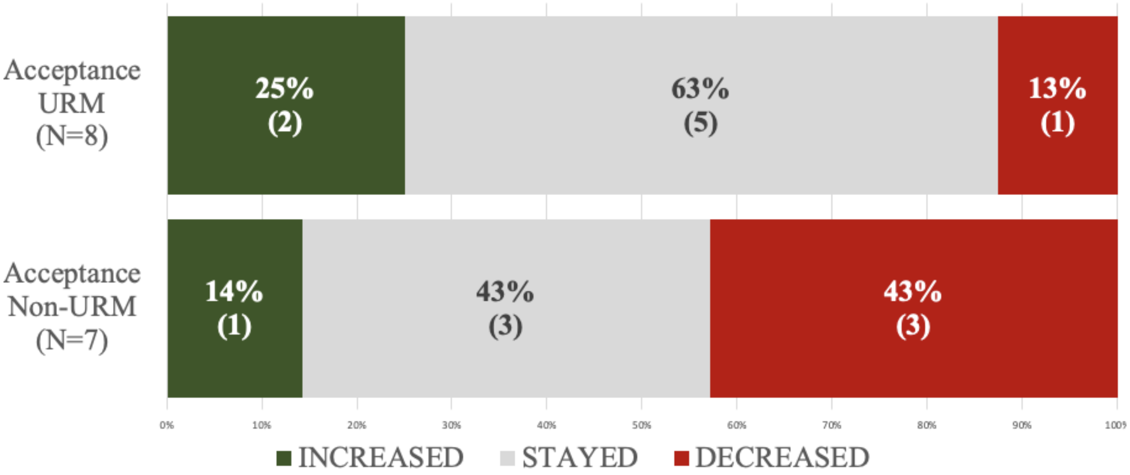
Changes in Affect Score by URM Status.



For the last component, Acceptance (see Figure 19), one URM female had an increase from a medium score of 4.70 to a high score of 7.20. One URM female had an increase in their Acceptance score from a low score of 3.00 to a medium score of 4.90. Four URM females stayed in the high range and one stayed in the medium range. One URM female decreased from a medium Acceptance score of 4.00 to a low score of 2.90. For the non-URM females, one had an increase from a medium score of 5.40 to a high score of 6.60. Three non-URM females stayed within the high range. One decreased from high Acceptance score 6.40 to low score 3.40. Two non-URM females decreased from high scores of 6.70 and 6.10 to medium scores of 4.00 and 5.30 respectively. As seen in the figure below, more non-URM females decreased in their Acceptance scores than URM females. Additionally, the majority of URM females maintained their Acceptance scores within the same range from the pre- and post-survey.

Figure 19.

Changes in Acceptance Score by URM Status.



Overall, the pre- and post-*MSoBS* scores did not change for the majority of the URM females, increased for the non-URM females in the Membership component, and decreased for the non-URM females in both the Trust and Affect components. Additionally, the majority of the non-URM females either stayed the same (i.e., three out of seven) or decreased (i.e., another three out of seven) in the Acceptance component. The only component for which the majority of all females had no increase nor decrease (i.e., stayed the same) was in the Desire to Fade component. Although the URM females did not have a majority increase in any of the *MSoBS* components, indicating their sense of belonging did not change over the time in the course, their performance in the second group test did surpass the non-URM males and was nearly the same as the URM males. The testing method, however, did not have the same impact on the non-URM females. The majority of the non-URM females had a decrease in both Trust and Affect, and also were the only subset of students to score lower on the group test than on the individual test. This finding may imply that the non-URM and URM female students had different experiences while engaged in the group tests. To understand more about how the female students felt about engaging in group testing, I examined their post-test reflection responses using the *MSoBS* components and questions as my guide (see Chapter 3 for analysis procedures). In the following section, I provide the qualitative results from the post-test reflections by race and group composition.

Qualitative Post-Test Reflection Results

To understand how the group testing environment may have affected the female students' sense of belonging I examined their post-test reflections. I first present the codes from the inductive and deductive coding process as a result of the qualitative analysis (see Chapter 3). Before presenting the themes that emerged from the analysis, I first discuss these eight codes that were compiled for the non-URM and URM females separately. I used a treemap chart to visualize the hierarchical ordering of the codes and the increase or decrease of the female students *MSoBS* that emerged for the students' post-test reflections (see Figure 20 for the non-URM females and Figure 21 for the URM females). Note in the treemap charts, rectangles with a green shade indicate an increase or positive theme, rectangles with an orange shade indicate a decrease or negative theme, and rectangles with a yellow shade indicate neutral theme.

Figure 20.

Treemap of Codes for non-URM Females (N =7) Post-Test Reflections.

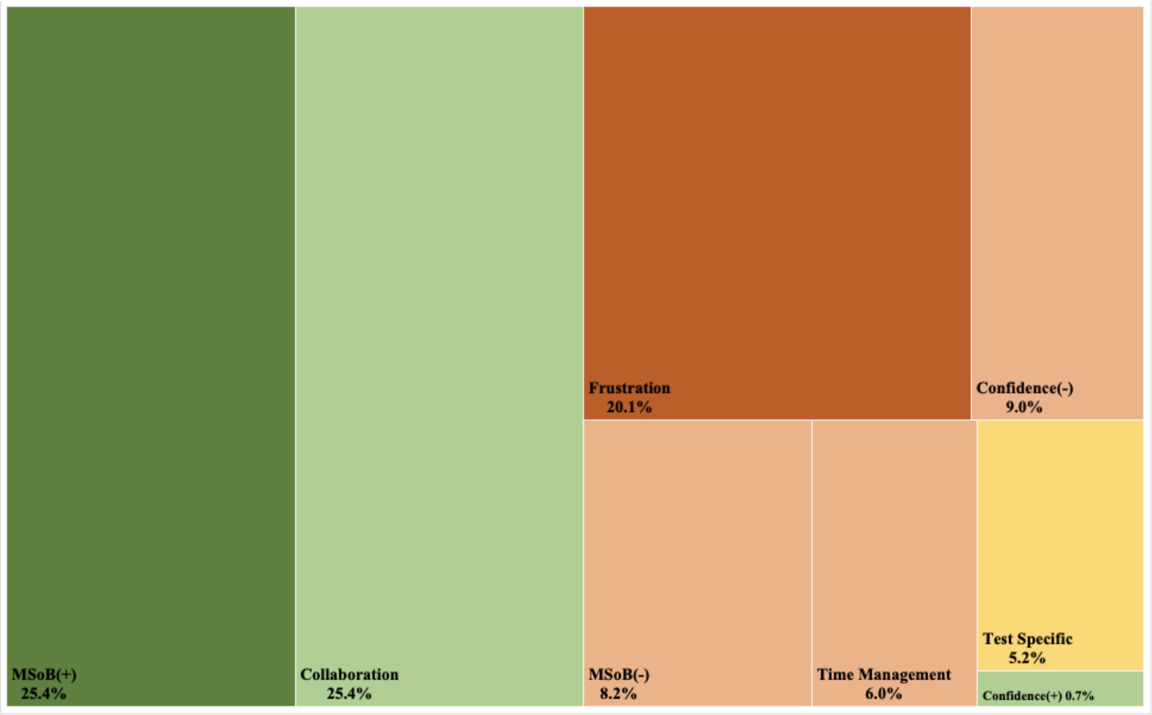
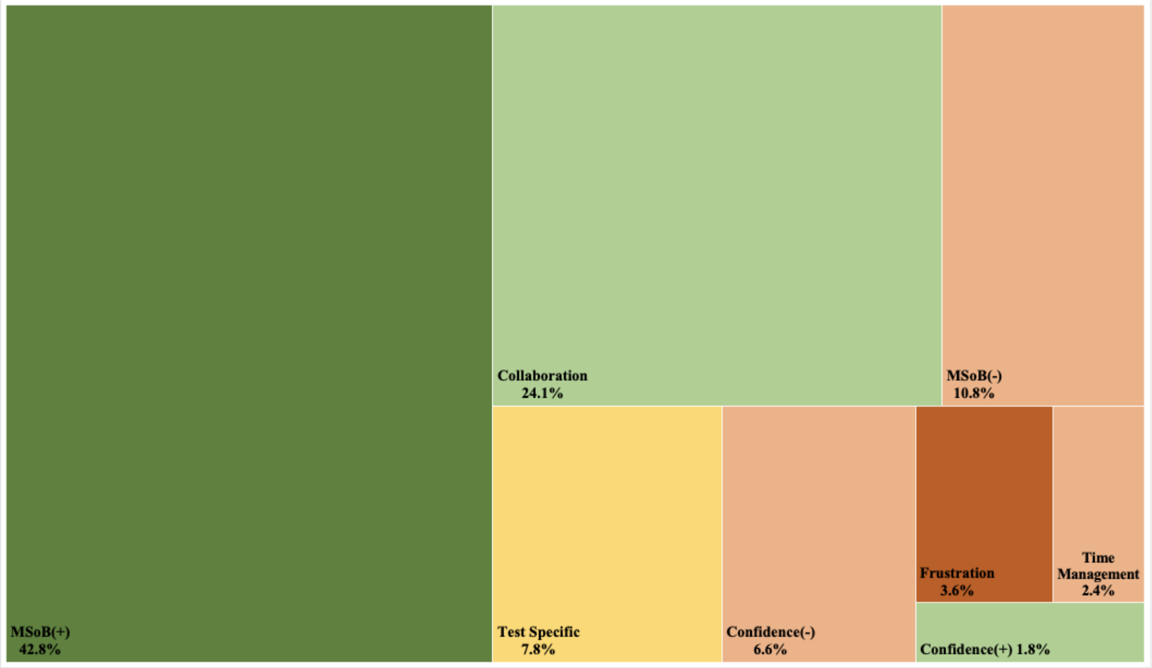


Figure 21.

Treemap of Codes for URM Females (N =8)Post-Test Reflections.



The results of the post-test reflections seen in the figures above may help to explain some of the decrease seen in the quantitative *MSoBS* results for the non-URM females. In the first figure, the non-URM had almost the same amount of negative and positive comments in their post-test reflections. Around 50% of their comments did indicate they had a good experience with respect to the collaboration in their group (i.e., Collaboration 25.4%) and feeling as if they were an accepted member of their group (i.e., *MSoBS* (+) 25.4%), the smallest positive theme was their confidence as only 0.7% of their comments indicated an increase in their confidence. The URM females had nearly the same frequency of positive comments regarding the benefit of the group's collaboration (i.e., Collaboration 24.1%), however, they had more positive comments regarding their affect, acceptance, and membership in their groups (i.e., *MSoBS* (+)

42.8%). Additionally, the URM females had a slightly larger increase in confidence (i.e., Confidence (+) 1.8%) than the non-URM females.

For the second level of coding, to refine the categories each coded comment from above was examined and similarities were identified and compiled as overarching themes. The female student responses from their post-test reflections illuminated four main themes: collaboration, confidence, testing, and control. Each theme directly addressed the purpose of the study (i.e., the possible effect of group testing on one's sense of belonging to mathematics), and each theme either positively or negatively impacted the female students' sense of belonging. The collaboration and confidence themes have positive and negative subthemes, the testing and control themes are both negative themes without any subthemes. In the following subsections, I present the definition for each theme and subtheme, and include examples from the students' post-test reflections in which each theme was present. I then provide the frequency of the occurrence of themes for the non-URM and URM females. The results presented are separated by race to better understand the differences previously seen in both the test performance and quantitative *MSoBS* results.

Collaboration

Collaboration as defined in Chapter 1 is when students work on a problem together without individual work time to build their mathematical ideas as a group. The collaboration theme has both a positive subtheme, *beneficial,* and a negative subtheme, *wishful.* The beneficial and wishful collaborative subthemes emerged from comments in which students referred to the conversations they had or did not have during the group test in terms of either voicing their understanding and mathematical ideas or receiving feedback. That is, either the group was collaborating during the test, or a student expressed a need for more collaboration with their group members during the test (i.e., wanting feedback). Comments in this theme directly relate to the measured *MSoBS* quantitative scores. The beneficial collaboration subtheme includes comments that were coded with membership (+), affect (+), acceptance (+), trust (+) and desire to fade (+). The wishful collaboration subtheme includes comments that were coded with membership (-), affect (-), acceptance (-), and desire to fade (-). The only component of the quantitative *MSoBS* scores not included in this theme was trust (-), as these comments were merged with the other comments regarding the testing material and will be seen later in this section.

The collaboration subtheme, referred to as beneficial collaboration, appeared from comments regarding positive occurrences of the groups working together, receiving feedback from group members, having others to talk through the test problems, and the group members coming together as a team. For example, one student commented “We all did talk through each problem before we started the test. I felt that I was able to share anything I needed to say and got great feedback.” In this comment, the student notes that her ideas were both shared and well received by her group members (i.e., “great feedback”) as they attempted to answer each problem. In another example, one student noted how her group acted as a team when she wrote “Our group really joined forces and worked together pretty well for this test.” The student elaborated on the “joined forces” comment by discussing the change in the group dynamics her group experienced during the course. She said “our group used to be very quiet, but I feel that we really powered through the test today as [a] full group.” This comment shows the benefits of the group coming together as a team and the change that can happen in the group when working together. Finally, consider Mary's response when asked if she felt she communicated her mathematical ideas during the test she said "Yes, I feel like when I had an idea or concept or even a question my group listened to what I had to say and responded with relevant and helpful comments." In this statement, Mary notes that not only did she voice her ideas but also questioned her group members' ideas. Additionally, she received feedback and explanations from her group members during the test. This comment illuminates the positive discussions and collaboration her group had during the test and how she was comfortable with asking questions to her group members.

The negative aspect of the collaboration theme, referred to as wishful collaboration, appeared from comments regarding occurrences of the group members not working well together, only one member speaking for majority of the time, and exclaiming a desire for others to share ideas more often during the test. For example, one student commented "I kept having to ask my team members what their thoughts were. They did not volunteer much information." This comment is a good example of how the student wished her group members would have collaborated more and talked more during the test. Another example of how the negative aspect of collaboration emerged from students who commented they would have liked feedback from their peers. Consider Annie's comment from the first post-test reflection, she stated:

> *My ideas and thoughts were not listened to and disregarded by all of my group members. Even if my answers were wrong, I would have liked it if someone could have told me that I was wrong and explained why.*

This comment shows how Annie would have liked her group to have responded to why she was wrong and explained what they were thinking (i.e., wanting collaboration) during the group test. Overall, the comments regarding collaboration were either about the benefit the students felt from working together (i.e., beneficial), or a wishful note for more conversation or feedback from their peers during the group test (i.e., wishful). The beneficial collaboration positively impacted the female students' sense of belonging as it made the women feel as a listened to and valued member of the group. Feeling as if one is a listened to member of a group relates to the acceptance component on the *MSoBS* (i.e., not feeling discarded, neglected, or excluded from the group). The wishful collaboration negatively impacted the female students' sense of belonging as it made them feel neglected or a disregarded member of the group. That is, the wishful collaboration relates to a lowered feeling of acceptance within the group.

Confidence

Confidence refers to one's personal belief in their own mathematical understanding and knowledge. The confidence theme has both a positive subtheme, *expressed*, and a negative subtheme, *need for*. The expressed confidence subtheme and a need for confidence subthemes emerged from comments in which students noted their own knowledge or abilities on the post-test reflections.

The *expressed* subtheme appeared from comments regarding one having a better understanding of the mathematics, feeling more prepared to take the individual test, and being more willing to voice one's ideas or questions. For example, one student stated in her second post-test reflection: "I had way more input and knowledge to add to the test this time rather than the first time around." To be able to offer more input, ideas, and information to the group this student showed an increase in confidence from the first to the second group test. Noting how she was able to offer more knowledge and gave more input was an expression of her confidence in her own knowledge and abilities to help her group. Another example, Mary stated “I felt like we had good conversations going over the concepts in a group test is a warm up for the individual test” which indicates she felt more prepared for the individual test after the group test. Last, consider what Pratibha wrote on her second post-test reflection “I think for second test I did about 40% contribution, I felt that I know something.” This comment is an example of a student feeling as if they have a good understanding of the mathematics (i.e., “I felt I know something”).

The *need for* subtheme appeared when subjects made a direct comment about having lowered confidence in one’s own abilities, fear of voicing one’s ideas, or only giving help if one was sure of the answer. There were a few comments that contained statements specifically about one’s own self-confidence after the test. For example, one student said “Group tests always manage to lower my self-esteem and confidence. I would be curious to see if people score better on individual exams when they are given first.” This comment was after the second group test and came from a student who had experience with group testing during high school. Although the comment does not address why the group test lowers her self-esteem and confidence, the impact was still noted as a negative impact on her sense of belonging due to her directly addressing her confidence. Additionally, after the second test another student commented “Telling each other what to do. Nitpicking. Causes high anxiety and impatience. Negatively impacts confidence for individual tests.” Unlike the first example, this student explained why the testing environment lowered her confidence (i.e., telling and nitpicking). Comments that did not explicitly state a decrease in confidence in which this theme was present include sentiments similar to the following comment:

> *Because this test was very daunting/intimidating because of the amount of material I was more quiet during this group test and I didn't want to speak in fear of confusing people or holding the group back. When I did communicate my ideas I felt they were taken into more consideration than the first group test. I do think I communicated pretty clearly.*

This student did not have the confidence in herself that she would not confuse her group members by voicing her ideas and questions. This comment is classified as *need for* because if she had stated an idea or question it may have changed the course of an answer for the group. However, her fear of confusing the group manifested during the second test and she was less confident in her mathematical ideas (i.e., "hold the group back"). Finally, on the first post-test reflection, when asked if she felt she contributed her mathematical ideas to the test Sophie wrote "I did if I understood the question and felt like my input would benefit the group." This response is an example of a student offering ideas only if they knew they were sure about the mathematics, and is an example of a student being unsure of their ideas (i.e., need for confidence) and not fully participating in a group discussion in fear of not benefitting the group's progress.

Overall, the comments regarding confidence were either regarding the noted increased confidence the students felt from working together (i.e., expressed), or a lowering or need for confidence in themselves and their mathematical understandings (i.e., need for). The expressed confidence positively impacted the female students' sense of belonging as it made the women feel confident in their own mathematical understanding. The need for confidence negatively impacted the female students' sense of belonging as it made them either stressed to take their individual test or unsure and fearful of holding their group back during the group test.

Testing

Testing refers to all general testing in mathematics and general group testing comments. Recall that trust of testing material to be unbiased and the instructor's commitment to helping all students learn is a subcomponent of students' mathematical sense of belonging. This theme has no subthemes: there is only one aspect to this theme and its relationship to the female students' sense of belonging is negative. Testing dislike emerged from comments about a general dislike for testing (i.e., both in groups and individually), general comments about the procedures required for the group testing for the study (i.e., using a pen), or comments about the time allowed to complete the test. The general dislike of tests arose from comments such as one Kathy's said "I just don't like exams, but I'm not sure if that counts as a negative comment" in response to the question "Please provide one negative comment about your experience in the group test." In her response, Kathy admits to not like testing in general. She explained in her interview that testing increases her anxiety, and that adding the group part of testing did not elevate her feelings of anxiousness. A couple of students who had experience with group testing in high school commented on their general dislike for group tests. For example, after the first group test Maryam wrote on her post-test reflection "I don't like group tests. I find them to always be hard (I had to do group tests in high school math and hated it)." When interviewed, I asked Maryam if she could elaborate on this comment and she said she could not explain why, but the test themselves just feel harder to her than individual tests and stated "I don't feel I get any benefit from them." This comment was in response to all group testing she had experienced both in this course and in high school.

Testing dislike that was centered around how the group test was implemented in this course was generally due to having only one pen and one paper. For example, one student wrote “[We] worked well as a team. Don’t like that we have to use a pen for a math test. Kind of stressful when we are discussing and crunching for time.” This comment is an example of both the use of a pen, as students are generally not used to using pens in mathematics classroom, and the allotted time given to complete the test (i.e., “crunching for time”). Other students' comments about the time management were more explicit. For example, Annie’s response for providing one negative comment to the second group was “time management was not taken into consideration.” During her interview Annie elaborated on her comment. She said that during the second test she placed herself as the time keeper for the group and told everyone they each had 15 minutes to spend on their problem. However, this did not happen as Isaac took over 20 minutes for the first problem and Sophie took over 20 minutes for the second problem. By the fourth and last problem, in which Annie was assigned as the writer, she was left with less than 10 minutes to complete her part of the test. She explained how her group not taking time management into consideration made her feel stressed and rushed during her problem on the second test.

Overall, the comments regarding testing were regarding general dislike for all tests or group tests, using a pen for mathematics, and time allotted for the group test in this course. These comments were considered to have a negative impact on the female students’ sense of belonging as they induced stress and made the students feel less at ease. Additionally, time management alone can make students feel more tense and more nervous in their community during the group test.

Control

Control refers to having power, authority, or influence over how something is completed. This theme does not have multiple subthemes and, instead, is a single negative theme with respect to the female students' sense of belonging referred to as lack of control. The lack of control emerged from comments about the frustration students had during the test. Frustration included comments about experiencing problems with group members, not being able to write on the paper, and the group disagreeing during the test. Not being able to write on the paper is included in the control theme and not the testing theme because this was part of the purposeful design of the group test for this study. That is, students were only allowed to hold the pen and paper for one specific problem and had to talk through any problem in which they were not the assigned writer. Recall these tests were designed to increase student-to-student communication. That is why this theme is specific for the group tests in this study and does not fall under the testing theme. For example, after the second group test one student wrote:

> *I personally have never felt more stressed and nervous before a test, not because I don't know the material but when [I] am only allowed to write one problem from a test some group members are struggling on basic algebra concepts during the test I get very frustrated and this test today caused a lot of tension in our group.*

This student directly states how the testing environment was frustrating particularly because she could not write on the test for more than one problem. However, if she was able to control how the mathematics was written for more problems (i.e., write on the test herself) she may have not noted the difficulty her fellow group members had with algebra. Another student commented about the frustration of having one writer per problem for the group test. On the first post-test reflection, this other student wrote “It was sometimes hard to explain your mathematical thinking without just showing them what you mean.” This frustration again of not being able to “just show” her work suggests if she had control over the paper, it would not have been as hard for her to explain her mathematical thinking. Last, the control over how the mathematics was present on the group test was also a frustration for some students. For example, Maryam “I also wish that only one person wrote so that all of it is done neatly. I liked mine and Georgia’s writing” Her comment still alludes to an attempt to control how the answers are put on the paper for the test, and hence is included in the lack of control theme.

The main difference between the testing and control theme is the setting of the group test for this study. That is, this theme (i.e., control) may have not emerged if students were able to work on their own papers and then come to a group consensus. However, the group tests were designed for collaborative and not cooperative group work as defined in Chapter 1. So, in order to increase student-to-student talk and have less occurrences of students just showing their work to their group members without talking about the mathematics, the students in this course were not allowed to write on the paper for more than one problem per test. Therefore, these comments regarding the frustration they felt about not being able to just write on the paper themselves caused a different frustration than seen in the general testing theme.

Summary of Themes

Analysis of the female students’ post-test reflection illuminated several themes that directly address the purpose of this study – namely, to understand the possible effect the social interactions during the group test may have had on female students' sense of belonging. Looking across all of the post-test reflections, two themes could positively impact students' sense of belonging (beneficial collaboration and expression of confidence), and four themes negatively impact students' sense of belonging (wishful collaboration, need for confidence, testing dislike, and lack of control). In the following section, I provide the frequency of themes present in the overall post-test reflections for the female students separated by both URM status and then group composition.

Frequency of Themes

The frequency of occurrence is the percentage of the number of times a comment was coded for each theme. Total counts were gathered for each of the five questions on the two post-test reflections. I am reporting the overall frequencies (see Table 14 for the number for each theme, or overall count, is also provided in the table) in order to gain a general understanding of how the group testing impacted the female students' sense of belonging. Again, the results are separated by URM status due to the differences already seen in non-URM and URM female student performance and quantitative sense of belonging measure (i.e., *MSoBS*). Table 14 shows that both the non-URM females and URM females had similar total counts for all of the themes (i.e., non-URM total of 102 and URM total of 96) with the highest occurrence in the beneficial collaboration theme. However, the URM females did have higher frequencies of beneficial collaboration (56%) than the non-URM females (38%). Additionally, the non-URM females had higher frequencies of wishful collaboration (14%), testing dislike (11%), and lack of control (16%) as compared to the URM females (i.e., 5%, 8%, and 5% respectively). Both non-URM and URM females had similar frequencies for the *expressed* and *need for* confidence, with the URM female having slightly higher frequencies in both categories. Specifically, the non-URM females had 10% in the *expressed* and 11% in the *need for* confidence, where the URM females had 12% in the *expressed* and 14% in the *need for* confidence.

Table 14.

Overarching Themes from Post-Test Reflections Regarding the Experience of Group Testing on Female Sense of Belonging by URM Status.

	Collaboration		Confidence		Testing	Control
	Beneficial[a]	Wishful[b]	Expressed[a]	Need for[b]	Testing Dislike[b]	Lack of Control[b]
Non-URM Females (N=7)	38% (39)	14% (14)	10% (10)	11% (11)	11% (11)	16% (17)
URM Females (N=8)	56% (54)	5% (5)	12% (11)	14% (13)	8% (8)	5% (5)

[a]Themes or subthemes that had a positive impact on their sense of belonging.
[b]Themes or subthemes that had a negative impact on their sense of belonging.

When combining the positive (beneficial collaboration and expressed confidence) and negative themes (wishful collaboration, need for confidence, testing dislike, and lack of control) in the Table 14, it appears the non-URM females seemed to have a more negative experience than the URM females. Specifically, 52% of the non-URM females' comments explain that group testing negatively impacted their sense of belonging, where URM females had 32% of their comments that showed a negative impact. That is, the URM females seemed to have a more positive experience (i.e., 68%) than the non-URM females (48%). Overall, the non-URM females had a more negative comments (wishful collaboration 14%, *need for* confidence 11%, testing dislike 11%, and lack of control 16% added to 52%) than the URM females (wishful collaboration 5%, *need for* confidence 14%, testing dislike 8%, and lack of control 5% added to 32%) indicating non-URM and URM female students had a different experience during the group test.

Overall, there was a large amount of beneficial collaboration for both URM and non-URM females during the group test. However, note the beneficial collaboration was a combination of the coded collaboration comments and the coded increase in *MSoBS* comments from their post-test reflections. As seen in Figure 20 and Figure 21 (i.e., the tree maps) the non-URM and URM females had roughly the same frequencies of collaboration but a large difference in the occurrence of *MSoBS*(+). Additionally, there was a larger frequency of negative impacts for the non-URM females. Specifically, the difference in the wishful collaboration and the lack of control themes may help explain the differences seen in the quantitative measure for their sense of belonging to a mathematics community. That is, recall the non-URM females had a significant decrease in their feelings of acceptance and in the trust component. The acceptance change may be explained by their need for more collaboration during the group test. Additionally, the decrease in the trust component for the non-URM females may be explained by the larger number of comments that fell under the lack of control theme. Finally, the majority of the URM female students' comments fell under the beneficial collaboration theme. This finding may help explain the increase in the URM females' feeling of acceptance to the mathematics community. Because the group test performance data also indicated that the group composition type may have impacted performance, I examined the distribution of the occurrence of themes across the different groups in which females were placed (i.e., 4(3)F, 2F-2M, and 3F-1M).

When examining the themes across group type (see Table 15, note the table includes the total counts for all 15 females) we can see that each group had approximately the same frequency of occurrence for the beneficial collaboration. Specifically, both the all-female group and the 2F-2M group had 44% and the 3F-1M group had 49% percent of occurrence for the beneficial collaboration. However, there is a difference when the positive themes are combined for each group type (i.e., beneficial collaboration and expressed confidence). The all-female group had only 50% of their comments indicate a positive impact and the 2F-2M group had 51% indicating positive impact. The 3F-1M groups had 66% of their comments indicating a positive impact on their sense of belonging. This is a 15% difference for the women in the 3F-1M groups to the other group compositions. However, recall the 3F-1M groups had five URM females (i.e., two in Group 7 and three in Group 9) and only one non-URM female. This means, I cannot conclude if the positive experience was because of the group composition or the number of URM females that were placed into the 3F-1M groups.

Table 15.

Overarching Themes from Post-Test Reflections Regarding the Experience of Group Testing on Female Sense of Belonging by Group Composition.

	Collaboration		Confidence		Testing	Control
	Beneficial[a]	Wishful[b]	Expressed[a]	Need for[b]	Testing Dislike[b]	Lack of Control[b]
All Females (N=15)	47% (93)	9% (19)	10% (21)	12% (24)	9% (19)	11% (22)
4(3)F Group (N=3)	44% (15)	6% (2)	6% (2)	12% (4)	9% (3)	23% (8)
3F-1M Groups (N=6)	49% (25)	10% (5)	17% (9)	12% (6)	8% (4)	4% (2)
2F-2M Groups (N=6)	44% (47)	15% (16)	7% (8)	16% (17)	6% (7)	12% (13)

[a] Themes or subthemes that had a positive impact on their sense of belonging.
[b] Themes or subthemes that had a negative impact on their sense of belonging.

The most common occurrence was in the beneficial collaboration when examined by race (Table 14) and group composition type (Table 15). This means each student indicated they did feel their group collaborated well and communicated their ideas with one another. This may imply that collaboration and increased student-to-student communication may not be as impactful for all female students’ sense of belonging to a mathematics community (i.e., their group). That is, notice the biggest difference in both of the tables above are in the wishful collaboration and lack of control themes. Additionally, the females in the 2F-2M group had the largest count (i.e., number of comments) for beneficial collaboration, but also the largest count for wishful collaboration, need for confidence, testing dislike, and lack of control. Additionally, the all-female group had the largest frequency of occurrence for lack of control. Overall, it may be that the group composition has more of an effect on the negative impact of the social interaction during a group test than it may have on the positive impact of female students' sense of belonging. However, because the number of URM and non-URM females were disproportionate in each group composition type (i.e., more than half of the non-URM females were in 2F-2M groups and half of the URM females were in 3F-1M groups) this conclusion cannot be made at this time.

Summary Sense of Belonging

The difference in the beneficial collaboration theme with the non-URM having 38% and URM having 56% may help to explain the difference in the change in their affect and acceptance measure seen in the previous quantitative results. That is, the URM females had less decreases in their feelings of being a heard and valued member of their community (i.e., quantitative measured *MSoBS*) and noted more beneficial collaboration in their groups. Whereas the non-URM females had a more decrease in their feelings of affect and less beneficial collaboration in their groups. Additionally, the large number of occurrences of testing dislike (11%) and lack of control (16%) for the non-URM women may explain the decline in the trust component of their quantitative *MSoBS* results. The decrease in the non-URM female students' sense of belonging may also help explain their decrease in performance for the second group test. These women may have not felt as if they were a listened to valued member of their group, and hence their performance during the group test decreased. Additionally, the women in the 3F-1M indicated the group test may have had more of a positive impact due to their feelings of confidence and the beneficial collaboration than the women in the other group types. However, recall the groups were purposely created to provide the students with working environments to help increase their social interactions. Because there was this difference in both performance and sense of belonging for the non-URM and URM women, and the majority of each were placed in differently composed groups, it is important to understand how the group members interacted. These interactions may help explain the differences seen in the analysis thus far.

Social Interactions within Groups

The previous results have already shown a difference in test performance between the non-URM and URM females (i.e., all of the non-URM females scored lower on the second group test than on the second individual test, whereas all the URM females scored higher on the group test). There was also a difference in the change, or lack thereof, in the female students’ sense of belonging for both categories of females. Additionally, the qualitative results showed the non-URM females expressed a need for more collaboration and confidence whereas the URM females expressed more beneficial collaboration and increased confidence. The majority of the URM and non-URM females were in differently composed groups; therefore, the social interactions that occurred during the group test are important to understand the possible association between the group’s communication patterns and the female students' sense of belonging.

To understand the possible influence the social interactions had on the female students' feelings of belonging, I examined the social interactions among the members of three differently composed groups during two group tests. Transcripts for each group were coded for individual student utterances (i.e., talk turns) as a response to the student who spoke before them. Each interaction was coded on three dimensions as developed by Chiu (2000) when identifying social and individual actions of the group problem-solving process (i.e., evaluation of previous actions, knowledge content, and invitational form, see Chapter 3 for more details).

The three research groups observed during the tests contained a different number of women and men in the groups. Recall from Chapter 2 a sense of belonging can be impacted by a presence of stereotype threat, and in situations where a woman's mathematical skills are exposed (i.e., testing) they may have to endure the extra stress of being stereotyped. To investigate the possibility that this stereotype threat was a contributing factor to the female students' sense of belonging, I examined groups that contained women with zero, one, or two men in a group (recall that no group consisted of three men and one woman). See Chapter 3 for more information about the three research groups, the coding process for the group tests, and sense of belonging measures. I provide the following table as a summary of the participants in each research group as a guide when reading this section. Specifically, Table 16 provides a breakdown for each of the research groups by participant name, gender, URM status, group and individual test scores, and pre- and post-*MSoBS* scores.

Table 16.

Summary of the Research Groups Demographics, MSoBS scores, and Test Scores.

Participant	Gender	URM (Y/N)	Pre-*MSoBS*	Post-*MSoBS*	GT_1	IndT_1	GT_2	IndT_2
2F-2M Research Group								
Isaac	M	N	7.07	6.60	73%	73%	59%	74%*
Carl	M	N	3.90	5.13	73%	80%*	59%	54.5%
Annie	F	Y	6.10	6.53	73%	49%	59%	58.7%
Sophie	F	N	7.07	7.00	73%	74%*	59%	67.4%*
3F-1M Research Group								
Kathy	F	Y	3.83	4.43	85%	60%	100%	63%
Mary	F	Y	5.07	5.53	85%	79%	100%	67%
Dorothy	F	Y	3.67	3.30	85%	71%	100%	66.3%
Kurt	M	Y	4.90	5.17	85%	74%	100%	77.2%
4(3)F Research Group								
Maryam	F	N	4.73	5.97	61%	71%*	63%	86.5%*
Georgia	F	Y	7.07	5.83	61%	49%	63%	43.5%
Pratibha	F	N	6.23	5.50	61%	32%	63%	82%*

*Indicates students who scored higher on the individual test than on the group test.

Before I describe the interactions for each group, I provide the reader with an overview of each group to gain a general idea of how the three groups talked through the two group tests. See Figure 22 and Figure 23 below for the relative frequency of on-task talk turns taken by each group member in each of the three groups. During the first group test the 3F1M group had the largest amount of on-task talk turns with 1271, the all female, 4(3)F, group had the second largest amount with 700 on-task talk turns, and the 2F2M group had the smallest with 675 on-task talk turns. Isaac spoke the most in the 2F2M group with 43% of the talk turns coming from him, with Carl speaking the second most frequently with 27% of the talk turns. The females in the 2F2M groups spoke less frequently during the first test with Sophie having 20% of the talk turns and Annie having 10% of the talk turns. In the 3F1M group, Mary spoke the most with 34% of the talk turns coming from her, and Kurt spoke the second most with 32% of the talk turns. Kathy had 20% of the total talk turns and Dorothy had 14%. In the 4(3)F group, Georgia spoke the most with 39% of the total talk turns coming from her, and Maryam spoke the second most with 30% of the talk turns. Pratibha had 22% of the total talk turns and Rin had only 9% of the talk turns for the first group test.

During the second group test the 3F1M group again had the largest amount of on-task talk turns with 1258 talk turns, though their total talk turns slightly decreased from the first group test. The all-female 4(3)F group had a slight increase in their amount with 826 on-task talk turns, and the 2F2M group had a large increase in talk with 1249 on-task talk turns. Similar to the first group test, Isaac spoke the most in the 2F2M group with 40% of the talk turns coming from him. Sophie, however, now spoke the second most frequent with 27% of the talk turns. Carl spoke less frequently during the second test with 20% of the talk turns. Annie increased her talk turns during the second test with 13% of the talk turns. In the 3F1M group, Kurt now spoke the most with 34% of the talk turns coming from him, and Mary spoke the second most with 31% of the talk turns. Kathy had a slight increase in her talk turns with 24% of the total talk turns and Dorothy slightly decreased with 11% of the total talk turns. The 4(3)F group talk turns were similar from the first group test even though the group lost the member Rin. During the second group test, again Georgia spoke the most with 40% of the total talk turns coming from her, and Maryam spoke the second most with 38% of the talk turns. Similar to test one, Pratibha had 22% of the total talk turns.

Figure 22.

Relative Frequency of On-task Talk Turns by Participant per Group for Group Test 1.

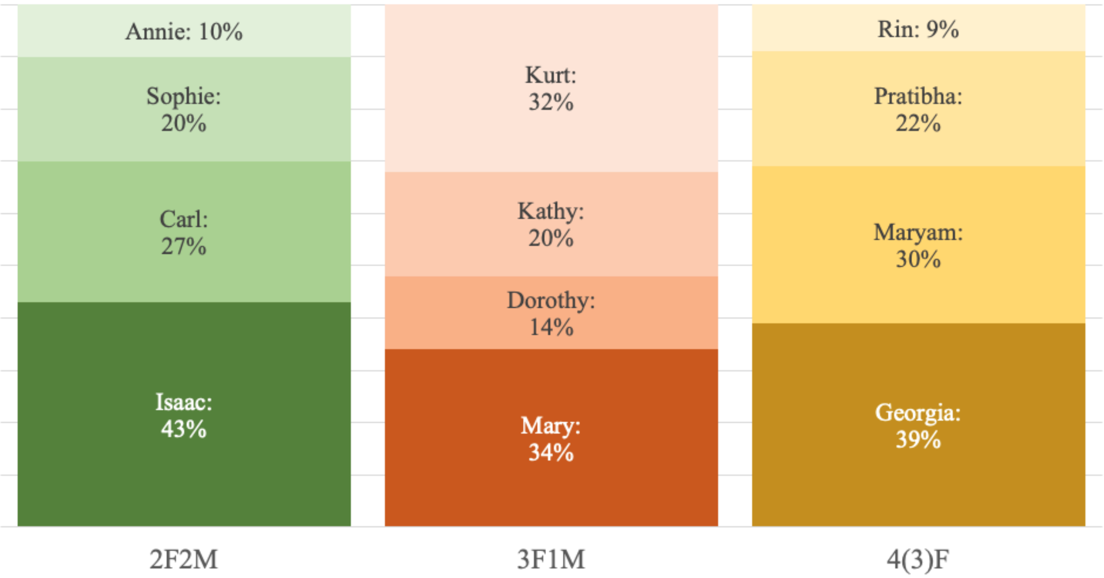
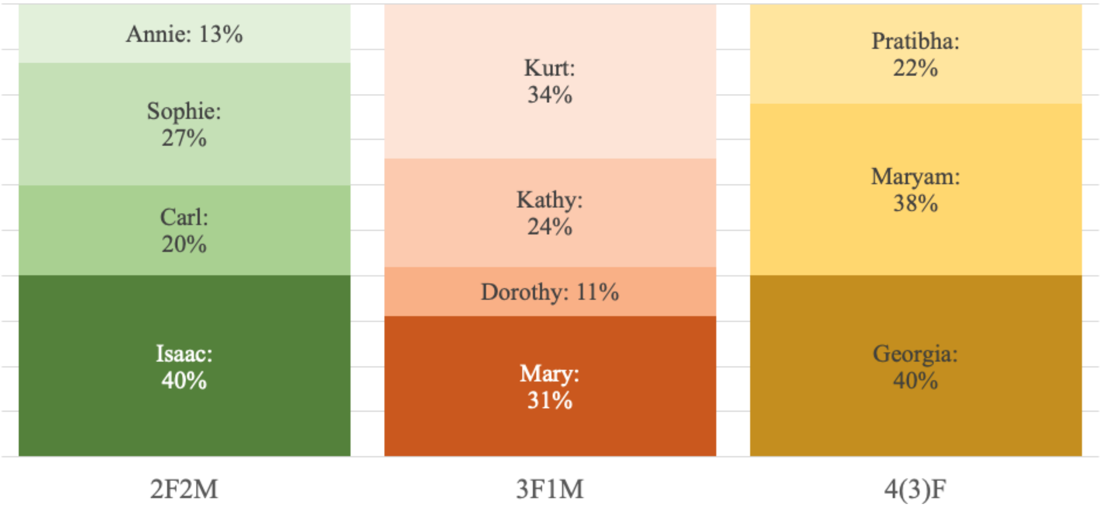


Figure 23.

Relative Frequency of On-task Talk Turns by Participant per Group for Group Test 2.



Although knowing who spoke during the test is helpful, it is also helpful to understand *how* the students spoke to one another. Overall, using the Group Problem

Solving Framework to understand how the students formed their mathematical answers (see Chapter 3), seven themes of talk during the group tests emerged. These seven themes are: (1) supportive talk containing mathematical content, (2) supportive talk without mathematical content, (3) critiques containing mathematical content, (4) critiques without mathematical content, (5) unresponsive statements, (6) commands, and (7) questions. The two figures below display the distribution of the themes per group member for each of the three research groups for both Group Test 1(Figure 24) and Group Test 2 (Figure 25).

Figure 24.

Frequency of Student Talk Themes per Participant for Group Test 1.

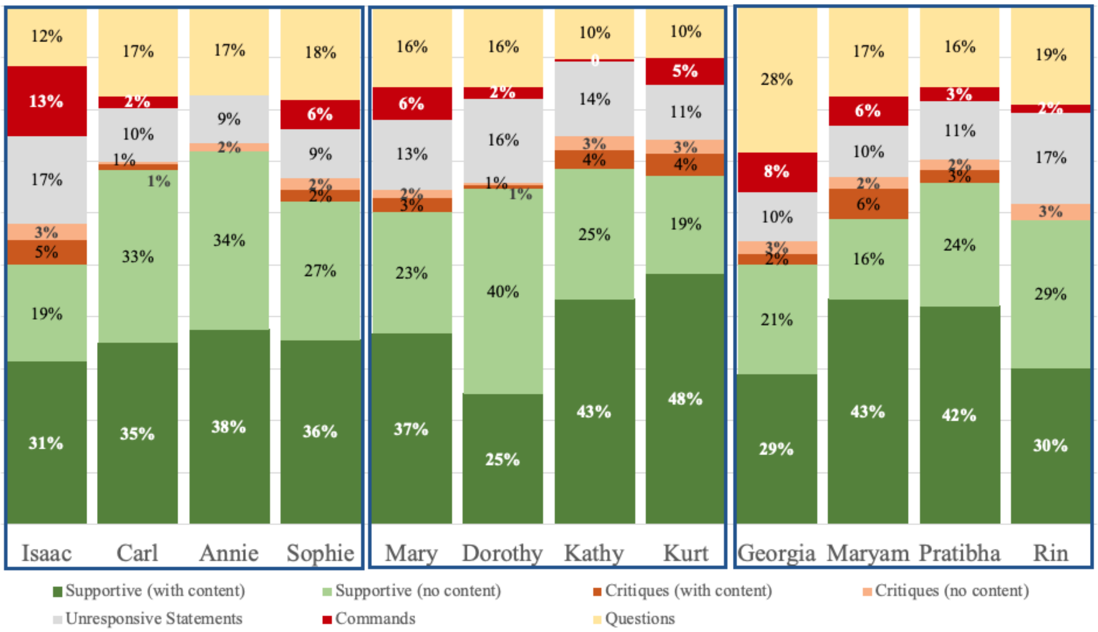
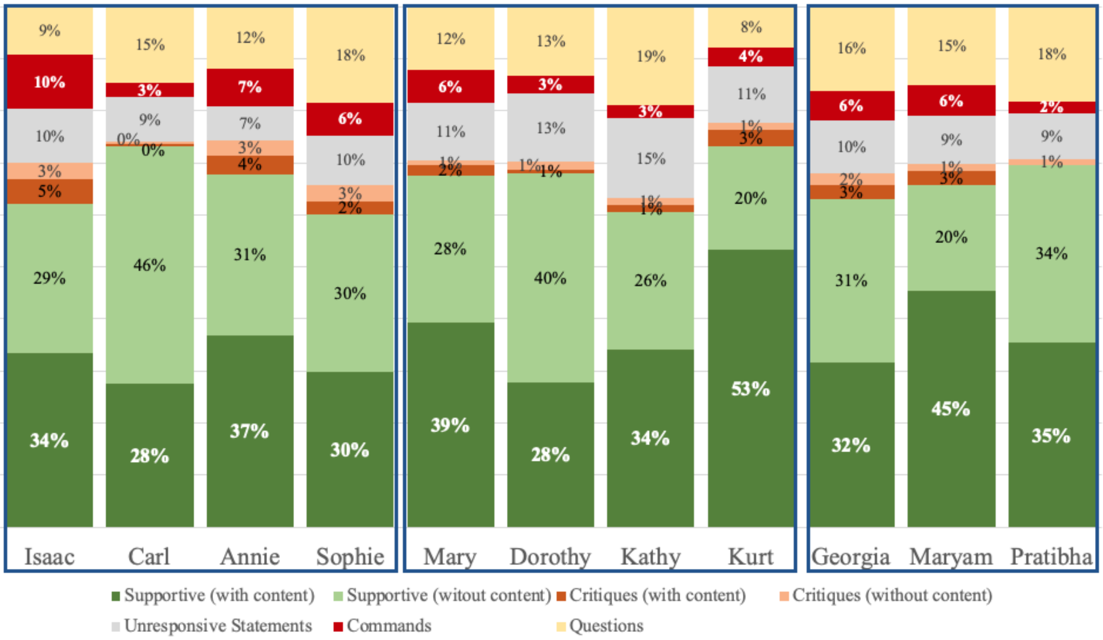


Figure 25.

Frequency of Student Talk Themes per Participant for Group Test 2.



As seen in Figure 24 and Figure 25, each of the three research groups did have some similar interactions, such as the majority of talk consisted of supportive statements shared between the students, but ultimately the three groups had different overall social interactions when answering the problems on the group tests. Because the study is concerned about how the social interactions possibly influenced the female students' sense of belonging, it is important to examine how the students spoke in response to one another, and not the overall compiled interactions of the group members. Therefore, in the following subsections, I describe the social interactions for each group during both the first and second group tests along the three dimensions of the Group Problem Solving Framework (see Chapter 2). Additionally, I include the performance of each group (i.e., grade received on the tests), and how the female students felt regarding their experience (i.e., comments from their post-test reflections or interviews).

2F-2M Research Group Social Interactions

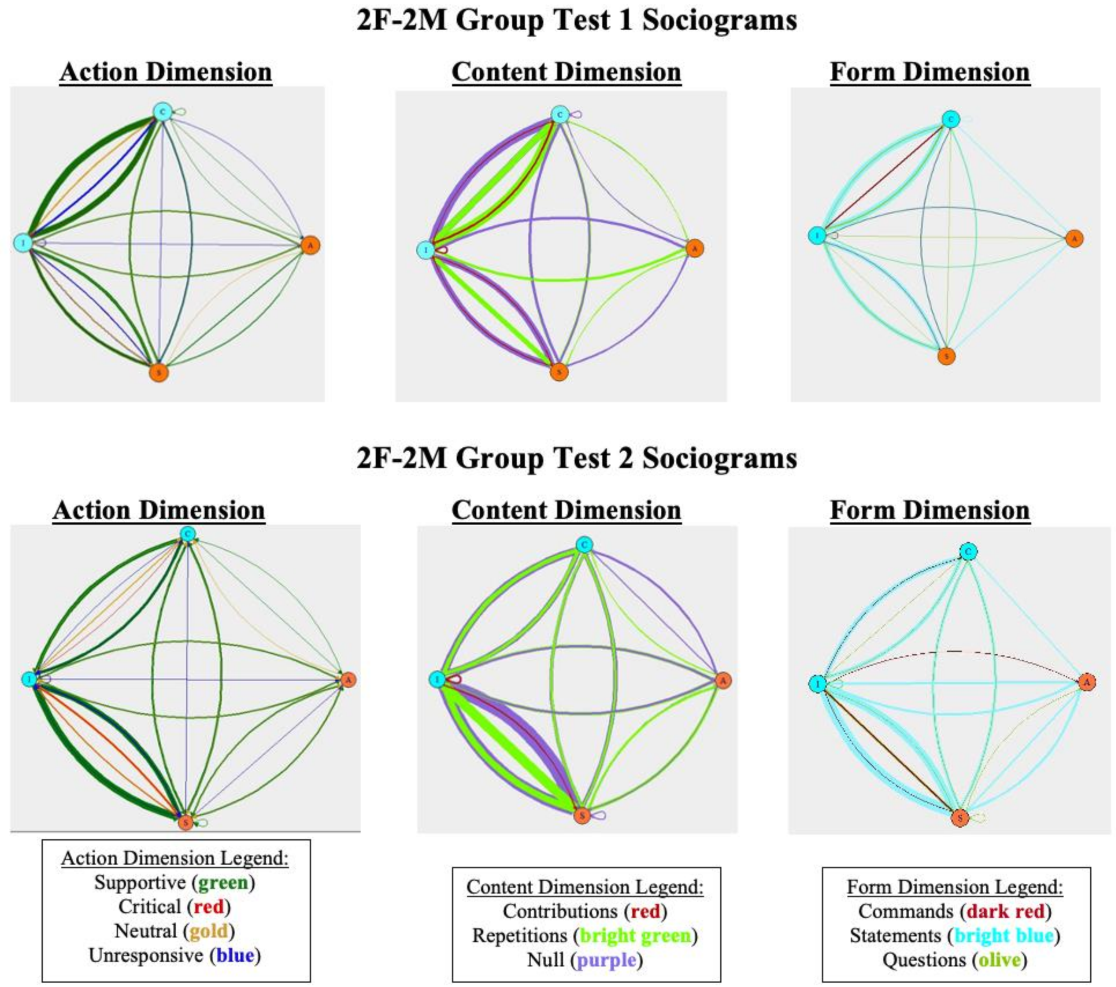
This group contained two non-URM males (Isaac and Carl), one non-URM female (Sophie), and one URM female (Annie). At the beginning of the course, Sophie and Annie had a high sense of belonging to mathematics (i.e., as measured by the *MSoBS* survey), and neither of their scores changed drastically after engaging in the two group tests. Specifically, Sophie had a slight decrease from her pre-*MSoBS* composite score of 7.07 to her post-*MSoBS* composite of 7.00. Annie had a slight increase from her pre-*MSoBS* composite score of 6.10 to her post-*MSoBS* composite of 6.53. Overall, because there was little change to Sophie's *MSoBS* composite score and a slight change in Annie's *MSoBS* composite score it may be speculated that the interactions during the group test did not influence a change to their sense of belonging to a mathematics community as quantitatively measured.

Interactions during Group Test 1

The group earned 73% of the possible points on the first group test. After the group test the students took an individual test in which Isaac received a 73%, Carl received an 80%, Sophie received a 74%, and Annie received a 49%. Annie, a URM female, was the only student in this group to receive a higher score on the first group test than on her individual test. The majority of the mathematical talk occurred between the two men in the group during the first test. The overall interactions during the first test along the three dimensions of the group problem solving framework (i.e., Action Dimension, Content Dimension, and Form Dimension) are displayed in Figure 26. The teal nodes in the sociograms represent the two males in the group (I node is for Isaac and C node is for Carl), the orange nodes represent the two females in the group (S node is for Sophie and A node is for Annie). The majority of the talk during the first test occurred between Isaac and Carl. This can be seen by the thickness in the lines along each dimension, particularly in the Action Dimension, the thickest lines are the green lines between the I and C nodes. This means the most frequent type of talk that occurred during the test was supportive talk between Isaac and Carl. The images are provided on the following page. I describe each sociogram in the paragraphs after the figure.

Figure 26.

Sociograms for the 2F-2M Group's Social Interactions during Group Test 1 and Group Test 2.



When considering the mathematical content spoken in the second image, the majority of the contributions (i.e., initial new ideas, red lines) stem from Isaac, Carl, and Sophie. Contributions are initial mathematical ideas that are brought up in conversation. Ideally, if one mathematical idea is offered by a member of the group and then the group discusses the idea there would be less contributions (i.e., red lines in the Content Dimension) and more repetitions (i.e., bright green lines in the Content Dimension).

However, in the Content Dimension for Group Test 1 (the second image) many new ideas were offered from Isaac to Carl, Carl to Isaac, Isaac to Sophie, Sophie to Isaac, and Isaac in response to himself. Notice as well in the figure there were not as many new ideas offered in response to Annie. This can be seen by the lack of any red line connecting a node to the A node in the Content Dimension for Group Test 1. Also, note in the same sociogram there is a large number of purple lines connecting all of the nodes. The purple indicates there was a large number of talk turns that did not contain any mathematical content. This type of talk would include simple agreements between group members such as “yeah that sounds good to me” or “yeah that’s what I was thinking.” This type of talk (i.e., simple agreements) was common between Carl and Isaac illustrating how Carl would often agree to most anything Isaac stated without adding any type of mathematical statement in his response. The simple agreements type of talk also occurred most often between Isaac and Sophie, indicating the talk between these two group members (i.e., Isaac and Sophie) as well often did not contain any mathematical ideas.

Last, considering the Form Dimension, there were a number of commands given during this test. Most of the commands were delivered by Isaac and can be seen in the third image by the red lines from Isaac to each group member. There were also a number of commands given from Sophie to Carl. Finally, considering all three images overall, the majority of the talk was shared between the two male participants, Isaac and Carl, then between Isaac and Sophie. Annie has the least number of lines indicating she spoke the least during the test and was spoken to the least during the first group test. Overall, the mathematical conservation for the first group test was between the two male participants (i.e., Isaac and Carl). The second most common conversation was between Isaac and

Sophie but contained less mathematical talk (i.e., more purple lines in the second image than green lines). Finally, Annie was the least spoken to directly which is evident from the lack of lines between Annie and any other group member. Examining Sophie's and Annie's post-test reflections helps us understand how these interactions impacted the sense of belonging for these two female students.

The interactions from the first group test had a negative effect on both Annie and Sophie as evidenced by their post-test reflections. However, Annie was more direct in her responses on how the interactions made her feel than Sophie. This may be because Sophie was more involved in the conversation although more as a supportive group member rather than a contributor to the mathematical ideas. For example, when both women were asked if they felt they communicated their ideas during the test they stated:

> *Annie: No, because I do not think that I connected the idea of what we were learning to the problems very well. Also, even when I did speak or try to contribute my ideas to the group, I felt like they just assumed I was wrong and disregarded what I said.*
>
> *Sophie: I did if I understood the question and felt like my input would benefit the group.*

Annie's comment shows not only did she feel the group did not listen to her, but that she did not have the self-confidence in her mathematical knowledge (i.e., not connecting her ideas to the problems) and her perception that her group members just assumed she was wrong when she did speak. Additionally, Sophie's response to the same question showed her lack of confidence in herself and her mathematical understanding (i.e., only speaking if she felt she understood the question or if it would benefit the group). Therefore, both women did not seem to have confidence in their mathematical knowledge during the first group test.

This lowered confidence of the women could indicate the presence of stereotype threat in the group, especially due to the majority of the conversation occurring between Isaac and Carl. In his interview Isaac noted how Carl was more likely to talk to him than to the women in the group. Specifically, Isaac stated:

> *[…] I feel like, um, Carl though would do a lot more talking to me than he would to like Sophie and Annie. Like even, like about like math stuff. Like I don't know if it's because, because he knows me more like, um, because we have had two classes together and like I there, there is also this idea that like Carl and Annie have of me or that I'm good at math, and that's not like my ego talking. Like you can hear them sometimes and they'll make jokes but like it's like a serious like. Because in precalc one [and] in precalc two I was like really good, and I was like able to get through things like really easily, and I was able to explain a bunch of different concepts to them. Like Carl would call me like an SPT in precalc two even though I was his classmate.*

This comment shows how Isaac noticed that Carl had positioned him (Isaac) as the intellectual authority of the group. Although he does not explicitly state that Carl thinks he (Isaac) is better at mathematics than the women in the group, he alludes to this sentiment in the excerpt above. Both Sophie and Annie echoed Carl's sentiment that Isaac was the intellectual authority of the group. During the pre-test conversation for the second group test when the recorders were on Sophie stated "Isaac *is* the group" as Carl compared Isaac to LeBron James.

Additionally, Annie had several comments that fell under the *need for* collaboration as well as a need for confidence, which directly affected her feeling of acceptance by the group. She stated "even if my answers were wrong, I would have liked it if someone could have told me that I was wrong and explained why." This comment illuminates her need for feedback and discussion during the test. She did attempt to offer ideas and engage the group in discussion during the test in which her ideas were never thoroughly discussed. Consider the following excerpt from the third question on the first test. The problem asked students to explain the topic of average velocity and to connect the topic to limits and derivatives. Sophie was the writer for this question, and Sophie and Annie were discussing how the question was similar to a previous homework problem.

1. *Annie: That was the one [\ that you helped me with, for\]*
2. *Sophie: [\ And it was like, \] it was like the, the velocity is equal to the derivative of that function and then [\\ and then \\]*
3. *Annie: [\\ It asks for \\] the acceleration where you had to take the derivative of the derivative.*
4. *Sophie: Yeah.*
5. *Annie: Do you remember?*
6. *Isaac: [\ that was acceleration \]*
7. *Sophie: [\ the acceleration \] was the [\\ derivative \\] of the derivative*
8. *Isaac: [\\ yeah \\]*
9. *Annie: [\\ But \\] yeah*
10. *Sophie: So, it's the, the velocity is the derivative of the original function.*
11. *Isaac: Well like the*
12. *Sophie: Because that's what it, was it was asking for the, it was like, S or like that right? [\ and it was asking for the \] it was asking for*
    the average velocity of that. I, you could, we could put that in there //
13. *Annie: [\ Yeah, remember, S \]*
14. *Isaac: \\You can put the slope of the line, um, or the slope of the secant line, uh. Um. Fuck dude. The, like the secant line is the, average velocity of the.*
15. *Isaac: Well, wait, what do you say before? What were you saying before?*
16. *Sophie: That it's [\ the derivative part? \] it's the derivative of the, of the original function*
17. *Carl: [\ the derivative of \]*
18. *Isaac: Alright, you could say it's like a derivative of that function like in between to, uh, like, two //*
19. *Carl: Left hand right hand [\ limits? \]*
20. *Isaac: [\ No not \] left hand right hand, but like within an interval*
21. *Carl: Oh alright*
22. *Sophie: [\ the derivative of \]*
23. *Isaac: [\ so yeah \] the slope of the secant line, um, or the derivative.*
24. *Isaac: No, I don't think that's right like writing it like that. I think the slope of this secant line, cross that out too. The slope of the secant line within an interval.*
    4 second pause in the conversation
25. *Isaac: Can you guys like give me something??*
26. *Sophie: I don't know, I don't know*

In the exchange above, Isaac does not acknowledge the current conversation at first and just offers an answer and gets frustrated during the conversation (lines 12 through 14). His offer of "secant line" is correct and Sophie and Annie did not have the correct idea for the answer. However, as noted by Annie's comment, no one in the group explained why her original idea was incorrect and Isaac just continued saying his idea without explaining what he was thinking (lines 23 and 24). Isaac instead just states his answer and instructs Sophie to correct the answer "cross that out too." We can see the lack of collaboration in the group with Isaac placed as the intellectual authority (i.e., his answer was placed on the test without discussion). Additionally, this excerpt illuminates Isaac's frustration in line 25 in which he asks his group to help, though he never acknowledges and discusses their initial idea of the derivative of the position function (lines 1 through 7). This excerpt also helps to explain where Annie's feeling of wishing for collaboration and wanting for someone to tell her why she was wrong happened during the first test.

Another example that helps gain insight into the group's dynamics during the first test happened during the time the group attempted the bonus problem. The bonus question nearly led to a disagreement; Isaac, however, took charge of the problem and represented the solution. For the bonus problem, I gave students a vending machine with different colored cans and prices. Students were asked to determine whether the machine would represent a function or not (Figure 27). When the group first read the problem, Annie was the first to attempt to give an answer and Sophie and Carl discussed her idea. The following conversation occurred:

1. *Isaac: Thoroughly explain your choice, list the domain and range of the function.*
2. *Sophie: We have 10 minutes still. Okay [\ well (inaudible) \]*

3. *Annie: [\ Well we definitely can say that it is, \] you put in money //*
4. *Isaac: \\ Wait we have 10 minutes left, is that [\ what you said? \]*
5. *Annie: [\ and it gives you what you want. \]*
6. *Isaac: Yes.*
7. *Annie: So yeah it is.*
8. *Carl: Yeah, your input value is a currency*
9. *Sophie: Yeah*
10. *Carl: And your output value is a [\ product that \] currency is worth so.*
11. *Sophie: [\ product \]*
12. *Isaac: You must give enough logical information to convince (the grader) of your choice function or not function. So (writing on test) it [\ is a function.\]*
13. *Carl: [\ do they want a literal function? \]*
14. *Annie: I mean //*
15. *Isaac: \\ I don't think it wants us to like write an equation*
16. *Annie: It like, just explain like how*
17. *Isaac: Yeah because like all of these are different inputs and they all give you different outputs so*
18. *Carl: mm-hmm*
19. *Annie: Right*
20. *Isaac: It's a function.*
21. *Carl: As long as none of the you know X values have [\ multiple Y's or something\]*
22. *Sophie: [\ Well actually wait, would those. \] Yeah.*
23. *Isaac: These are all different*
24. *Annie: Yeah*
25. *Sophie: Those would all be X values*
26. *Isaac: Yeah. Cuz these [\ are your inputs, you input 45 yeah \]*
27. *Annie: [\ and you get out with different price \]*
28. *Sophie: Right right right*
29. *Carl: Output is the drink*
30. *Isaac: (writing on test) For each X value*

Figure 27.

Bonus Problem from Group Test 1.

Bonus (for up to 10 points):
Is the following a function? Using mathematical terms explain why the machine is or is not a function. Thoroughly explain your choice, list the domain and range of the function based off the given machine, and create a graph to support your answer. You must give enough logical information to *convince the grader* of your choice (function or non-function). Since you are trying to convince the grader, you must all agree on one approach to explain this problem.

The group started the problem all discussing the question (lines 3 through 29), and Carl, Sophie, and Annie quickly turned to talking off task as Isaac wrote their explanation on the test. While the group was talking off task Isaac had questioned if the answer was supposed to be a discrete function. He decided it was for the group and began writing this on the test. While he was writing it was hard for the other group members to see the paper. Isaac noted in his interview that he is left-handed and that was the reason he held the pen and paper the way he did during the test. He was sitting on the right side of Annie, so she had the least view of the paper when Isaac was the writer. The time was running out and after the off-task talk ended, Annie tried to see what Isaac was writing. The following conversation occurred when this happened:

1. *Annie: What are you writing? The inputs are 45, [\ the inputs in a line? \]*
2. *Carl: [\ There are a set of \]*
3. *Isaac: [\ (holds hand up to Annie) wait wait wait wait wait \] shut up shut up.* Isaac turns to Annie

4. *Isaac: I'm sorry, I'm sorry, I shouldn't have said shut up that was really mean. I had it and then I lost it that was really mean I'm sorry guys*
   Isaac turns to the camera
5. *Annie: That was really mean*
   Annie and Sophie laugh
6. *Isaac: I [\ cuz I'm like really bad my memories trashed \] dude yeah*
7. *Carl: [\ Like you got to be writing when you say sorry \]*
8. *Carl: 5 minutes*
9. *Isaac: (writing on test) So, there are [\ set inputs \] 45, 65, 70, etc. with set outputs however, uh, so you cannot uh input values in between those inputs. 43 cents, 62 cents, 39 cents, etc. there (tosses pen on table) I [\\think we got it guys \\]*

This exchange demonstrates the need for collaboration the two women in the group had alluded to in their post-test reflection. Annie was trying to see how Isaac was representing the group's answer and Isaac got frustrated with her interruption (line 3). However, note Isaac may have benefitted from collaboration as well to help with his memory problem (line 6). That is, if the group was crafting the answer together then the answer would not depend on one person's memory.

Overall, during the first group test the conversation revolved around Isaac with Carl directly responding in agreement with Isaac. Sophie and Isaac had the second most talking exchanges, and Annie was directly spoken to the least during the test (Figure 14). The two women in the group had lowered confidence in themselves and expressed a need for collaboration during the first group test. Annie specifically brought up a need for feedback and a need for feeling heard by her group members. Isaac also noted how Carl spoke more often in response to him than to the women in the group. Because of the dynamics observed during the first test, and as the instructor of the course, I did speak with the group members before the second test in an attempt to help them communicate more as a group. Note, I did not conduct this conversation as the researcher, but instead as the instructor of the course. After the “shut up” comment by Isaac at the end of the first test I felt it was my duty as their instructor to help the group be more inclusive while solving problems and make sure each person was allowed to speak and ask questions. On Sophie’s second post-test reflection, she explained that this conversation helped and there was a positive change in the dynamics during the second group test.

Interactions during Group Test 2

For the second group test, there was both a shift in performance (i.e., all test scores) and the group dynamics from the first test setting. That is, on the second group test, the group earned 59% of the possible points, and for the second individual test Isaac and Sophie received grades higher than the group test score (i.e., 74% and 67.4% respectively) and Carl and Annie received grades slightly lower (i.e., 54.5% and 58.7% respectively). The majority of the mathematical talk occurred between Isaac and Sophie during the second test. In the following subsections I provide details about the group’s interactions during both tests and the impact those interactions had on the female students in the group with regard to their sense of belonging. There is a shift in the group dynamics during the second group test that can be seen in the sociograms in Figure 26.As seen in these sociograms for Group Test 2, the majority of the lines are still connected to the node representing Isaac. However, the majority of the talk during the second test was now between Isaac and Sophie.

As seen in the sociograms for Group Test 2 in Figure 26, the most common type of talk that occurred during the second test was repetitive supportive and null talk between Isaac and Sophie. There was an increase in critical talk as well between Isaac and Sophie. This increase indicates the students correcting each other’s ideas during the test (see Chapter 3). In the Content Dimension there more green lines connecting each node which indicates more mathematical talk to and from each participant during this test. However, when considering the initial mathematical ideas spoken in the second image, the majority of the contributions (i.e., red lines) only come from Isaac. In Form Dimension for Group Test 2, the group members often still spoke in a commanding way and many of these commands were given by Isaac. Recall, a command is a directive given from one student to another such as “write this down” or “just put” in which the student does not suggest a possible answer but instead insists on what is to be written. However, Sophie was also giving commands to Isaac during the second test, this is indicated by the two red lines between the I and S nodes. That is, the red lines between Isaac and Sophie indicates multiple commands were delivered by both participants to each other.

Overall, the mathematical conversation for the second group test was mainly between Isaac and Sophie more than any other participant. The second most common conversation was between Isaac and Carl. Additionally, Sophie and Carl had more exchanges about the mathematics during the second test. Finally, Annie was spoken to more often than she was during the first test and spoke more about the mathematics as indicated by the increase in green lines on the Content Dimension for Group Test 2 from the A node. Even though the talk around Annie is still less than the talk with the other participants, she is now more involved in the mathematical discussion. Because one’s sense of belonging is both fluid and context specific, this experience for Annie was a drastically different experience than the first group test. Therefore, the effect of the interactions during the second group test was now positive for Annie’s sense of belonging to this specific mathematics community with respect to feeling heard and valued by her community (i.e., her group). Specifically, Annie stated in her interview:

> *I feel like [...] it definitely was a huge change for me from the first test to the second one. [...] I knew what I was talking about and like people listened to what I said and agree-, well not like went along with, but like agreed with what I said.*

In her comment, it is evident that Annie felt better about voicing her ideas (i.e., "knew what I was talking about") and felt heard by her community (i.e., "[they] agreed with what I said"). Additionally, the interactions from the second group test had a positive effect on both Annie and Sophie as evidenced by their post-test reflections. Both of the women in this group noted a change in the group's dynamics from the first to the second test. Additionally, both women commented on the increase in discussion during this test:

> *Annie: We all were heard and understood each other this time. Everyone was involved and contributed.*
>
> *Sophie: We bounced off more ideas and felt more open about voicing opinions.*

Both Annie's and Sophie's comments show each woman felt the group communicated more during the second test. Additionally, Annie felt more listened to more during the second group test than the first test, and that there was beneficial collaboration (i.e., "involved and contributed") in the group. Sophie showed an increase in confidence by feeling more open to voicing her opinion (i.e., "felt more open") and also noted the increases in discussion ("bounced off more ideas"). Both women did speak more about the mathematics, and even offered ideas without being asked by their group members. For example, during the first problem in which Isaac was the writer, the two women recalled the problem from a previous one they observed the day before. The problem was asking students to find the vertical and horizontal tangent lines of a given implicit curve.

The following conversation occurred when Isaac began the test:

1. *Isaac: (reading problem 1) how many points on the curve have horizontal tangent lines, show as much work as possible, so Sophie: Remember, yeah, this is what we did yesterday at the um*
2. *Annie: Yeah*
3. *Isaac: We differentiated for Y right in the first one?*
4. *Sophie: mm-hmm*
5. *Annie: yeah you have to use the product rule for X and Y*
6. *Isaac: yeah, yeah, yeah, so it's um, 4 oh, I'll write down my original first. (writing on test) Plus X Y, plus X squared, (mumbles – starts writing the derivative) times Y prime plus um (tapping on table)*
7. *Annie: I'm pretty sure it's x um times*
8. *Isaac: It'd be, yeah,*
9. *Annie: [\ dy over dx plus y, yeah \]*
10. *Sophie: [\ it's x plus y and then yeah \]*
11. *Isaac: Would then*
12. *Sophie: Yeah [\ x plus y \] and then the x would be the*
13. *Isaac: [\ yeah x plus \]*
14. *Isaac: Times Y prime yeah, plus two x equals zero*
15. *Annie: I forgot you're write Y prime and not DY DX*
16. *Isaac: Yeah, dy/dx is too much, you know.*
17. *Isaac: Um, okay so (writing on test) Minus 2 X minus Y (10 sec pause) and then factor out the Y prime (note no one is talking as he's doing the work). Okay, so then I have horizontal tangent lines, so uh when the derivative is zero, I have a horizontal tangent line. So*
18. *Carl: mm-hmm*
19. *Isaac: (writing on test) Zero equals negative 2 X minus Y over 4 Y plus X, and then um*
20. *Sophie: Wouldn't you cross multiple [\ I mean multiply \] by the reciprocal that's what I mean*
21. *Isaac: [\ No it's just \]*
22. *Isaac: Yeah*
23. *Sophie: That's what I meant, and then it would be just the numerator*
24. *Isaac: mm-hmm*
25. *Sophie: And remember we're not actually looking for values*
26. *Isaac: Yeah, cuz then you just plug it back in*
27. *Sophie: mm-hmm*
28. *Isaac: So then*
29. *Annie: You take that and plug it back into the original equation*
30. *Isaac: Yeah*

31. *Sophie: (to Annie) look at you contributing, I feel like a proud mom.*[8]

In this excerpt, both Sophie and Annie offering help while Isaac takes the derivative. Also, notice Sophie's comment to Annie (line 31) about contributing. Annie and Sophie had talked about their concerns of not contributing during the first group test.

Although Sophie and Annie perceived a change in collaboration during the second test, they did not exhibit an increase in confidence when answering problems. The conversation for this problem continued for over 20 minutes. However, there was a slight change in dynamics as the conversation continued for this problem:

1. *Isaac: How many points yeah that's all (flips page-reading test) how many points on a curve have vertical tangent lines, so this is the inflection point right*
2. *Carl: mm-hmm*
3. *Isaac: Um, so I have to take my original derivative which was Y prime equals negative 2 X minus okay (flips page again) wait I thought it was 4X minus Y, or 4Y minus X. So then, take the derivative of this, uh the quotient rule. Oh wait.*
4. *Sophie: Don't you just //*
5. *Isaac: \\ I have to do //*
6. *Sophie: \\ don't you just set the denominator equal to zero?*
7. *Isaac: Do what?*
8. *Annie: Yeah*
9. *Sophie: Cuz that's [\ how you find the vertical asymptotes \]*
10. *Isaac: [\ oh yeah yeah yeah true true \] um so*
11. *Sophie: cuz don't you remember, that's when she was looking for yesterday.*
12. *Isaac: (writing on test) 4Y plus X equals 0, um and I need X values so, or so then I should, (writing) minus 4 Y, so X equals negative 4Y and then I have to plug this back into my original*
13. *Carl: mm-hmm*
14. *Isaac: (writing on test) That's 2Y squared plus Y plus X square, 14 Y squared equals 112 divided by 14, Y squared, so that equals radical 8, plus or minus radical 8*

    12 second pause / silence as Isaac writes on the test
15. *Isaac: So um, well no, when, didn't she have to set that equal in order to find asymptotes that wasn't for (...) Because this is asking for vertical tangent lines*

---

[8] Note this group had a friendly dynamic and joked during class that Isaac and Sophie argued like a married couple and were the mom and dad of their group.

and so if you have a uh a vertical tangent line is equal to zero and that would be the inflection point.

16. *Carl: Yeah cuz that's when there*
17. *Isaac: It's the second derivative*
18. *Carl: You're switching from one to*
19. *Sophie: mm-hmm*
20. *Isaac: Well but that's when the second derivative is equal to zero so I have to do the quotient rule //*
21. *Sophie: \\ mm-hmm //*
22. *Isaac: \\ for this actually*
23. *Sophie: Shit my bad*
24. *Isaac: No, you're good*
25. *Carl: Well, I mean at least we caught it.*

During the continuation of the problem, Sophie had told Isaac in order to find the vertical tangent lines he had to set the denominator of the derivative equal to zero. (lines 6 through 11). Isaac at first agrees, but then decides the problem was looking for the inflection points instead (line 15), Carl immediately agrees with him (line 16), and Sophie apologizes for being incorrect (line 23). If Sophie or Annie had more confidence, they may have been able to argue their point with Isaac. Also, note at no time during the course did students ever take a second derivative of an implicit equation. This could have been one point the women could have made to help convince Isaac their idea was correct. As Isaac began to take the second derivative of the implicit equation, Sophie, Annie, and Carl began talking to each other off task. Notice when Isaac stated that he believed the question was looking for inflection points (line 15), no one in the group questioned his idea (i.e., still placing Isaac as the intellectual authority of the group).

In summary, although the women in the group spoke more during the second test, the conversation and answers for the test still revolved around Isaac's stated mathematical ideas with Carl agreeing. Sophie and Annie did contribute more to the mathematical conversation and felt their contributions were well received by the group as evidenced from their post-test reflections. Both women felt the group collaborated more during the second test and did not recognize areas in which they could have mathematically argued with their group members (i.e., the last example from test two).

Summary of the 2F-2M Research Group Social Interactions

Overall, each group member spoke more often during the second group test than during the first test. The majority of the mathematical conversation switched from between Isaac and Carl in the first test to Isaac and Sophie in the second test. Though Isaac was still positioned as the intellectual authority of the group for both tests. The two women in the group noted an increase in collaboration and feelings of being heard by their group members from the first to the second test. Although both Sophie and Annie indicated they had a more positive experience during the second group test, they did not show evidence of confidence in their mathematical knowledge during the second test.

The presence of the two men in the group may have activated Sophie's and Annie's lowered confidence due to stereotype threat (i.e., men are better at math than women). This was evidenced by Carl agreeing with Isaac even when Isaac was incorrect and Sophie and Annie had already presented correct mathematical ideas to the group. The possible presence of stereotype threat, however, did not seem to lower the perceived positive impact of the beneficial collaboration for the two women in this group, nor did it significantly impact their sense of belonging to a mathematics community. Conversely, the group performance decreased from the first (73%) to the second (59%) test, which may indicate the lowered confidence of the women impacted the group's performance (i.e., rather than their sense of belonging). Recall from the conceptual framework in Chapter 2, the group tests were created to measure participatory learning (i.e., students constructing knowledge as a group). Overall, the 2F-2M group interactions did increase among the group members from the first to the second test, yet this increase alone did not impact the students group performance (i.e., test scores) nor the sense of belonging measure (i.e., *MSoBS* scores) of the women in the group.

3F-1M Research Group Social Interactions

This group contained one URM male (Kurt) and three URM females (Kathy, Mary, and Dorothy). At the beginning of the course, Kathy and Dorothy had a low sense of belonging to mathematics and Mary had a medium sense of belonging to mathematics (i.e., as measured by the *MSoBS* survey). Kathy was the only woman in the group to have a change in her sense of belonging from a low to a medium score. Specifically, Kathy had an increase from pre-*MSoBS* composite score of 3.83 to post-*MSoBS* composite of 4.43. Dorothy had a slight decrease from her pre-*MSoBS* composite score of 3.67 to her post-*MSoBS* composite of 3.30. Mary had a slight increase from her pre-*MSoBS* composite score of 5.07 to her post-*MSoBS* composite of 5.53. Because Kathy's *MSoBS* composite score increased from low to medium, the interactions during the group test could have had a positive impact on her sense of belonging to a mathematics community. Since Dorothy's and Mary's scores remained in the same classification (i.e., low and medium), however, any conclusion cannot be time.

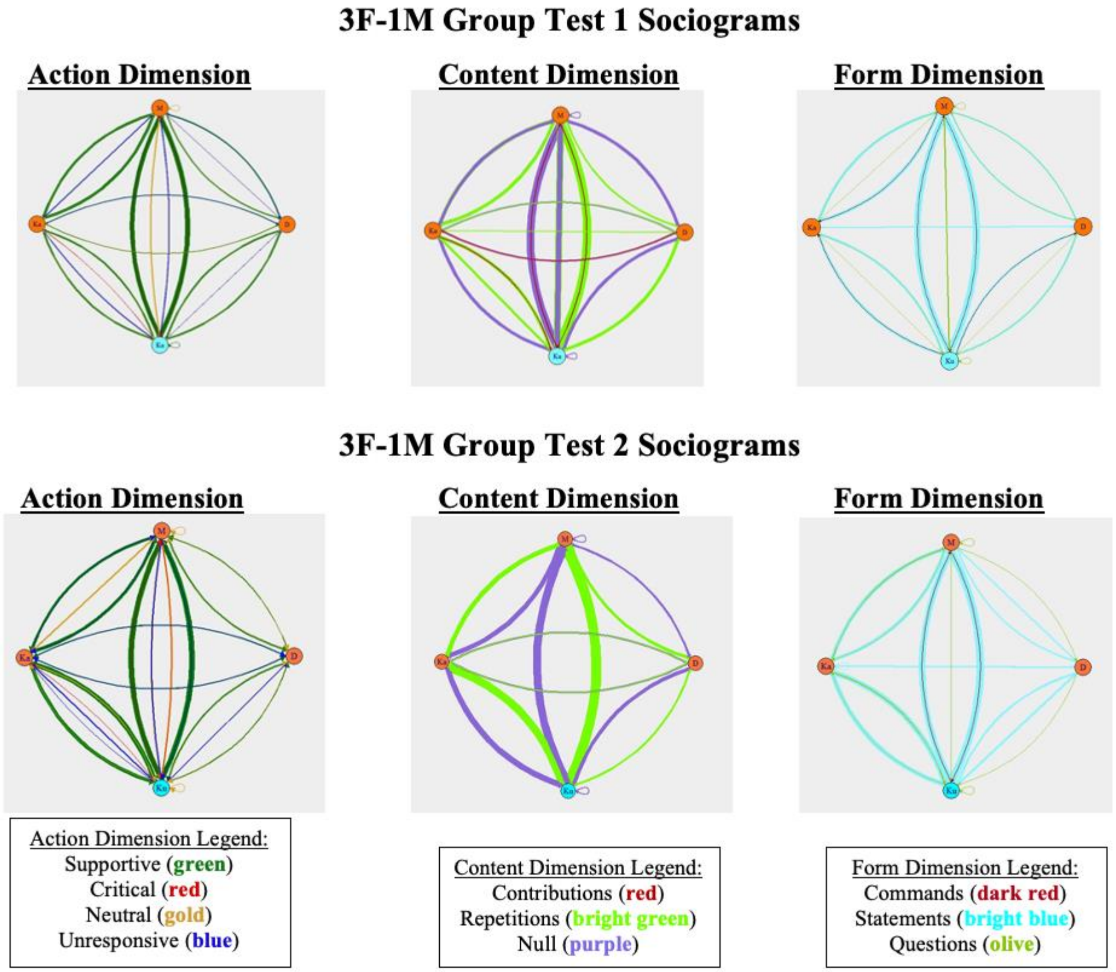
Interactions during Group Test 1

The group earned an 85% on the first group test. On the corresponding individual test Kurt received a 74%, Kathy received a 60%, Mary received a 79%, and Dorothy received a 71%. Every student in the group scored higher on the first group test than on the first individual test. The majority of the conversations during the first test happened between Mary and Kurt. Each member of the group, however, was involved in all mathematical conversations that occurred during the first group test.

The interactions during the first test along the three dimensions of the group problem solving framework (i.e., Action Dimension, Content Dimension, and Form Dimension) are displayed in Figure 28 he teal nodes in the sociograms represent the male student in the group (Ku node is for Kurt), the orange nodes represent the three females in the group (Ka node is for Kathy, M node is for Mary, and D node is for Dorothy). The majority of the talk during the first test occurred between Mary and Kurt. This can be seen by the thickness in the lines along each dimension between the M and Ku nodes. Notice, however, the thickness of the lines between Mary and Kurt are not that much thicker than the lines between each other participant. That is, in each picture there is clearly a colorful circle connecting each of the four nodes. This means each of the four students was engaged in the conversation for relatively the same number of talk turns during the first test.

Figure 28.

Sociograms for the 3F-1M Group's Social Interactions during Group Test 1 and Group Test 2.



Examining he sociograms, particularly in the Action Dimension (i.e., the first image), there are green lines going from each participant with the thickest between Mary and Kurt, then between Mary and Kathy. This means the most frequent type of talk that occurred during the test was supportive talk between Mary, Kurt, and Kathy. Also, consider the Content Dimension (i.e., the second image) there are thick bright green lines connecting each of the nodes indicating all group members participated in the mathematical discussion during the test. Additionally, notice the red lines in the Content Dimension for Group Test 1 stemming from each node to another, all of similar thickness. This indicates each student offered an initial idea (i.e., contribution) roughly the same number of times as their peers during the test. Finally, in the Form Dimension (i.e., third image) there are a number of each statement and questions said by each member of the group. The only commands given in the group stemmed from Mary and Kurt and each group member was spoken to in a commanding way relatively the same number of times (i.e., the red lines in the image are the same thickness). This image is what I would consider equitable distribution of talk turns (see Chapter 3 Purpose 2). Although it appears Mary and Kurt talk more to each other during the test as a whole, there is no member who is left out of the conversation (as previously seen in the 2F-2M group). Examining Kathy's, Mary's, and Dorothy's post-test reflections helps uncover how these interactions impacted the sense of belonging for these three female students.

The interactions from the first group test had a different effect on each of the women in the group. This was particularly evidenced by their individual replies to the question "Do you feel you communicated your mathematical ideas during the group test?" Kathy stated "Kinda sorta, but not really. I feel as though the way I communicated was not clear" which indicates she did not have confidence in herself or the way she was communicating her mathematical ideas (i.e., was not clear) during the test. Conversely, Dorothy wrote "I do feel I communicated my mathematical ideas with my group during the exam as I felt more confident in my understanding [of] the material" in which she directly states she did feel more confident in expressing her mathematical knowledge. Finally, Mary said "Yes, I feel like when I had an idea or concept or even a question my group listened to what I had to say and responded with relevant and helpful comments" in which she notes how she felt the group listened to her even when she was questioning a peer's mathematical ideas.

Although each of the women felt differently with respect to how they communicated their ideas, they all mentioned the beneficial collaboration they experienced in the group. Consider their replies when asked to provide one positive comment about their experience:

> *Kathy: The group exam seemed less strenuous because we could bounce ideas off each other.*
>
> *Mary: I like that when I had questions or don't know exactly what we were looking at my group took the time to explain a concept to me. I think it helped us move forward together and answer future questions more effectively.*
>
> *Dorothy: We have learn[ed] how to communicate more effectively since the first week of class so allowing each person to state thoughts and allowing someone to talk when they do not understand or disagree with something.*

Kathy felt the discussions during the group test made the test seem easier or more manageable (i.e., "less strenuous"). Mary felt comfortable enough with her group members that she felt she was able to voice all of her ideas and questions. Dorothy adds to Mary's comment with the idea that it took the group some time to learn how to communicate. Recall students were placed into their permanent group at the end of the first week of class and the first group test was at the end of the third week of class. Mary further explained in her interview how her group members all respected and listened to one another. Specially, I asked Mary if she could explain how this type of interaction started in their group and she stated:

> *How do we listen to each other? Maybe like, we all do this thing where we like internalize what the person just said and then like come back at it because we all have attitudes like we're all pretty sassy we're not like mean attitudes but we have*

> *like sass if that makes sense. So, if someone says like okay well if we know two plus two is four then blah blah blah I can tell like everybody in the group is thinking does two plus two equal four how does it relate to this and then how can I reply. So, like we're always like a little bit of like we have like a couple seconds before we answer our questions.*

She brings in this idea of patience her group members have with allowing each person to internalize each other's responses. This type of interaction (i.e., allowing a person to contemplate and question another's idea) was apparent in each problem for this group during the test, and was most noticeable when the group disagreed on a problem.

During the first test, the group only disagreed on the answer for the second problem. This problem asked students if the limit existed for a given graph and to rationalize their response using the epsilon delta definition of limit. The three women all agreed the limit did exist and Kurt did not agree. Consider the following excerpt from the group discussing this problem:

1. *Mary: So, ah, according to the epsilon-delta definition of a limit, um, oh fuck what did I say? F of, F of A is within reach of the limit approaching [\ the left and the limit \] approaching from the right*
2. *Kathy: [\ the left and the right \]*
3. *Mary: That make sense?*
4. *Dorothy: [\ Yeah \]*
5. *Kathy: [\ Yeah \]*
6. *Mary: Kurt?*
7. *Kurt: I don't know if I, I mean if I disagree I could write my own thing //*
8. *Mary: \\ I know, but what do you think about what I said like why do you [\ disagree \]*
9. *Kurt: [\ I just \] don't think it exists in generalize that's*
10. *Mary: But epsilon, the epsilon-delta definition says it has to be within like [\ within \]*
11. *Kurt: [\ Yeah \] I understand that, can we [\ just get in a middle ground \] cause like (chuckles)*
12. *Kathy: [\ Yeah, that's according the definition \]*
13. *Kathy: No, because //*
14. *Kurt: \\ No, no no [\ not of us, not of us \], I meant like like like for like the disagreeing like //*
15. *Kathy: [\ that's the, that's, okay \]*

16. *Mary: \\ So wait, like, why do you disagree though?*
17. *Kurt: I just don't think it exists*
18. *Mary: Because you think coming from the left //*
19. *Kurt: \\ Yeah //*
20. *Mary: \\ But this does exist. So, if you think about this way like all these are [\ leading to \]*
21. *Kurt: [\ this this \] does exist. But like this [\ value doesn't \]*
22. *Kathy: [\ You're also thinking /] I wou-, say, you're thinking in terms of continuity. We not, we not concerned about [\ continuity \]*

As seen in the excerpt above, because Kurt does not agree with the group's answer (line 7), instead of letting him just answer himself Mary asks him what he is thinking (line 16). Additionally, Kathy provides a mathematical reason for what she believes Kurt is thinking (line 22). The group discussed the problem for a few more minutes until moving on to the third problem. They had decided to come back to problem two at the end of the test and Kurt ended the following discussion by saying "*[I] probably won't change my opinion just cuz it's my, I mean, I'm stubborn though so."* By the end of the test, Kurt decided to sign the agreement line on the test for the answer provided by the women (i.e., that the limit did exist). This interaction is very different from what was previously seen in the 2F-2M group. The disagreement revolved around asking about the other students' understanding, instead of just providing an idea without explaining why one person thought that the idea was correct.

Overall, during the first group test the mathematical conversation included each group member's questions and ideas. If one member was unsure of an answer provided by another member, the group explained the mathematics to them until everyone understood or agreed with the answer. Mary and Kurt appeared to be the dominant voices during the first test, but all students felt comfortable voicing their ideas and engaged in the mathematical conversation. Even when the group did not agree on an answer, each member was encouraged to explain what they were thinking. All of the women noted the collaboration was beneficial in their post-test reflections, however the impact of the collaboration was different for each of the women. Kathy exhibited a need for confidence in her own mathematical understanding where Dorothy expressed her confidence. Finally, Mary felt comfortable with her group members and felt she could learn from them by asking questions. The group's performance increased on the second test and their dynamic had only a slight change from the first to the second group test.

Interactions during Group Test 2

For the second group test, this group performed similarly as they did during the first test with a few notable changes. The group earned a 100% on the second group test, and were the only group in the class to earn a perfect score. Again, after the group test the students took an individual test in which Kurt received a 77.2%, Kathy received an 63%, Mary received a 67%, and Dorothy received a 66.3% on the individual test. Similar to the first pair of tests, every student in the group scored higher on the second group test than on the individual test. The majority of the conversations for the second test happened among Kathy, Mary, and Kurt almost equally. While Dorothy spoke less during the second group test, she was still included in all mathematical conversations. Notice the slight shift in the group dynamics from Group Test 1 sociograms to Group Test 2 sociograms in Figure 28

As seen in Figure 28, the talk increased between Mary and Kathy, and Kurt and Kathy. While this is similar to the group's interactions during the first test, there is an increase in the thickness of the green lines connecting the M and Ka nodes and the Ku and Ka nodes. This would appear that Kathy was more involved in the conversation during the second test than she was during the first test. Also, notice the lines connected to the node representing Dorothy are lighter in Figure 28 (i.e., 3F-1M Group Test 1 interactions). This would indicate Dorothy was less engaged in the second group test. Recall, however, the lines are representing the most frequent occurrence of interactions for talk turns that happened at least five times. Also, there is still a clear circle connecting each node indicating all of the group members were still involved in the conversation. Considering the Content Dimension for Group Test 2, there is an increase in thickness of the purple and green lines. Also, there are no red lines in the second image indicating a decrease in initial ideas (i.e., contributions) offered. These two changes together suggest the group discussed one person's ideas more in the second test instead of offering new mathematical ideas. Finally, in the Form Dimension for Group Test 2, there is also a decrease in the number of commands (i.e., red lines) from the first test. Recall, in the first test both Mary and Kurt had red lines stemming from their nodes to each other group members' nodes. During the second test, the commands given were primarily between Mary and Kurt. That is, the red lines between Mary and Kurt in the Form Dimension for Group Test 2 indicates multiple commands were delivered by both participants to each other.

In their second post-test reflections, all three women in this group wrote about the benefit of the group members' collaboration. Additionally, none of these women provided a negative comment about their experience during the second group test. In her post-test reflection, Kathy stated "*My group was very cooperative and this made the exam seem less stressful.*" In her comment she both notes the collaboration (i.e., very cooperative) and the greater feeling more at ease (i.e., less stressful) than she had during the test. Mary noted that she did not speak as much during the second test but that she felt what she said was listened to more during the second test. Specifically, she stated "*I think my contributions were taken into consideration more seriously than in the last test. I felt like I was heard and part of my team. Even though I was less talkative*." Mary expanded on this comment in her interview explaining that she felt she was questioning everyone too much during the first test. Even though she felt she did speak less, her feeling of being a valued member of the group (i.e., part of the team) increased during the second test. Finally, Dorothy said "*I do feel I communicated [my] mathematical ideas clearly during the group test as when I spoke (or anyone from the group spoke) everyone was willing to listen to others*." In her comment, Dorothy notes that not only herself was listened to during the test but all of the group members were listened to when they spoke.

Although only one person was allowed to be the writer for each problem, each group member in this specific group would talk through all of the problems. To assign a writer for the problem in the group, the group members would look at the problem and give the writer position to the person who felt most comfortable writing the answer. Consider the following exchange from the first problem on the second test:

1. *Kurt: Alright. So who's doing number one?*
2. *Kathy: Number 1, it's the one we did yesterday.*
3. *Mary: I don't feel comfortable with this one.*
4. *Kurt: Okay*
5. *Mary: Like, like I couldn't riddle it all out right now. Anybody?*
6. *Kurt: Yeah, I can do it.*
7. *Dorothy: Yeah.*
8. *Kathy: Okay.*
   Paper passed to Kurt – becomes writer for Problem 1.
9. *Kurt: Alright, how many points, so we have that curve.*
10. *Kathy: How many, oh, I was gonna say don't worry about how many points it's worth.*
11. *Mary: (reading problem) [\ how many points on the curve \]*

12. *Kurt: [\ So we ha::::ve t:::o do \] Do implicit!*
13. *Kathy: Yep.*
14. *Mary: For something to have a horizontal tangent line*
15. *Kurt: I'm gonna scoot up (Kurt moves his desk in closer to the group)*
16. *Kathy: Wait so //*
17. *Mary: \\ It's where the critical points are*
18. *Kathy: So, brain not working*
19. *Mary: mm-hmm*
20. *Kathy: So when you're doing, remember in Precalc 2 when you're doing asym-. No, Precalc 1 when we were doing asymptotes. Which one, verticals with on the bottom right?*
21. *Kurt: If I set the bottom equal to zero*
22. *Dorothy: Yeah*
23. *Kathy: Great.*
24. *Kurt: So, I'm going with that (starts writing on test – doing implicit)*
25. *Mary: But were not needing asymptotes for this*
26. *Kathy: Well yeah, but*
27. *Mary: Okay.*
28. *Kathy: It still tells you [\ where you set, which one you set for zero \]*
29. *Dorothy: [\ Where we find the tangent lines \] Yeah*
30. *Kathy: So it's [\ how, cuz you set the top for the one \]*
31. *Dorothy: [\ You got to, the top equal to zero \] for the horizontal*
32. *Kathy: And then the bottom is*
33. *Mary: (sassy tone) horizontal*
34. *Kathy: And then your bottom is vertical*
35. *Mary: I got ya*
36. *Kurt: (writing on test) F equals X, D equals*

This excerpt shows how the group discussed the mathematics even if one person is unsure of the group's approach. Also, note Mary claimed she spoke less in her post-test reflection, but she still voiced her concerns about what the group was thinking (line 25). After she voiced her concern Kathy and Dorothy explained why the analogy Kathy had previously given was still a correct approach for the answer (lines 28 through 34). Additionally, this conversation was occurring while the writer was representing the group's answer on the paper. The group interactions followed this pattern all throughout the second group test.

Although all of the women had a positive collaborative experience during the first test, Kurt was a little more on edge being placed in this particular group. In the first week of class Kurt had said he was unhappy with his group placement because he did not know Mary, Kathy, or Dorothy before this course. He was not upset about being placed in a group with three women, but had wanted to be placed in a group with women he already had experience working with in previous mathematics classes. During his interview at the end of the course, Kurt explained his feelings around his group and how they interacted. He stated:

> *At first [I was] uncomfortable and uneasy about it, but kind of accepting it. That it's, I mean I'm not gonna change my group obviously. […] So, I was like alright then. I mean I'm stuck [with] this, so I might as well make the most of it. […] If I was like, well, I'm in this study this is about females and I don't want to come off as sexist. […] I'm always afraid of it and it's my personality that I have to be accepted. That's just me […]. I felt like I didn't want to look like I came off […] like I control. Look, I want to control like, like, I want to like lead.*

His comment is a great example that when a person is placed into a position where the possibility stereotype threat is evident, this can help reduce the threat on women. That is, Kurt realized he was in a position in which he could have been seen as sexist. Therefore, he was more cautious about how he spoke to his group members. This feeling may have only been a byproduct of the study, or because he was the only male in the group. A conclusion on why he felt that way cannot be made at this time.

Overall, during the second group test the mathematical conversation included each group member's questions and ideas. Similar to their interactions during the first test, if one member was unsure of an answer provided by another member, the group explained the mathematics to them until everyone understood or agreed with the answer. The dominant voices in the second group test included Kathy, Mary, and Kurt. Dorothy did speak less in the group test, but was still involved in the conversation around the mathematics for each problem. Finally, all of the women noted the collaboration was beneficial in their post-test reflections, and all women felt they were a valued member (i.e., listened too) of their community (i.e., group).

Summary of the 3F-1M Research Group Interactions

Overall, each group member was included in the mathematical conversation during both the first and second group test. The three women in the group noted that they felt heard by their group members for both tests and expressed how their group worked well together especially during the second test. Additionally, Kathy had an approaching significant increase in her *MSoBS* score. Also, note Dorothy's score decreased specifically in the membership category. However, when we discussed the word "community" in her interview she said she felt she was part of a community in her group because every member helps one another. This increase in feeling as if she was a part of a community was not represented in her quantitative measure because she said she was not considering her group as the "mathematics community" in question on the survey. Specifically, she chose "strongly disagree" (i.e., a score of 1) for all of the questions in the membership component. Therefore, her feelings of belonging to the *group* were not represented in her quantitative *MSoBS* score.

The women of the group had increased feelings of collaboration and the group also increased in performance from the first to the second test (i.e., 85% on Group Test 1 and 100% on Group Test 2). Although the general focus of the study was women in calculus, it is important to note Kurt's feelings, particularly because his behavior may have had an impact on the women in his group. He stated how uncomfortable he was in the group, that he did not “like” all of his group members, and that he was afraid to come across as sexist in his group. Perhaps as the only male in the group and one with a concern about being classified as sexist, he did not activate the stereotype threat for the women in his group as it may have been activated for the women in the 2F-2M group. That is, in the previous group there was a possible impact on the women’s confidence due to activating a stereotype threat response that is not seen in the 3F-1M group. However, even though the group worked well together and performed very well on the two tests, and the women benefited from their experience, it should be noted that Kurt was not at ease in his group.

4(3)F Research Group Social Interactions

This group contained one URM female (Georgia) and two non-URM females (Maryam and Pratibha). The group also had one additional non-URM female (Rin) who was only present for the first group test. At the beginning of the course, Georgia and Pratibha had a high sense of belonging to mathematics and Maryam had a medium sense of belonging to mathematics (i.e., as measured by the *MSoBS* survey). Rin’s scores and post-test reflection results are not reported as she was removed from the course and did not take the second group test. Rin is included in this section because she was present during the first group test and cannot be parsed out from the data described.

The remaining three females had different types of changes to their sense of belonging to a mathematics community. Maryam was the only woman in the group to have an increase in her sense of belonging from a medium to approaching high score, Pratibha and Georgia both had decreases from high to medium sense of belonging. Specifically, Maryam increased from pre-*MSoBS* composite score of 4.73 to post-*MSoBS* composite score of 5.97. Pratibha decreased from pre-*MSoBS* composite score of 6.23 to post-*MSoBS* composite score of 5.50. Georgia decreased from pre-*MSoBS* composite score of 7.07 to post-*MSoBS* composite score 5.83. Although the three women in this group started with different overall sense of belonging scores (i.e., one low, one medium, and one high), they all had a medium to approaching high score at the end of the course. Additionally, this is the only group who's test performance was consistent between both tests (i.e., their score on the first and second group test was nearly the same).

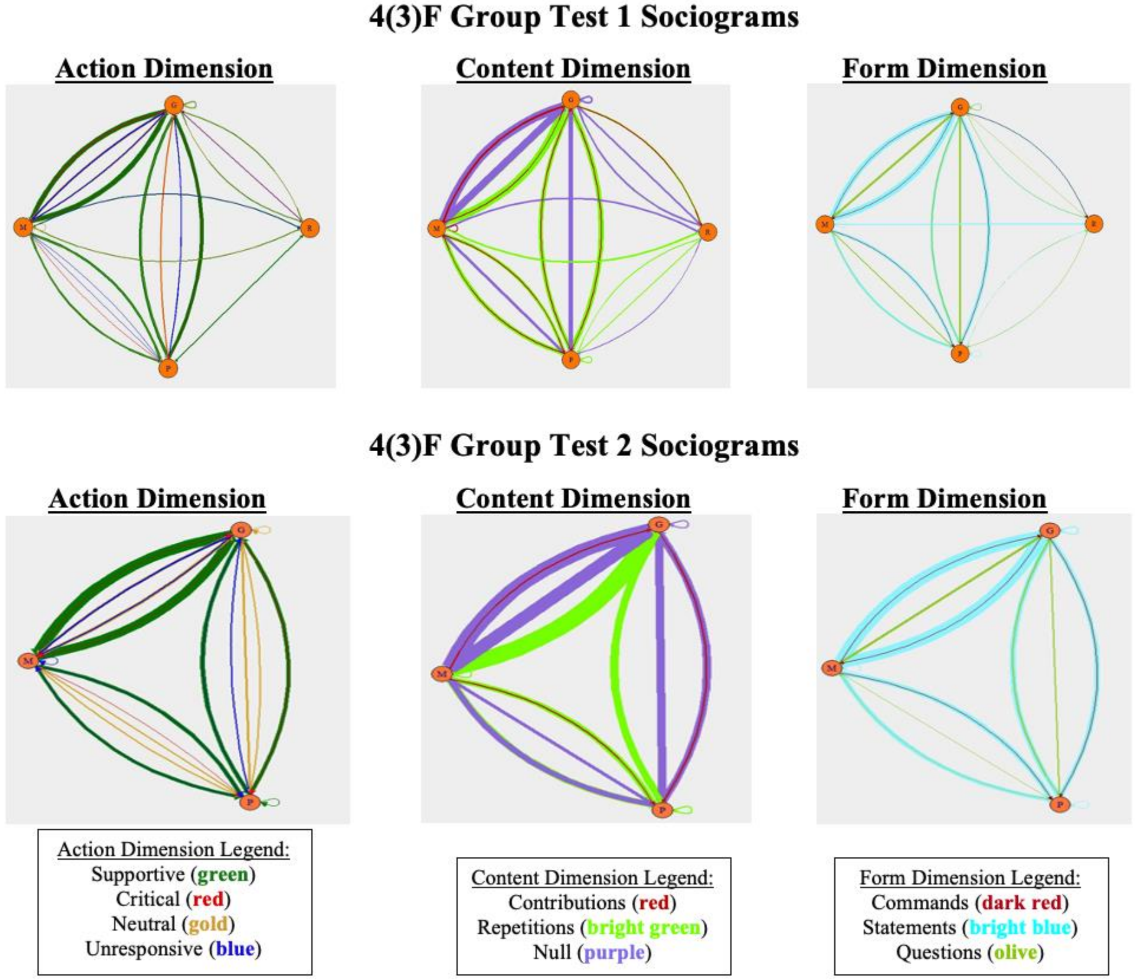
Interactions during Group Test 1

The group earned a 61% on the first group test. After the group test the students took an individual test in which Maryam received a 71%, Georgia received a 49%, and Pratibha received a 32% on the individual test. Maryam was the only student in the group who scored higher on the first individual test than on the first group test. The majority of the mathematical talk occurred between Maryam and Georgia.

The interactions during the first test along the three dimensions of the group problem solving framework (i.e., Action Dimension, Content Dimension, and Form Dimension) are displayed in Figure 29.There are no teal nodes in the sociogram as there were no men in the group. All group members are represented by the orange nodes in the sociograms. The M node is for Maryam, G node is for Georgia, P node is for Pratibha, and the R node is for Rin. Figure 29 shows the majority of the talk during the first test occurred between Maryam and Georgia. This can be seen by the thickness in the lines along each dimension, particularly in the Action Dimension the thickest lines are the green lines between the M and G nodes. This means the most frequent type of talk that occurred during the test was supportive talk between Maryam and Georgia. The second most talk was between both Maryam and Georgia to Pratibha. Rin spoke the least and was spoken to the least during the first group figure.

Figure 29.

Sociograms for the 4(3)F Group's Social Interactions during Group Test 1 and Group Test 2.



In the sociograms, particularly in the Action Dimension for Group Test 1, there are green lines going from each participant with the thickest between Maryam and

Georgia then Georgia and Pratibha. This means the most frequent type of talk that occurred during the test was supportive talk between Maryam, Georgia, and Pratibha. There are very few lines connected to and from Rin indicating she did not speak as much as her group members. Considering the Content Dimension for Group Test 1 there are thick bright green lines connecting the M, G, and P nodes. These lines indicate the majority of the mathematical talk was between Maryam, Georgia, and Pratibha. Additionally, notice in the second image there are many thick red lines stemming from each node except Rin's node to another that are all relatively the same thickness. This indicates Maryam, Georgia, and Pratibha offered multiple new mathematical ideas (i.e., contribution) roughly the same number of times to each other during the test. Finally, in the Form Dimension for Group Test 1 there are a number of commands given by Georgia to all other participants, and given by Maryam to both Georgia and Pratibha. That is, the only commands given in the group stemmed from Maryam and Georgia and each group member was spoken to in a commanding way relatively the same number of times (i.e., the red lines in the image are the same thickness). Considering the three images together, it appears the women did respond to each other in supportive ways, however, they also offered and questioned many of their mathematical ideas during the first test. Examining Maryam's, Georgia's, and Pratibha's post-test reflections helps us understand how these interactions impacted the sense of belonging for these three female students.

The three women in this group all felt they did communicate their mathematical ideas during the test, but also felt that their contributions were not necessarily received by their group members. Consider their responses to the question "How do you feel your contributions were received in the group during the test?":

> *Maryam: I feel they were received well by Georgia, but Pratibha and Rin seemed to think faster than they could write and had to cross or scribble things out multiple times.*
>
> *Georgia: They took in a few answers I suggested but not all.*
>
> *Pratibha: My contribution was received, but, not the way my group want[ed]. I think I should be more contribute than I was. All over all I was not prepare very well.*

From the comments above show that all of the women did not fully feel they were heard by their group. Each of these comments reveals a negative impact on their sense of belonging, but in different ways. For example, Maryam did feel her and Georgia worked well together, but that did not extend to the rest of the group. That is, Maryam exhibited a need for more collaboration with Pratibha and Rin. Georgia expanding on her comment said "not everything I suggested about doing the test neatly was taken" which exhibits a lack of control during the test (i.e., controlling how the answers were written on the paper). Maryam's comment echoes the lack of control Georgia stated about how the mathematics was presented on the group test (i.e., scribble things out). Last, Pratibha showed in her comment a lack of confidence in her mathematical understanding (i.e., not prepared). This illuminated three different negative impacts (i.e., wishful collaboration, lack of control, and *need for* confidence) for each group member.

Often during the first test, Maryam and Georgia would instruct the writer for the problem on how to present the group's answer. Consider the following exchange from the third problem on the first test, in which Pratibha was the assigned writer for this problem:

1. *Pratibha: Okay, state the definition of derivative using mathematical notation and your own words. What does [it] mean and it tells you so, finding the tangent line, right?*
2. *Georgia: Of the deriva-*
3. *Maryam: Well*
4. *Georgia: the derivative definition?*

5. *Pratibha: Yeah, tangent line.*
6. *Maryam: it says using mathematical notation. Doesn't that mean like actually writing stuff out? Like the equation and stuff?*
7. *Georgia: Isn't that, wait, wait don't write it yet, but um the F of X plus H*
8. *Maryam: Yeah. [\ The limit \]*
9. *Georgia: [\ minus F of X \] over H*
10. *Maryam: With the before it H approaches zero.*
11. *Georgia: So limit as H approaches, the limit as H approaches zero. What is that?*
12. *Pratibha: Limit*
13. *Georgia: Why don't you write just L I M?*
14. *Pratibha: Okay*
15. *Georgia: [\ And then the \]*
16. *Maryam: [\ And then F \] parenthesis X plus H.*
17. *Pratibha: F right?*
18. *Maryam: F before the parenthesis cuz F of, like how we do the function*
19. *Pratibha: Okay you want me to write prime?*
20. *Maryam: Like this, no we don't need prime. In the parenthesis put F plus H.*
21. *Pratibha: In parenthesis like this?*
22. *Georgia: mm-hmm*
23. *Pratibha: X plus H*
24. *Georgia: Minus F of X, over H.*
25. *Georgia: Make sure that's like, make sure that can be seen as an H, I just.*
26. *Pratibha: That's fine.*
27. *Georgia: Can you put an X over that, it looks like it's mixed in //*
28. *Pratibha: \\ X?*
29. *Georgia: Uh, no, ex over it [\ so she knows it, yeah scribble over it \]*
30. *Maryam: [\ No, cross that, scribble over it \] cuz it's going to look like it's part of the equation.*

Notice both Georgia and Maryam are instructing Pratibha how to write the answer on the paper. Specifically, Georgia questions how Pratibha is writing the limit (line 11 through 13) and Maryam instructs Pratibha to correct what she has written (line 30). Additionally, Pratibha attempted to offer an initial idea (i.e., tangent line) in which her group members did not directly respond to her. Georgia and Maryam appear to be leading the group, and Pratibha does not feel as if she is being heard by her group members. Although, neither Georgia or Maryam question what Pratibha meant by "finding tangent line," this could have been a moment where the women discussed the offered idea. If a discussion around

Pratibha’s idea occurred, perhaps she would not have internalized her feeling of not being heard as her not expressing her knowledge and being unprepared for the test.

The group's interactions during the first test followed the pattern as seen in the excerpt above. Maryam and Georgia attempted to lead the group and were very particular on how the answers should be written. Pratibha would often offer ideas or single word utterances that were not discussed by her group members. Last, Rin’s voice was rarely heard during the test. These interactions help explain the feelings of frustration (i.e., lack of control) and a need for collaboration (i.e., wishful collaboration) in the group. The group’s interactions slightly changed for the second group test. Recall, Rin was not present for the second group test. Additionally, Pratibha did speak much more often and was involved more in the mathematical conversation during the second test as will be discussed in the next section.

Interactions during Group Test 2

For the second group test, the group performance was similar to the first test and there was a slight change in the group dynamics. That is, on the second group test, the group earned a 63% on the group test, and for the second individual test Maryam and Pratibha received grades higher than the group test score (i.e., 86.5% and 82% respectively) and Georgia received a grade lower (i.e., 43.5%). The majority of the mathematical conversation still occurred between Maryam and Georgia during the second test. In this section I provide details about the group’s interactions during the second group test and the impact those interactions had on the female students with regard to their sense of belonging. There was a slight shift in the group dynamics during the second group test which is evidenced in the sociograms (Figure 29). In the sociograms for Group

Test 2, there was an increase in lines connecting the three nodes for Maryam, Georgia, and Pratibha from Group Test 1 sociograms to Group Test 2 sociograms. This indicates the conversation was among all three of the women in the group during the second test. However, the majority of the mathematical talk during the second test was still between Maryam and Georgia. This is evidenced by the thicker lines connecting the M and G nodes. Though, Pratibha did speak more and was more involved in the mathematical conversation.

Considering the Action Dimension in Figure 29 for Group Test 2, the majority of the talk was supportive between all three women, but particularly between Maryam and Georgia. A number of the talk turns in response to and from Pratibha were neutral. This is shown in the first image with the number of yellow lines connected to the P node. In the Content Dimension for Group Test 2, there is an increase in thickness of the purple and green lines. The purple lines are thicker than the green lines indicating more talk during the test did not contain any mathematical statements or ideas. Additionally, the contributions (i.e., initial ideas) originated only from Maryam and Georgia. This can be seen from the red lines in the Content Dimension stemming from the M and G nodes. Finally, during the second test, the commands given by both Maryam and Georgia to each other and to Pratibha. That is, the red lines in the Form Dimension for Group Test 2 between Maryam and Georgia in the third image indicates multiple commands were delivered by both participants to each other. And both Maryam and Georgia gave commands in response to Pratibha. The increase in talk among all three of the females in this group was noted in each of their individual post-test reflections.

Maryam, Georgia, and Pratibha all noted the collaboration their group had during this second test in their post-test reflections. Additionally, none of the women provided a negative comment that was not specific to the mathematics on their second post-test reflections (i.e., they all stated they were unsure about one of the problems). On her post-test reflection, Georgia noted how each member of the group was involved during the second test. Specifically, in her post-test reflection, Georgia stated "*My group makes sure they put in work plus [we] don't just rely on each other without making sure something is correct.*" In this comment, Georgia explained how every member was involved in the test instead of one person leading the group. Pratibha also noted her increase in contributing to the test, she stated "*I think for [the] second test I did about 40% [of the] contribution, I felt that I knew something.*" During her interview Pratibha elaborated on this comment and explained that she felt more prepared for the second group test and was able to discuss the problems with the group more than she could during the first group test. Maryam also noted how her group members "added more" to her ideas during the second group test.

In her second post-test reflection, Maryam still acknowledged her dislike for group tests (i.e., she stated "*[I] personally do not like group [tests]. Always hard. I don't feel any benefit from them. (I even had them in high school.)*"). She explained in her interview that she would rather take tests by herself because she feels individual tests are "easier." She explained this feeling of "easier" was because she can go through the test and pick what parts of the problems to start with until she "gets stuck" and move on. When discussing the group test, Maryam stated "*I can never like pick out the parts, like what information [is] actually needed.*" This comment is the extent of her explanation of why she dislikes group tests, and therefore it was left unclear to me as the researcher why she felt this way about the tests. However, this comment alludes to the lack of control she may have felt during a group test as she was not able to do something in the group test (i.e., pick out parts) that she could do if she was taking the test by herself.

In her interview, Maryam did explain how much she enjoyed working and talking with Georgia during the group setting and test, but that these feelings did not extend to Pratibha and Rin. In her interview, Maryam stated:

> *[during the tests] me and Georgia [would] like read the questions and we try to come to our final answer before we write anything down. And they [Pratibha and Rin] are like let's write it and we're like no wait, let's make sure first*

In this excerpt, Maryam explains how her and Georgia attempted the problems in a similar manner (i.e., discuss before writing). Georgia also explained that she liked working with Maryam as they both were concerned about the neatness of the test.

Georgia felt Maryam contributed the majority of the ideas for the group tests, and she (Georgia) felt she could learn from Maryam. Specifically, during her interview Georgia stated:

> *I feel like Maryam is the one, like she knows more. Meaning like she gives a lot of her answers and most of the time it's right. It's awful like we all put our input but [...] she like beats us to the punch. But it's like she usually knows everything like you know off the spot. I can really learn from her basically is what I'm saying.*

This comment illuminates Georgia's desire to be an active participant in her community (i.e., group) and her willingness to listen to and learn from her community members (i.e., Maryam). Considering both Maryam and Georgia's comments, their community did not seemingly include Pratibha. That is, both Maryam and Georgia explicitly spoke only about each other when talking about the benefits of the group's collaboration during their interviews. On the other hand, Pratibha was positioned, by herself and Maryam and Georgia, as a non-contributing member to the group. This is evidenced by Pratibha mentioning how she felt embarrassed and specifically stated *"I feel like I am dumb and like I am stupid, I do not know how to do this"* when referring to working with a group as well as the few interactions she had during the group tests.

Summary of the 4(3)F Research Group Interactions

This was the only research group to score similarly on both of the group tests (i.e., 61% on Group Test 1 and 63% on Group Test 2). Overall, the majority of the mathematical conversation for both tests occurred between Maryam and Georgia. Pratibha, however, was more involved in the conversation during the second test and she did feel she contributed more to the mathematical ideas for that test. The three women in the group noted the benefit of the collaboration they had during the second test. Maryam noted on both tests that she did not enjoy nor felt she benefitted from group tests, however, she had an increase in her *MSoBS* score at the end of the course. Additionally, even with this increase Pratibha and Georgia both had a lower *MSoBS* score at the end of the course. Maryam and Georgia both explained that they enjoyed working with each other, but did not speak the same about Pratibha.

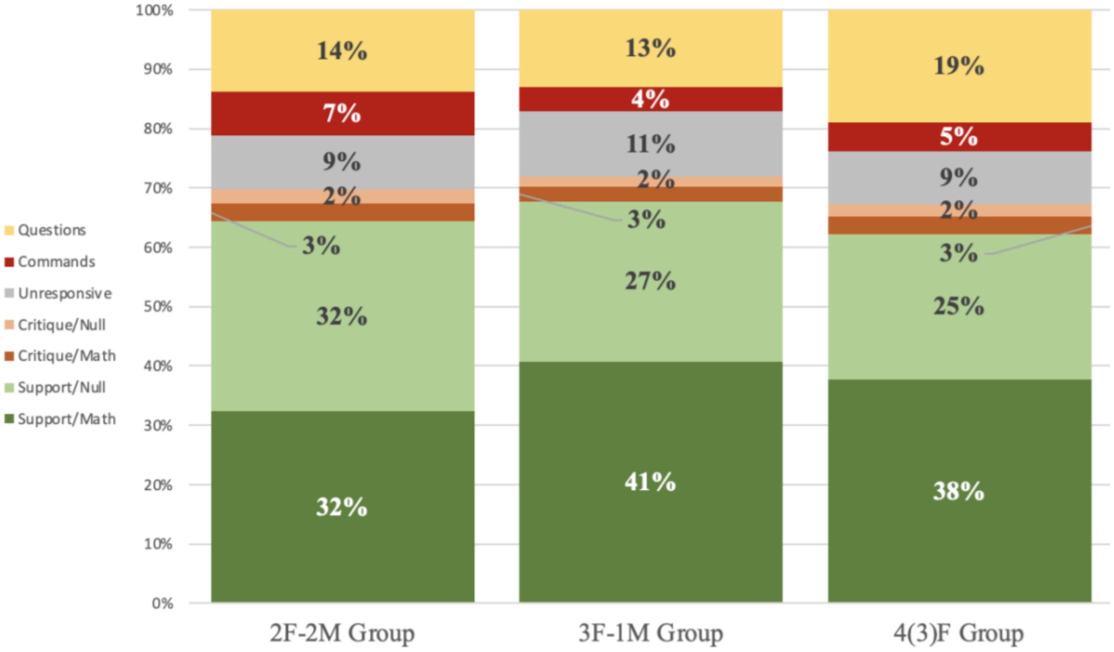
The increase in Maryam's *MSoBS* and decrease in Georgia's *MSoBS* scores could be attributed to Georgia's positioning of Maryam as the intellectual authority of the group as indicated by her statement of wanting to learn from Maryam. However, this opens more questions on the impact of the social interactions to the women's sense of belonging as changes in sense of belonging for Maryam (a non-URM female) and Georgia (a URM female) did not follow the patterns of the other groups nor the majority of the non-URM and URM females in the study.

Comparison of the Impact of Group Interactions

The overall group interactions were condensed into seven main types of talk turns: supportive with mathematical content, supportive without mathematical content, critiques with mathematical content, critiques without mathematical content, unresponsive statements, commands, and questions. Figure 30 displays the percentage of each theme per group in order to compare the three research groups social interactions during the two group tests. As seen in the groups' sociogram descriptions, the group with the most supportive talk among all members that contained mathematical content was the 3F-1M group (ie.,41% of their total talk was supportive with mathematics content) . The 2F-2M group contained the most supportive talk without mathematics content (i.e., 32% of their total talk was supportive without mathematical content). All three groups had the same amount of critical talk with mathematical content (3%) and without mathematical content (2%). Among the groups, the 3F-1M group had the most unresponsive talk (11%), the 2F-2M group had the most commands (7%), and the 4(3)F group had the most questions (19%). Considering the overall themes per group, only the 3F-1M and 4(3)F groups had the largest amount of on-task talk turns as supportive talk centered around the mathematics. That is, for the dark green in the figure below, the largest percentage for these two groups (i.e., 38% and 41%) were the largest percentage for these two groups. The 2F-2M group had the same amount of supportive talk with and without mathematical content (32%).

Figure 30.

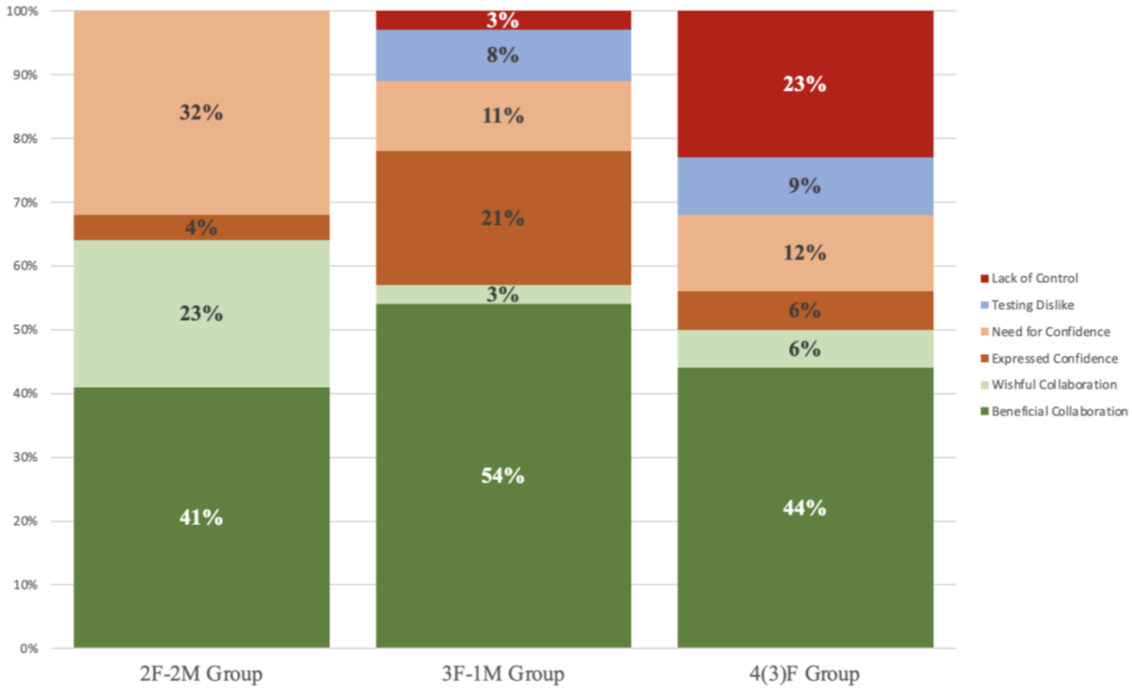
Distribution of Social Interaction Themes Per Research Group.



Differences were found when comparing the groups interactions to the females' sense of belonging themes from their post-test reflections per group (see Figure 31). Although the largest occurrence of sense of belonging themes is beneficial collaboration for all the females, the women in the 3F-1M group had the largest occurrence for this theme (i.e., 54%). The 3F-1M women also had the largest amount of expressed confidence (21%) among the women in three groups. The 2F-2M group had the most commands from their interactions during the test (7%) and the female students had the most occurrence of both wishful collaboration (23%) and need for confidence (32%). These differences may be explained by the difference in the frequency of questions, commands, and supportive mathematical talk in each of the three research groups.

Figure 31.

Distribution of Sense of Belonging Themes Per Research Group.



Note that themes present in the all-female group represent the total number women in the study placed for this specific type of group composition (i.e., there was only one all-female group). Additionally, the themes present in the 3F-1M group may proportionately represent the themes for all of the women placed in this type of group composition (i.e., there were two 3F-1M groups in the course). Interestingly, however, there was a different pattern of themes for the sense of belonging of the female students than seen previously in this chapter (i.e., Table 15). When considering the overall themes from all of the females in 2F-2M groups, the six females (i.e., all the women placed in 2F-2M groups) expressed a larger amount of frustration (i.e., 12% lack of control); that was not present in the observed 2F-2M research group.

Chapter Summary

In this chapter, I presented the results from an exploratory mixed methods study describing the test performance for both group and individual tests across gender and across the intersection of race and gender, the changes in quantitative *MSoBS* for the female participants across race, the qualitative sense of belonging themes that emerged for the females organized across participant race and group composition, and the social interactions that occurred in the three research groups. Additionally, I provided possible explanations for the differences in the qualitative sense of belonging themes for the females in the three research groups using the differences in the three groups' social interactions during both of the group test settings. The findings from the test data indicated that the majority of the students in the course performed better when tested in groups for all students except the non-URM females. The findings from the sense of belonging quantitative data indicated no change in sense of belonging for the majority of the URM females, and a decrease in the trust and affect component for the majority of the non-URM females. The qualitative sense of belonging data were classified into four main themes: collaboration, confidence, testing and control. The non-URM females had a less positive experience (i.e., beneficial collaboration) than the URM females. Additionally, the non-URM females had a more negative experience with respect to the testing and control theme than the URM females. Finally, the social interactions during the group tests were classified into seven main themes: supportive (with or without mathematical content), critiques (with or without mathematical content), unresponsive statements, commands, and questions. Although all of the research groups had over 60% of supportive interactions during the two tests (i.e., a combination of supportive talk with and without mathematical content), the 3F-1M group had the most supportive talk containing mathematical content. The females in the 3F-1M group also reported the most benefit from the collaboration and expressing confidence in their post-test reflections. The 2F-2M group had the most commanding statements and the females reported the most need for confidence in their post-test reflections. The 4(3)F group had the most questions posed during the group test, and the females reported the most lack of control in their post-test reflections. The results from this study were not aligned with my original hypothesis, but I am unsure whether the results for the non-URM females are due to the natural variability of the data due to the small sample size or due to the intersectionality of the students' identity. I provide a summary and discussion of the results in the next chapter.

CHAPTER FIVE: DISCUSSION

Introduction

The main purpose of my study was to examine the social interactions of the group members during a group test, the differences in the interactions for groups with different compositions, and how the interactions might explain changes in the women's sense of belonging to mathematics in a reform Calculus I course. I utilized group testing, a student-centered assessment method, as the catalyst for the students' interactions during this course. Specifically, my study is situated at the intersection of the research on undergraduate calculus reform, women in mathematics, equitable education techniques, and group testing. There has been an increase in use of equitable instructional techniques (i.e., student-centered activities) in mathematics classrooms and group testing has been noted as a possible equitable testing method in some mathematics classrooms (Goetz, 2005; MacArthur, 2019). Group testing has been shown to increase student communication and collaboration, and help students focus more on learning and less on the grade (i.e., reducing the competitive nature of the course; Duane & Satre, 2014; Jang et al., 2017; Lusk & Conklin, 2003; Revere et al., 2008). With many students working and being tested together instructors must be aware of the possible threat this may pose to women's sense of belonging in mathematics classrooms (Good et al., 2008; Oswald & Harvey, 2000; Spencer et al., 1999). In this study, I hypothesized that by increasing female students' social interactions within a group, their sense of belonging would increase, and ultimately lead to increased collaboration, performance, and persistence in the STEM pipeline (Oswald & Harvey, 2000; Thorsen et al., 2019).

I chose to focus on women in calculus because female students have many obstacles to overcome if they intend to enter the STEM workforce such as fighting traditional gender roles (Shapiro et al., 2015), stereotype threat (Cadaret et al., 2017; Blickenstaff, 2005; Good et al., 2008), and having lowered confidence in their mathematics skills (Ellis et al., 2016; Shapiro et al., 2015). Particularly, the purpose of this study was to investigate women's sense of belonging and group interactions in a reformed Calculus I course while engaged in group testing. I investigated the following three research questions:

1. In a reform Calculus I course, how do students' performance on group tests compare with their performance on individual tests; and how do gender, and the intersection of race and gender contribute to the differences, if at all?
2. How does the sense of belonging to a mathematics community change for female students after engaging in group testing in a reform Calculus I course, if at all?
3. How are social interactions within the group composition different based on the composition of the group, if at all, and can any differences be used to explain differences in the female students' sense of belonging?

In this chapter, I present a review of methodology utilized and a brief summary of the results of the study. This review will be followed by a discussion of the results of the study, which will include connections to prior research, practical implications, and recommendations for future research.

Review of Methodology

I used an exploratory mixed method design to investigate female students' sense of belonging and social interactions in a Calculus I reform class through the use of group testing. This allowed me to explore the viewpoints of multiple participants involved in the same experience (i.e., a group test) while collecting both quantitative and qualitative data. I examined the test scores of 30 students who had been placed intentionally in eight permanent groups. I selected 15 female students and examined their sense of belonging. I selected 3 of the groups (a total of 11 students) and examined the social interactions that occurred during the group tests. Finally, the 8 women in those three groups provided reflections designed to explain changes in their sense of belonging to mathematics over the course. I collected data for three main purposes for the study. First, I collected data to arrange the students in their permanent working group. That is, consideration was carefully taken to place each student in their group for the course based on their responses to the *Working Group Survey*. The remaining collected data were analyzed to answer the research questions. Multiple sources of data, including individual and group test scores, pre- and post-*Math Sense of Belonging Scale* (*MSoBS*) survey scores, two post-test reflections, and video, audio, and LiveScribe recordings for three groups during both group tests were analyzed to understand the possible link between the group's interactions to the female students' sense of belonging. Additionally, I presented the test scores for 30 students arranged in eight groups to answer the first research question. In the next section, I present a summary of the results for the three research questions.

Summary of Results

In the following subsections I provide an overview of the pertinent findings that answer each of the three research questions. For a full description of the results please see Chapter 4.

Testing Performance

To answer the first research question, I presented the individual and group test data (i.e., student performance) for the 30 students who were arranged into eight groups. There were little differences in performance among the students on the first pair of tests. The majority of students scored higher on Group Test 1 than an Individual Test 1. Only one non-URM male and one non-URM female scored notably higher on Individual Test 1 than Group Test 1 (i.e., score at least one letter grade higher as in from a D to a C). The results of the student test performance on the second pair of tests indicated that the intersection of race and gender may have contributed to differences in student performance. The non-URM females were the only subset of students (i.e., by race and gender) to score lower on the second group test than on the second individual test. All but one of the non-URM females scored notably higher on the Individual Test 2 than on Group Test 2, and one non-URM female scored nearly the same on both tests (i.e., a difference of only 0.3% points). The only other students to score higher on Individual Test 2 were two non-URM males.

Note that the majority of the non-URM female students were placed in 2F-2M groups (i.e., four of the seven non-URM women), two were in the all-female group, and only one non-URM woman was placed in a 3F-1M group. Although no official comparison of test performance by group was provided as the sample size was too small, it is apparent this composition would have similar results (i.e., differences) for the second pair of tests due to the placement of the non-URM women. Additionally, the two non-URM male students who scored higher on Individual Test 2 than on Group Test 2 were also placed in two different 2F-2M groups. Overall, I cannot determine if the difference in performance is due to the group composition or the number of non-URM females in the group. To answer the first research question, however, I can say that both the intersection of race and gender and the group composition may be factors that contributed to student performance on the group tests.

Sense of Belonging

To answer the second research question, I presented the quantitative and qualitative data (i.e., *MSoBS* scores and post-test reflections) for the 15 female students by race. The comparison was examined across URM status (i.e., race) due to the previously seen difference in group test performance for the non-URM and URM women on the second group test, and the need for more studies to consider multiple dimensions of students' identities (Leyva, 2017). The changes in the five components of the *MSOBS* measure (i.e., membership, affect, acceptance, trust, and desire to fade) were different for nearly all categories for the non-URM and the URM females except for the desire to fade component. That is, comparing changes across race, the non-URM females had more decreases in all categories except membership, and the URM females had more increases in all categories except membership and desire to fade. The majority of the URM females' sense of belonging stayed the same throughout the course for both the overall score and each of the five subcomponents. The only notable increase in sense of belonging was in the membership component for the non-URM females, in which the majority increased in feelings of membership. The majority of non-URM females had a decrease in their feelings of affect, where the majority of the URM females did not change in their sense of belonging along the five components. Additionally, the majority of non-URM females either decreased or stayed the same in the acceptance component. These results could indicate that the URM female students' sense of belonging did not change over the time in the course, but the non-URM female may have changed. Specifically, more non-URM females felt as if they were part of a mathematics community (i.e., membership component), but that they were not a valued member (i.e., acceptance component), they were not comfortable (i.e., affect component), and did not have trust in the instructor or testing materials (i.e., trust component).

The analysis of the post-test reflections illuminated four main themes that may help explain the changes in students' sense of belonging: 1. collaboration, 2. confidence, 3. testing, and 4. control. The non-URM students had less occurrence of beneficial collaboration and overall more occurrences indicating a negative experience (i.e., testing and control) than the URM females. All of the women in the different composed groups noted near the same amount of beneficial collaboration in their groups. The women in the 2F-2M groups, however, had fewer feelings of self-confidence and noted a wish for more collaboration than those in the 3F-1M groups. Additionally, the all-female group had more feelings centered around the lack of control (i.e., frustration) experienced during the group tests. Overall, the non-URM females did not benefit from the group testing experience in terms of confidence and frustration, and felt the test to be more biased than the URM females. The results from the post-test reflections do help explain the decrease in the quantitative trust component for the non-URM female students. The main concern seemed to be the method of testing as unfair which is directly asked in one of the questions on the *MSoBS* survey.

Group Interactions

To answer the third research question, I presented the qualitative data of the group interactions for the three research groups. Analysis of the group interactions illuminated seven main themes: (1) supportive talk containing mathematical content, (2) supportive talk without mathematical content, (3) critiques containing mathematical content, (4) critiques without mathematical content, (5) unresponsive statements, (6) commands, and (7) questions. Overall, the groups shared many similarities in their interactions (i.e., a large number of supportive statements among members), but ultimately did not interact the same when discussing the mathematics (i.e., differences in unresponsive statements, commands, and questions). For each of the three research groups over 60% of the total social interactions (i.e., combined codes from Group Test 1 and Group Test 2) consisted of supportive talk. The 3F-1M group had the largest amount of supportive mathematical talk (41%), the 4(3)F group had the second largest amount of supportive mathematical talk (38%), and the 2F-2M group had the smallest amount of supportive mathematical talk (32%). Since each group did have a large amount of supportive talk, the comparison of the groups interactions was conducted by looking at the differences seen among the three research groups.

Comparing the group interactions to the women's' qualitative sense of belonging themes (i.e., their post-test reflection responses), three main findings emerged from the differences among the three groups. The women in the 3F-1M group had the largest amount of beneficial collaboration and expressed confidence and their group's interactions had the largest amount of supportive talk (with or without mathematical content). The women in the 2F-2M group had the largest amount of need for confidence and their group had the largest number of commands generally given by the one student in their group who was considered the intellectual authority (i.e., Isaac). The women in the 4(3)F group expressed the largest amount of lack of control (i.e., frustration) and this group had the largest number of questions in their interactions. The results from the social interactions of the three groups indicate the more women in the group the more the group talk supportively around the mathematics. The presence of one man in the group, however, may increase the women's confidence, lower their frustration, and possibly lower the number of questions asked by the women in the group.

Discussion of Results

In this study, I examined the performance and sense of belonging to a mathematics community for women who engaged in group testing in a calculus course while placed in non-randomized groups. Specifically, I investigated the relationships between students' sense of belonging, group interactions, and their participatory learning measure (i.e., group test performance). The findings from this study contribute to the literature in three main ways. First, I discuss how group testing provides students a space to construct knowledge collaboratively and instructors with a way of assessing students' participatory learning. Second, the results of this study indicate a decrease in two main components of women's sense of belonging after engaging in a group test (i.e., trust and affect) and one that should be considered in future work (i.e., acceptance). In particular, I discuss the importance of building trust, affect, and acceptance for the women in the undergraduate calculus course. Finally, I discuss how group testing may be the factor associated with chances in female students' sense of belonging: specifically, how the group testing environment may have differentially impacted URM and non-URM females' sense of belonging. Therefore, in contrast to results indicated by previous literature, group testing might not be equitable for all students when considering only a single identity (i.e., gender). Expanding on the possible inequity of group testing, this contribution relates to the intersectionality of student identities that should be considered as part of promoting equitable learning and assessment in undergraduate mathematics (and STEM) classrooms, especially within the use of student-centered learning activities.

Group Testing as a Learning Space

The testing is a student-centered assessment method in which students are required to communicate and collaborate to solve problems. This method of assessment has previously been explored across multiple disciplines: nursing (Eastridge, 2018; Duane & Satre, 2014; Lusk & Conklin, 2003; Sandahl, 2010), psychology (Curless, 2012; Pandey & Kapitanoff, 2011; Vogler & Robinson, 2016), physiology (Cortright et al., 2003; Giuliodori et al., 2008; Rao et al., 2002), and biotechnology (Srougi et al., 2013), general science education course for undergraduate students (Gilley & Clarkston, 2014), non-mathematics majors’ statistics courses (Kapitanoff & Pandey, 2018; Revere et al., 2008), engineering students in a mechanics course (Jang et al., 2017), and mathematics (Berry & Nyman, 2002; Goetz, 2005; MacArthur, 2019; Paterson et al., 2013). Furthermore, Hattie and Timperley (2007) suggested that testing methods can be a type of learning space as assessments can influence student learning. Therefore, it is within this learning space that I situated my study.

Since student learning can also be influenced by the assessment methods used in the classroom (Roediger & Pyc, 2012), if students are only assessed on the acquisition of their internally constructed knowledge, I argue this would not maintain the balance needed for student's holistic learning of mathematics. The group testing environment allowed for students to "bounce ideas off of each other" (i.e., collaborate) and come to a consensus to answer conceptual problems about calculus topics (i.e., participatory). Whereas the individual test assessed students' skills on performing procedures directly related to those calculus concepts represented in the corresponding group test (i.e., acquisition). The student-to-student interactions during the group testing environment provided a learning space for students to construct and express their knowledge during the test (e.g., Cobb, 1994; 1995), and allowed for the instructor to assess the participatory learning of each group. Therefore, the use of both individual and group tests for this study called for a framework to recognize both acquisition (i.e., individually constructed knowledge) and participatory (i.e., participation within a community) learning metaphors (i.e., Sfard, 1998).

Both scholars and policy (NRC, 1993; Roediger & Pyc, 2012; Webb, 1995) recommended alignment between the learning space and assessment methods used in mathematics classrooms. In the present study, the classroom learning environment included both group work and individual work. As such, the testing methods used included both individual and group tests. Group testing has been shown to reduce the competitive nature of STEM courses and allow students to focus on their collaborative efforts, increase their social interactions through these effects, and increase their confidence (Revere et al., 2008). Additionally, Paterson et al. (2013) and Pedersen and

Liu (2003) noted that students who engaged in group tests were more focused on their communication with their fellow students than on the outcome of the test (i.e., their grade). Considering just the student performance in this study, the majority of the students scored higher when tested in a group (as compared to individual tests).

Overall, 27 out of the 30 students in the course scored higher on their Group Test 1 than on the Individual Test 1, and 22 students scored higher on Group Test 2 than on the Individual Test 2. Of the eight students who scored higher on their second individual test (rather than on the second group test), six were non-URM females (discussed further below). Because the majority of the students in the course performed better when tested as a group, in which the group members had to collaboratively construct their knowledge to answer the questions, group testing may be considered a learning space. If we consider only the performance on the second group test, however, in which six of the seven non-URM female students did not perform better in the group, the environment may be considered biased across gender. That is, not all male and female students performed similarly. This finding is in contrast to previous findings in which no statistical differences were found on testing performance based on gender (MacArthur, 2019).

If group testing environments are to be considered a learning space, then we must also consider how this environment might be experienced differently by different types of students: in this study male and female. Recall from Chapter 2 the ten needs of the learner that must be met for learning to occur (Sfard, 2003). In this study, I specifically focused on the need for female students' sense of belonging. More importantly, the Balanced Learning Needs framework suggests that when a student's learning need is not met, the system becomes imbalanced and learning cannot be achieved holistically. That is, lowered sense of belonging could impact student performance (Good et al., 2012) especially in a testing environment (Danaher & Crandall, 2008).

Group Testing Differentially Impacts Female Students' Sense of Belonging

A sense of belonging within collaborative learning environments has previously been shown to improve students' performance and retention in undergraduate STEM courses (Diekman et al., 2015; Good et al., 2012). Recall in the Balanced Learning Needs framework, both social interactions and sense of belonging are considered social needs of the learner (i.e., learning towards the participatory metaphor for learning). Therefore, in order for this learning need to be met, there must be a social context (i.e., people, group, or community) with which the learner engages. The group placement (i.e., the group in which students were placed) could be considered the social context for social interactions and sense of belonging, as the social interactions depend on the people placed in each group. Goodenow (1993) defined a student's sense of belonging as their "sense of being accepted, valued, included, and encouraged by others (teachers and peers) in the academic classroom setting and of feeling oneself to be an important part of the life and activity of the class" (p. 25). This definition aligned well with Good et al.'s (2012) five components: membership, affect, acceptance, trust, and desire to fade. Therefore, I used these five components to measure and understand the change in the sense of belonging for the women in the study. The combination of these components encapsulated the overall students' sense of belonging to a mathematics community. Moreover, a sense of belonging is context specific with regard to one's community (Strayhorn, 2019), therefore the mathematics community for this study was the classroom and the group in which each student was placed.

When considering all of the females (i.e., both URM and non-URM together), the findings from this study suggest that there were no overall changes in URM females' sense of belonging throughout the course. As noted by Leyva (2017), however, we must attend to the intersectionality of students' identities (i.e., race and gender) when considering performance and affective gains. Therefore, I tried to understand the change in belonging along the intersection of gender and race. In the present study, membership was the only sense of belonging subcomponent to increase for the non-URM females, and trust and affect were the only sense of belonging subcomponents that decreased during the course for the majority of the non-URM female students. Additionally, for the acceptance subcomponent the majority of the non-URM females either decreased or remained the same. Although membership, trust, affect, and acceptance were the only notable changes in the *MSOBS* measure, the lack of changes for the URM females may have been due to the program in which the students were enrolled. That is, recall from Chapter 2, this study was conducted in a program for STEM students in which six of the eight characteristics of successful calculus programs were implemented (Bressoud & Rasmussen, 2015; Hagman, 2019). The program was created to diversify the STEM majors at the university and students went through the first year together in order to help retention and build a community feeling (Carver et al., 2017; Van Sickle et al., 2020). Therefore, this could help explain why there were few changes for the URM women. If the program was created to help students feel as if they belong in a STEM domain, however, the changes in the trust, affect, and acceptance components then may be attributed to the group testing environment (as this was not a part of the original program).

Prior to the study, I hypothesized that changes in students' *MSoBS* would be related to the social interactions experienced in their groups. As a lowered sense of belonging can be reinforced through social isolation (Strayhorn, 2019). Therefore, providing a learning space that encourages student-to-student interactions (i.e., group tests) could reduce this feeling of isolation and increase one’s sense of belonging (Diekman et al., 2015; Good et al., 2012; Murphy et al., 2007; Walton & Cohen, 2007; Walton et al., 2012). While women generally have a lowered sense of belonging in mathematics fields and classrooms (Chavatzia, 2017; Ellis et al., 2016; Good et al., 2012), providing them with opportunities to engage with their peers while constructing knowledge and being assessed could improve their overall sense of belonging. This hypothesis, however, did not hold true for all of the women in this study. While the majority of the URM females did maintain their initial measured sense of belonging and their post-test reflections indicated they had beneficial collaborations with their group members, this was not true of the non-URM females.

The non-URM female students did also note the beneficial collaboration they felt they had with their group members during the group tests, but they had a lowered sense of belonging with respect to trust, affect, and acceptance. Additionally, the majority of the non-URM females scored higher on the second individual test (i.e., six out of the seven females) than on the second group test. These findings suggest that the nature of the group interactions in terms of collaboration alone are not enough for improving sense of belonging and performance.

The *trust* component of *MSoBS* decreased for the majority of the non-URM females. This component is centered around beliefs of the instructor and the testing materials. The main cause of the decrease for the non-URM females was due to the feeling that the test was biased. That is, the statement with the largest decrease for the non-URM females on the *MSoBS* survey was “I trust the testing materials to be unbiased.” Interpreting this decrease in trust alongside their feelings of “lack of control” during their group tests, might help explain their lower group test performance compared to their performance on the second individual test. That is, the non-URM females' decrease in trust could be related to their feeling about the test and instructor bias, whereas the lack of control focused on the frustration they felt with regards to their group interactions during the group test.

The *affect* subcomponent of *MSoBS* decreased for the majority of the non-URM females. This subcomponent refers to feeling comfortable, content, at ease, and as a member that “fits in” with the community. The feeling of the non-URM students of not being at ease (i.e., anxious, inadequate, tense, and nervous) may be explained by the frustration or lack of control they stated in their post-test reflections. Master and Meltzoff (2020) discussed how a sense of belonging could rest upon cues from a community, such as being excluded or “lack of control”, could threaten their sense of belonging. In the present study, the lack of control felt by the non-URM females suggests that their belonging was threatened in their group in terms of not being comfortable in the setting, therefore possibly explaining their higher individual test scores compared to group test scores.

The *acceptance* subcomponent of *MSoBS* decreased for three of the seven non-URM females and remained the same for three of the seven. This subcomponent refers to feeling as a valued, accepted, respected, and appreciated member of the female students’ mathematics community. For non-URM females, the decrease in their feelings of acceptance was expressed through their post-test reflections as not feeling heard and wanting more feedback from their peers. That is the non-URM females expressed a need for collaboration. The majority of the non-URM females were placed in a 2F-2M group. Therefore, with more males in the group, there may have been a possible presence of stereotype threat that was not explicitly investigated in the study. Consider, however, Thorson et al. (2019) who suggested that females who have experienced stereotype threat in the past (i.e., from being one of few women in mathematics) might have more productive collaborative interactions, especially with other females. The finding from Thorson et al. (2019), however, does not hold when considering the all-female group in this study. That is, the group had only female students, two of whom were non-URM and scored higher on the individual test. Although, similar to Thorson et al. (2019) the women in this group noted the benefit of their collaboration more than the women in the 2F-2M group, but also noted the largest amount of frustration (i.e., lack of control).

Additionally, many of the non-URM females did express a need for confidence. The women in the 2F-2M research group had the largest need for confidence out of the three research groups. This need though did not seem to impact the females' sense of belonging as neither of the females in this particular group changed from their pre- to post-*MSoBS* measure. Both of these women in this group had a high measure for their sense of belonging, yet still noted a large need for confidence. This finding is in contrast to Good et al. (2012) in which they suggested their measure provides evidence that the sense of belonging is a good predictor for one's confidence in mathematics.

Group Testing is not Equitable for All Students

As Leyva (2017) recently noted in their review of gender studies in mathematics education, there are few studies that examine performance and participation of students at the intersection of race and gender. While previous studies have shown an increase in student achievement, knowledge, or skills (Bookman & Friedman, 1994, 1999; Ganter & Jirovtek, 2000; Hurley et al., 1999; Meel, 1998; Penn, 1994; Roddick, 1993; Schwingendorf et al., 2000; Smith & Star, 2007; Williams, 1998), student attitudes and beliefs and affective gains (Bookman & Friedman, 1994, 1999; Laursen et al., 2014; Smith & Star, 2007), and performance and persistence (Bookman & Friedman, 1994, 1999; Hurley et al., 1999; Laursen et al., 2014; McDonald et al., 2000; Schwingendorf et al., 2000), recent studies have begun to explore the inequities in some collaborative learning environments (Johson et al., 2020; Maries et al., 2020). These studies, however, do not consider the gains, or lack thereof, of the students along the intersection of their race and gender.

The results from this dissertation contribute to this literature base by showing that the intersectional nature of students' identities can lead to fewer affective gains (i.e., sense of belonging) and decreases in performance (i.e., test scores). For example, even though student-centered activities such as group testing are considered as equitable (MacAurther, 2019); as seen in this study group testing in particular can still lead to non-URM females feeling as a lesser valued member in a mathematics community. The results of this study showed a difference in performance for students on the second group test based on the intersection of race and gender. The initial difference of students based on gender led to the discovery that non-URM women were the only group of students to *be harmed* from group testing (i.e., lowered their feelings of trust and affect). This finding is inconsistent with the equitable instruction literature and the group testing literature reviewed in Chapter 2.

Findings from the present study suggest that when females are purposefully placed into groups, race and gender should be considered a factor. The non-URM female students in this study did not feel as accepted valued members within their groups whereas the URM females did feel as accepted members. This was indicated by the female's post-test reflections. The feeling that one is an accepted member can affect their desire to maintain membership in that group (Good et al., 2012) and can impact their performance and perceptions of their mathematical abilities (Ellis et al., 2016). This suggests that when considering gender in mathematics classrooms, researchers and educators must be more cognizant to include the students' race (and possibly other identities; Leyva, 2017), as the non-URM women did not seem to have the same experience in the group setting as the URM women.

Note the intersection of gender and race is not the only important finding. There was also a difference in performance and participation based on group composition in this study, so the difference in the results may not only be based only on race and gender for the non-URM women. Recall each group composition contained a different number of men and women (i.e., all male,2F-2M, 3F-1M, and all female). Group composition can influence the dynamics, particularly group interactions, within a working group (Wiedmann et al., 2012). In particular, Spencer et al. (1999) and Woolley et al. (2010) suggest that the more women present in a group can positively influence the dynamics and performance of that group. In this dissertation, I found that groups with a different number of women (i.e., two, three, or four women) performed differently from each other group type (i.e., Group Test 2 scores), and each group exhibited different interactions among group members during the test.

Recall from Chapter 2, Good and colleagues (2012) stated “when sense of belonging is protected by learning environments that convey a malleable view of intelligence, students may be less vulnerable to the impact of negative stereotypes on achievement ...” (p. 714) suggesting that their feeling the test to be biased or having to constantly prove themselves (i.e., to the instructor or to their group members), the non-URM females may have been more vulnerable to the negative impacts associated or possible stereotype threat in their groups. Although stereotype threat was not explicitly investigated in this study[9], the placement of the females in their groups could have implicitly induced a stereotype threat due to the environment. The underperformance of the non-URM women in groups with men is consistent with the finding from Thorson et al. (2019). Recall from Chapter 2, Thorson and colleagues (2019) found that women working in pairs with a male partner performed worse than the men even if there was not an explicit threat of being stereotyped. This finding for the non-URM females also supports the initial research on stereotype threat by Spencer et al. (1999) who stated that women in environments where mathematical skills are exposed to judgement (i.e., a test) they can experience a stress response due to stereotype threat, and that stereotype threat can affect both women’s performance on mathematics tests (Cadaret et al., 2017) as well as influence women’s sense of belonging in mathematics (Good et al., 2012, Master &

---

[9] Researchers need to consider if it is possible to study stereotype threat on women in mathematics without explicitly inducing a stereotype threat condition on the women in the study.

Meltzoff, 2020). Finally, note the URM females did not change in their feelings of acceptance as evidenced by their *MSoBS* and post-test reflections. In particular, there were few instances of wishful collaboration in their post-test reflections. Recall that the URM females were placed into a 3F-1M group, thereby reducing the possibility of stereotype threat (i.e., Thorson et al., 2019) which could influence the women's sense of belonging (Diekman et al., 2015; Good et al., 2012; Dar-Nimrod & Heine, 2006; Good et al., 2008; Spencer et al., 1999), and the majority of the non-URM females were placed in groups 2F-2M groups.

While it is not possible to conduct a direct comparison between my study and those reported in Chapter 2 (as group composition was not reported), these findings contribute to the literature by suggesting that we should consider how we place students in groups. This is especially true for the women in our mathematics classrooms. Additionally, even though most teachers would agree with the statement "not all groups work the same" (Weaver et al., 2014) we must be aware of productive, disruptive, and equitable group interactions. In conclusion, the findings from this study suggest that student-centered activities, particularly group testing, are not equitable across all demographics with respect to performance and students' sense of belonging.

Implications and Recommendations for Future Research

Future research that investigates students' performance and participation while working in groups in mathematics classrooms should consider the composition of the groups and intersectionality of students' identities (Leyva, 2017). That is, group composition matters because it influences the interactions the students have in the groups which in turn can impact female students' sense of belonging to that group. We, as mathematics educators, should be cautious of using randomized groupings especially for our female students. Considering group composition is one attempt educators can make to help reduce possible stereotype threat as this is a threat to women's learning in mathematics (Diekman et al., 2015; Good et al., 2012; Dar-Nimrod & Heine, 2006; Good et al., 2008; Spencer et al., 1999). The social interactions and the negotiations of authority within a group, especially between the men and women, have the potential to induce stereotype threat on the women if they have ever experienced it in the past (Thorson et al., 2019), and hence lower their sense of belonging (Cech et al., 2011; Cheryan et al., 2009; Cheryan & Plaut, 2010; Cheryan et al., 2017; Good et al., 2012; Walton & Cohen, 2007). Finally, if we think about what randomization means in terms of group composition, we are acknowledging the fact that we would be okay if one or more students are at a disadvantage for the "greater good". This dissertation however, only provided a small amount of evidence against randomization of students into groups. Therefore, in the future, I plan to explore a quasi-experimental design study to compare randomization to purposefully placed groups for students in mathematics classes. Additionally, as the groups with a differing number of women performed differently in this study, I would like to explore more possible group compositions while considering the race and gender of all students. That is, I only had one non-URM female in a 3F-1M group which exhibited the most equitable talk around the mathematics.

In this study, there was some implicit evidence of a connection between the possibility of stereotype threat (i.e., need for confidence) and participatory performance measured by the group test for non-URM women. The possibility of stereotype threat being present might help to explain the lowered performance and need for confidence in the 2F-2M groups. Recall from chapter 2, stereotypes can negatively impact a woman's sense of belonging in mathematics (Good et al., 2012) which can impact their perceived ability and performance. If one does not feel as if their group belongs to a certain situation (i.e., women's lower mathematical sense of belonging) then the threats of stereotyping could additionally affect their performance and confidence in that situation (i.e., performance on mathematics tests). Therefore, we as educators need to be cognizant of the possibility of our women experiencing stereotype threat in our classrooms or in the groups in which they are placed.

Finally, the disheartening results for the non-URM women (i.e., white women) bring to light the need to consider the intersection of race and gender in our work in mathematics education. Recent studies on gender in mathematics have noted benefits for the women in the study (Laursen et al., 2014; MacArthur, 2019), and some researchers have noted there is still a gender gap (Johnson et al., 2020; Maries et al., 2020); however, few of the studies look at this intersection (i.e., race and gender; Leyva, 2017). Just as mathematics education has shifted from a focus of equality to equity (Fennema, 1979; Meyer, 1989), perhaps it is time to shift again from equity to culturally responsive pedagogy, by acknowledging each identity and culture that students bring to the classroom (Gay, 2018). This is from the discussion above, in which white women were harmed (i.e., decreased sense of belonging and lowered performance in a group) in what was traditionally thought of as an equitable learning opportunity. This result, however, may be linked to the group interactions that emerged from the 2F-2M group setting. That is, there could have been an implicit presence of stereotype threat in the 2F-2M groups, and most non-URM women were placed in this type of composed group. This could be the reason the lowered group performance, the expressed need for confidence, and the decreased sense of belonging were prominent in the results of this study. Again, more research on group compositions with larger sample sizes need to be conducted to make a stronger conclusion.

I would like to conclude with a statement from Leyva (2017), who stated “despite shifts in the conceptualization and empirical study of gender in mathematics education, intersections of gender with other dimensions of students’ identities generally remain minimally explored in analyses across achievement and participation studies” (p. 420). There are multiple studies about gender, race, and some studies specifically about URM women, yet we need to increase the number of studies looking at the intersection of race and gender for all students. I would, however, like to caution our community that when examining race and gender we should not let the whiteness overpower the female aspect of our non-URM women in mathematics. With a shift to a culturally responsive pedagogy, we need to acknowledge each part of a student’s identity and consider all aspects that they bring into the classroom (i.e., race, gender, sexual orientation, disabilities, etc.).

REFERENCES

Adler, J. (1997). A participatory-inquiry approach to the mediation of mathematics knowledge in a multilingual classroom. *Educational Studies in Mathematics, 33*, 235-258.

Amit, M., & Fried, M. N. (2005). Authority and authority relations in mathematics education: A view from an 8th grade classroom. *Educational Studies in Mathematics, 58*, 145-168.

Aydin, S., & Mutlu, C. (2013). Students' understanding of the concept of limit of a function in vocational high school mathematics. *The Online Journal of Science & Technology*, *3*(1), 145-152.

Banchefsky, S., Lewis, K. L., & Ito, T. A. (2019). The role of social and ability belonging in men's and women's pSTEM persistence. *Frontiers in Psychology*, *10*, 1-16.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117,* 497–529.

Beede, D. N., Julian, T. A., Langdon, D., McKittrick, G., Khan, B., & Doms, M. E. (2011). Women in STEM: A gender gap to innovation. *Economics and Statistics Administration Issue Brief*, (04-11). Retrieved from: https://files.eric.ed.gov/fulltext/ED523766.pdf

Bénéteau, C., Fox, G., Xiaoying Xu1, 6., Lewis, J. E., Ramachandran, K., Campbell, S., & Holcomb, J. (2016). Peer-led guided inquiry in calculus at the University of South Florida. *Journal of STEM Education: Innovations & Research, 17*(2), 5-13.

Berry, J., & Nyman, M. A. (2002). Small-group assessment methods in mathematics. International *Journal of Mathematical Education in Science and Technology, 33*, 641-649.

Bezuidenhout, J. (2001). Limits and continuity: Some conceptions of first-year students. *International Journal of Mathematical Education in Science & Technology, 32*(4), 487-500.

Bishop, J. P. (2012). "She's always been the smart one. I've always been the dumb one": Identities in the mathematics classroom. *Journal for Research in Mathematics Education, 43*(1), 34-74.

Blackburn, H. (2017). The status of women in STEM in higher education: A review of the literature 2007–2017. *Science & Technology Libraries, 36*, 235-273.

Blickenstaff, J. C. (2005). Women and science careers: Leaky pipeline or gender filter?. *Gender and Education, 17*, 369-386.

Blikstad-Balas, M. (2017). Key challenges of using video when investigating social practices in education: contextualization, magnification, and representation. *International Journal of Research & Method in Education, 40*, 511-523.

Boaler, J. (1998). Open and closed mathematics: Student experiences and understandings. *Journal of Research in Mathematics Education, 29,* 41-62.

Boaler , J., & Greeno, J.G. (2000). Identity, agency, and knowing in mathematical worlds. In J. Boaler (Ed.). *Multiple perspectives on Mathematics Teaching and Learning* (pp. 171 – 200). Westport, CT: Albex.

Boaler, J., & Staples, M. (2008). Creating mathematical futures through an equitable teaching approach: The case of Railside School. Teachers College Record, 110(3), 608-645.

Bookman, J., & Friedman, C. P. (1994). A comparison of the problem solving performance of students in lab based and traditional calculus. *Research in Collegiate Mathematics Education I, 4*, 101-116.

Bookman, J., & Friedman, C. P. (1999). The Evaluation of Project Calc at Duke University, 1989-1994. *MAA NOTES*, 253-256.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn* (Vol. 11). Washington, DC: National Academy Press.

Bressoud, D. M. (2011). The worst way to teach. *MAA Launchings, July*. (Accessed on September 29, 2016).

Bressoud, D. (2015). Insights from the MAA national study of college calculus. *Mathematics Teacher, 109*, 179-185.

Bressoud, D., Mesa, V., & Rasmussen, C. (Eds.). (2015). *Insights and recommendations from the MAA national study of college calculus*. Washington, DC: MAA Press.

Bressoud, D., & Rasmussen, C. (2015). Seven characteristics of successful calculus programs. *Notices of the American Mathematical Society, 62*, 144-146.

Brookstein, A., Hegedus, S., Dalton, S., Moniz, R., & Tapper, J. (2011). Measuring student attitude in mathematics classrooms. *Dartmouth: Kaput Center for Research and Innovation in STEM Education, University of Massachusetts.*

Bruner, J. (1997). Celebrating divergence: Piaget and Vygotsky. *Human development, 40*(2), 63-73.

Cadaret, M. C., Hartung, P. J., Subich, L. M., & Weigold, I. K. (2017). Stereotype threat as a barrier to women entering engineering careers. *Journal of Vocational Behavior, 99*, 40-51.

Cardwell, M. (2014). *Dictionary of psychology*. Routledge. Abingdon, United Kingdom.

Carter, J. D., Helliwell, D., Henrich, A., Principe, M., & Sloughter, J. M. (2016). Improving Student Success in Calculus at Seattle University. *PRIMUS, 26*(2), 105-124.

Carver, S., Van Sickle, J., Holcomb, J. P., Quinn, C. M., Jackson, D. K., Resnick, A. H., … & Marquard, A. M. (2017) Operation STEM: Increasing success and improving retention among first-generation and underrepresented minority students in STEM. *Journal of STEM Education, 18*, 30-39.

Cavanagh, M. (1996). Student understandings in differential calculus. In P. C. Clarkson (Ed.), *Technology in mathematics education* (Proceedings of the 19th annual conference of the Mathematics Education Research Group of Australasia, pp. 107-114). Melbourne: Mathematics Education Research Group of Australasia.

Cech, E. A., Rubineau, B., Silbey, S., & Seron, C. (2011). Professional role confidence and gendered persistence in engineering. *American Sociological Review, 76*, 641–666.

Chavatzia, T. (2017). Cracking the code: Girls' and women's education in science, technology, engineering and mathematics (STEM). Paris, France: UNESCO. Retrieved from http://unesdoc.unesco.org/images/0025/002534/253479E.pdf

Chiu, M. M. (2000). Group problem-solving processes: Social interactions and individual actions. *Journal for the Theory of Social Behaviour, 30*(1), 26-49.

Chen, X. (2013). STEM Attrition: College Students' Paths into and out of STEM Fields. Statistical Analysis Report. NCES 2014-001. *National Center for Education Statistics*.

Chen, X. (2015). STEM attrition among high-performing college students: scope and potential causes. *Journal of Technology and Science Education, 5,* 41-59.

Cheryan, S., & Plaut, V. C. (2010). Explaining underrepresentation: A theory of precluded interest. *Sex Roles, 63*(7-8), 475-488.

Cheryan, S., Plaut, V. C., Davies, P. G., & Steele, C. M. (2009). Ambient belonging: How stereotypical cues impact gender participation in computer science. *Journal of Personality and Social Psychology, 97*(6), 1045.

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin, 143*(1), 1–35.

Clarke, D., Breed, M., & Fraser, S. (2004). The consequences of a problem-based mathematics curriculum. *The Mathematics Educator*, *14(*2), 7-16.

Cobb, P. (1995). Mathematical learning and small-group interaction: Four case studies. In P. Cobb & H. Bauersfeld (Eds.), *The emergence of mathematical meaning: Interaction in classroom cultures* (pp. 25–129). Hillsdale, NJ: Erlbaum.

Cobb, P., & Bowers, J. (1999). Cognitive and situated learning perspectives in theory and practice. *Educational Researcher*, *28*(2), 4-15.

Cohen, E. (1994). *Collaborative learning: higher education, interdependence, and the authority of knowledge*. Baltimore: John Hopkins University Press.

Corbett, C., & Hill, C. (2015). *Solving the equation: The variables for women's success in engineering and computing.* American Association of University Women. 1111

Sixteenth Street NW, Washington, DC 20036. Retrieved from: https://files.eric.ed.gov/fulltext/ED580805.pdf

Cortright, R. N., Collins, H. L., Rodenbaugh, D. W., & DiCarlo, S. E. (2003). Student retention of course content is improved by collaborative-group testing. *Advances in Physiology Education, 27*(3), 102-108.

Cory, B. L., & Martin, T. (2018). Active vs. traditional learning in calculus I. In (Eds.) A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, and S. Brown, *Proceedings of the 21st Annual Conference on Research in Undergraduate Mathematics Education*, (pp. 1682-1683), San Diego, California.

Curless, J. M. (2012). *Relationships between adult learners' anxiety levels and cooperative exam scores.* [Doctoral dissertation, Walden University, Minneapolis, Minnesota.] ProQuest Dissertations Publishing.

Danaher, K., & Crandall, C. S. (2008). Stereotype threat in applied settings re-examined. *Journal of Applied Social Psychology, 38*, 1639 –1655.

Dar-Nimrod, I., & Heine, S. J. (2006). Exposure to scientific theories affects women's math performance. *Science, 314(*5798), 435-435.

DeHaan, R. L. (2005). The impending revolution in undergraduate science education. *Journal of Science Education and Technology, 14*(2), 253-269.

Dasgupta, N. (2011). Ingroup experts and peers as social vaccines who inoculate the self-concept: The stereotype inoculation model. *Psychological Inquiry, 22*(4), 231-246.

Diekman, A. B., Weisgram, E. S., & Belanger, A. L. (2015). New routes to recruiting and retaining women in STEM: Policy implications of a communal goal congruity perspective. *Social Issues and Policy Review*, *9*(1), 52-88.

Discourse (n.d.). In *Merriam-Webster's online dictionary* (11th ed.). Retrieved from https://www.merriam-webster.com/dictionary/discourse

Douglas, R. (Ed.). (1986), *Toward a lean and lively calculus: Report of the conference/workshop to develop curriculum and teaching methods for calculus at the college level* (MAA Notes No.6). Washington, DC: Mathematical Association of America.

Duane, B. T., & Satre, M. E. (2014). Utilizing constructivism learning theory in collaborative testing as a creative strategy to promote essential nursing skills. *Nurse Education Today*, *34*(1), 31-34.

Draskovic, I., Holdrinet, R., Bulte, J., Bolhuis, S., & Van Leeuwe, J. (2004). Modeling small group learning. *Instructional Science*, *32*(6), 447-473.

Eastridge, J. A. (2018). *Relationship between use of collaborative testing and nursing student success.* (Doctoral dissertation Concordia University-Portland, Portland, OR.) Retrieved from: https://commons.cu-portland.edu/cgi/viewcontent.cgi?article=1233&context=edudissertations

Edwards, D. (1993). But what do children really think? Discourse analysis and conceptual content in children's talk. *Cognition and Instruction*, *11*, 207-225.

Edwards, D., & Potter, J. (1992). *Discursive psychology*. Newbury Park, CA: Sage.

Ellis, M. W., & Berry III, R. Q. (2005). The paradigm shift in mathematics education: Explanations and implications of reforming conceptions of teaching and learning. *The Mathematics Educator, 15*(1), 7-17.

Ellis, J., Fosdick, B. K., & Rasmussen, C. (2016). Women 1.5 times more likely to leave STEM pipeline after calculus compared to men: Lack of mathematical confidence a potential culprit. *PLOS ONE, 11*(7), Retrieved from: https://doi.org/10.1371/journal.pone.0157447

Engage (n.d.). In *Merriam-Webster's online dictionary* (11th ed.). Retrieved from https://www.merriam-webster.com/dictionary/engage

Esmonde, I. (2009). Mathematics learning in groups: Analyzing equity in two cooperative activity structures. *The Journal of the Learning Sciences, 18*(2), 247-284.

Esmonde, I., & Langer-Osuna, J. M. (2013). Power in numbers: Student participation in mathematical discussions in heterogeneous spaces. *Journal for Research in Mathematics Education*, *44*(1), 288-315.

Fairweather, J. (2008). Linking evidence and promising practices in science, technology, engineering, and mathematics (STEM) undergraduate education. *Board of Science Education, National Research Council, The National Academies*, Washington, *DC.*

Farabaugh, R. (2007). 'The isle is full of noises': Using wiki software to establish a discourse community in a Shakespeare classroom. *Language Awareness, 16(1),* 41-56.

Fennema, E. (1979). Women and girls in mathematics—equity in mathematics education. *Educational Studies in Mathematics, 10*(4), 389-401.

Fernandez-Plaza, J. A., & Simpson, A. (2016). Three concepts or one? Students' understanding of basic limit concepts. *Educational Studies in Mathematics*, *60*(3), 315-332.

Ferrini-Mundy, J., & Guether-Graham, K. (1991). An overview of the calculus reform effort: issues for learning, teaching and curriculum development. *American Mathematical Monthly, 98*(7), 627–635.

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, *12*(2), 219-245.

Forman, E. A., Larreamendy-Joerns, J., Stein, M. K., & Brown, C. A. (1998). “You're going to want to find out which and prove it”: Collective argumentation in a mathematics classroom. *Learning and Instruction, 8*(6), 527-548.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, *111*(23), 8410-8415.

Ganter, S. L. (2001). *Changing calculus: A report on evaluation efforts and national impact from 1988–1998.* Washington, DC: Mathematical Association of America.

Ganter, S. L., & Jirovtek, M., R. (2000). The need for evaluation in the calculus reform movement: A comparison of two calculus teaching methods. In Dubinsky, E., Schoenfeld, A. H., & Kaput, J. J. (Eds.), *Research in Collegiate Mathematics*

Education IV (pp 16-41). Providence, RI: American Mathematical Society in cooperation with Mathematical Association of America.

Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice*. New York: Teachers College Press.

Gayles, J. G., & Ampaw, F. (2016). To Stay or Leave: Factors that impact undergraduate women's persistence in science majors. *NASPA Journal About Women in Higher Education, 9*, 133-151.

Gerasimova, D., Matson, K., Sachs, R., Hjalmarson, M. (2017). Conceptual understanding of differential calculus: A comparative study. In (Eds.) A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, and S. Brown, *Proceedings of the 20th Annual Conference on Research in Undergraduate Mathematics Education* (pp. 1517-1529), Pittsburgh, PA.

Gilley, B. H., & Clarkston, B. (2014). Collaborative testing: Evidence of learning in a controlled in-class study of undergraduate students. *Journal of College Science Teaching, 43*(3), 83-91.

Giuliodori, M. J., Lujan, H. L., & DiCarlo, S. E. (2008). Collaborative group testing benefits high-and low-performing students. *Advances in Physiology Education, 32*(4), 274-278.

Goetz, A. (2005). Using open-ended problems for assessment. *Mathematics Teacher*, *99*, 12-17.

Gonzalez, H. B., & Kuenzi, J. J. (2012). *Science, technology, engineering, and mathematics (STEM) education: A primer.* Washington, DC: Congressional Research Service. Retrieved from: https://fas.org/sgp/crs/misc/R42642.pdf

Good, C., Aronson, J., & Harder, J. A. (2008). Problems in the pipeline: Stereotype threat and women's achievement in high-level math courses. *Journal of Applied Developmental Psychology, 29*(1), 17-28.

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

Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools, 30*(1), 79-90.

Gutiérrez, R. (2013). The sociopolitical turn in mathematics education. *Journal for Research in Mathematics Education, 44*(1), 37-68.

Gutiérrez, R. (2018). When mathematics teacher educators come under attack. *Mathematics Teacher Educator, 6*(2), 68-74.

Hagman, J. E. (2021). The eighth characteristic for successful calculus programs: Diversity, equity, & inclusion practices. *Primus, 31*(1), 70-90.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112.

Harre, R., & Gillet, G. (1995). *The discursive mind.* Thousand Oaks, CA: Sage.

Hassi, M. L., & Laursen, S. L. (2015). Transformative learning: Personal empowerment in learning mathematics. *Journal of Transformative Education, 13*(4), 316-340.

Heid, M. K. (1988). Resequencing skills and concepts in applied calculus using the computer as a tool. *Journal for Research in Mathematics Education, 19*, 3-25.

Hewlett, S. A., Sherbin, L., Dieudonne, F., Fargnoli, C., & Fredman, C. (2014). *Athena factor 2.0: Accelerating female talent in science, engineering & technology.* Executive Summary, New York: Center for Talent Innovation.

Hickernell, F., & Proskurowski, W. (1985). The use of microcomputers in the teaching of calculus. *Collegiate Microcomputer, 3*(2), 111-25.

Hill, C., Corbett, C., & St Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics*. American Association of University Women. 1111 Sixteenth Street NW, Washington, DC 20036. Retrieved from: https://files.eric.ed.gov/fulltext/ED509653.pdf

Honey, M., Pearson, G., & Schweingruber, H. (2014). *STEM Integration in K–12 Education.* Washington, DC: National Academies Press.

Hurley, J. F., Koehn, U., & Ganter, S. L. (1999). Effects of calculus reform: Local and national. *The American Mathematical Monthly, 106(*9), 800-811.

Ilhan, K., Bulent, G., & Erdem, C. (2011). A cross-age study of students' understanding of limit and continuity concepts. *Bolema : Boletim De Educação Matemática, 24*(38), 245-332.

Jang, H., Lasry, N., Miller, K., & Mazur, E. (2017). Collaborative exams: Cheating? Or learning?. *American Journal of Physics, 85*(3), 223-227.

Johnson, E., Andrews-Larson, C., Keene, K., Melhuish, K., Keller, R., & Fortune, N. (2020). Inquiry and gender inequity in the undergraduate mathematics classroom. *Journal for Research in Mathematics Education, 51*(4), 504-516.

Johnson, D. W., & Johnson, R. T. (1999). Making cooperative learning work. *Theory into Practice, 38*(2), 67-73.

Johnson, D. W., & Johnson, R. T. (2004). *Assessing students in groups: Promoting group responsibility and individual accountability.* Corwin Press. Thousand Oaks, California.

Joiner, K. F., Malone, J. A., & Haimes, D. H. (2002). Assessment of classroom environments in reformed calculus education. *Learning Environments Research, 5*(1), 51-76.

Joseph, N. M., Hailu, M., & Boston, D. (2017). Black women's and girls' persistence in the P–20 mathematics pipeline: Two decades of children, youth, and adult education research. Review of Research in Education, 41(1), 203-227.

Judson, P. T. (1988). *Effects of modified sequencing of skills and applications in introductory calculus* (Doctoral dissertation). University of Texas at Austin, Austin, Texas.

Judson, P.T. (1990). Elementry business calculus with computer algebra. *Journal of Mathematical Behavior, 9*, 153-157.

Kabael, T. (2014). Students' formalising process of the limit concept. *Australian Senior Mathematics Journal, 28*(2), 23-38.

Kapitanoff, S., & Pandey, C. (2018). Collaborative testing in statistics: Group interaction, anxiety, and class performance. *Statistics Education Research Journal, 17*(2), 51-67.

Kidron, I. (2015). Is small, small enough? Students' understanding the need for the definition of the derivative as a limit. *International Journal for Technology in Mathematics Education, 22*(1), 31-41.

King, P. E., & Behnke, R. R. (2005). Problems associated with evaluating student performance in groups. *College Teaching, 53*(2), 57-61.

Kearns, J., & Bolyard, J. (2018). Assessing group learning opportunities in a first semester calculus course. In (Eds.) A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, and S. Brown, *Proceedings of the 21th Annual Conference on Research in Undergraduate Mathematics Education* (pp. 1400-1406), San Diego, California.

Kogan, M., & Laursen, S. L. (2014). Assessing long-term effects of inquiry-based learning: A case study from college mathematics. *Innovative Higher Education, 39*(3), 183-199.

Kozulin, A. (1990). The concept of regression and Vygotskian developmental theory. *Developmental Review, 10*(2), 218-238.

Kuh, G. D., Cruce, T. M., Shoup, R., Kinzie, J., & Gonyea, R. M. (2008). Unmasking the effects of student engagement on first-year college grades and persistence. *The Journal of Higher Education, 79*(5), 540-563.

Lampert, M. (1990). When the problem is not the question and the solution is not the answer: Mathematical knowing and teaching. *American Educational Research Journal, 27*(1), 29-63.

Langer-Osuna, J. M. (2011). How Brianna became bossy and Kofi came out smart: Understanding the trajectories of identity and engagement for two group leaders in a project-based mathematics classroom. *Canadian Journal of Science, Mathematics and Technology Education, 11*(3), 207-225.

Langer-Osuna, J. M. (2016). The social construction of authority among peers and its implications for collaborative mathematics problem solving. *Mathematical Thinking and Learning, 18*, 107-124.

Laursen, S., Hassi, M. L., Kogan, M., Hunter, A. B., & Weston, T. (2011). Evaluation of the IBL mathematics project: Student and instructor outcomes of inquiry-based learning in college mathematics. Boulder, CO: *University of Colorado Boulder, Ethnography & Evaluation Research.*

Laursen, S. L., Hassi, M. L., Kogan, M., & Weston, T. J. (2014). Benefits for women and men of inquiry-based learning in college mathematics: A multi-institution study. *Journal for Research in Mathematics Education, 45*(4), 406–418.

Laursen, S. L., & Rasmussen, C. (2019). I on the prize: Inquiry approaches in undergraduate mathematics. *International Journal of Research in Undergraduate Mathematics Education*, 5(1), 129-146.

Lave, J., & Wenger, E. (1991) *Situated learning: Legitimate peripheral participation.* Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511815355

Le Roux, C. J. (2011). *A critical examination of the use of practical problems and a learner-centred pedagogy in a foundational undergraduate mathematics course* (Unpublished doctoral dissertation). University of the Witwatersrand, Johannesburg, South Africa.

Lewis, K. L., & Hodges, S. D. (2015). Expanding the concept of belonging in academic domains: Development and validation of the Ability Uncertainty Scale. *Learning and Individual Differences, 37*, 197-202.

Lewis, K. L., Stout, J. G., Finkelstein, N. D., Pollock, S. J., Miyake, A., Cohen, G. L., & Ito, T. A. (2017). Fitting in to move forward: Belonging, gender, and persistence in the physical sciences, technology, engineering, and mathematics (pSTEM). *Psychology of Women Quarterly, 41*(4), 420-436.

Leyva, L. A. (2017). Unpacking the male superiority myth and masculinization of mathematics at the intersections: A review of research on gender in mathematics education. *Journal for Research in Mathematics Education, 48*(4), 397-433.

Lithner, J. (2004). Mathematical reasoning in calculus textbook exercises. *The Journal of Mathematical Behavior, 23*(4), 405-427.

London, B., Rosenthal, L., Levy, S. R., & Lobel, M. (2011). The influences of perceived identity compatibility and social support on women in nontraditional fields during the college transition. *Basic and Applied Social Psychology, 33*(4), 304-321.

Lovric, M., & Burazin, A. (2018). Analyzing narratives about limits involving infinity in calculus textbooks. In (Eds.) A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, and S. Brown, *Proceedings of the 21st Annual Conference on Research in Undergraduate Mathematics Education* (pp. 474-483), San Diego, California.

Lusk, M., & Conklin, L. (2003). Collaborative testing to promote learning. *Journal of Nursing Education, 42*(3), 121-124.

M Russell. (2018, December 25). Do my classroom rules help or hurt learning? [Web log post]. Retrieved from: https://www.middleweb.com/39369/do-my-classroom-rules-help-or-hurt-learning/

M Russell. (2019, February 19). Experimenting with math testing in groups [Web log post]. Retrieved from: https://www.middleweb.com/39734/experimenting-with-a-group-test-in-math-class/

MacArthur, K. (2019). Re-humanizing assessments in university calculus II courses. Preliminary Report in Weinberg, A., Moore-Russo, D., Soto, H., & Wawro, M. (Eds.). *Proceedings of the 22nd Annual Conference on Research in Undergraduate Mathematics Education* (pp. 953-958), Oklahoma City, Oklahoma.

Maries, A., Karim, N. I., & Singh, C. (2020). Active learning in an inequitable learning environment can increase the gender performance gap: The negative impact of stereotype threat. *The Physics Teacher, 58*(6), 430-433.

Marx, D. M., Stapel, D. A., & Muller, D. (2005). We can do it: the interplay of construal orientation and social comparisons under threat. *Journal of personality and social psychology, 88*(3), 432-446.

Master, A., Cheryan, S., & Meltzoff, A. N. (2016). Computing whether she belongs: Stereotypes undermine girls' interest and sense of belonging in computer science. *Journal of educational psychology, 108(*3), 424.

Master, A., & Meltzoff, A. N. (2020). Cultural stereotypes and sense of belonging contribute to gender gaps in STEM. *International Journal of Gender, Science and Technology, 12*(1), 152-198.

McDonald, M. A., Mathews, D. M., & Strobel, K. (2000). Understanding sequences: A tale of two objects. In Dubinsky, E., Schoenfeld, A. H., & Kaput, J. J. (Eds.), *Research in Collegiate Mathematics Education IV* (pp 77-102). Providence, RI:

American Mathematical Society in cooperation with Mathematical Association of America

McGarvey, C. (1981). Mathematics and miseducation: toward the next school curricula. *The Mathematics Teacher*, *74*(2), 90-95.

McLeod, D. B. (1994). Research on affect and mathematics learning in the JRME: 1970 to the Present. *Journal for research in Mathematics Education, 25*, 637-647.

Meel, D. E. (1998). Honors students' calculus understandings: Comparing Calculus & Mathematica and traditional calculus students. In Dubinsky, E., Schoenfeld, A. H., & Kaput, J. J. (Eds.), *Research in Collegiate Mathematics Education III* (pp.163-215). Providence RI: American Mathematical Society in cooperation with Mathematical Association of America.

Meyer, M. R. (1989). Equity: The missing element in recent agendas for mathematics education. *Peabody Journal of Education, 66*(2), 6-21.

Micari, M., Pazos, P., Streitwieser, B., & Light, G. (2010). Small-group learning in undergraduate stem disciplines: Effect of group type on student achievement. *Educational Research and Evaluation, 16*, 269-286.

Murphy, M. C., Steele, C. M., & Gross, J. J. (2007). Signaling threat: How situational cues affect women in math, science, and engineering settings. *Psychological Science*, *18*, 879-885.

National Research Council. (1993). *Measuring Up: Prototypes for Mathematics Assessment*. Washington, DC: The National Academies Press. Retrieved from: https://doi.org/10.17226/2071

National Research Council. (2012). A framework for K-12 science education: *Practices, crosscutting concepts, and core ideas*. National Academies Press.

National Science Foundation, Division of Science Resources Statistics. (2017). *Women, Minorities, and Persons with Disabilities in Science and Engineering: 2017*. Nsf 09-305. Retrieved from: https://www.nsf.gov/statistics/2017/nsf17310/static/downloads/nsf17310-digest.pdf

O'Brien, L. T., Garcia, D. M., Adams, G., Villalobos, J. G., Hammer, E., & Gilbert, P. (2015). The threat of sexism in a STEM educational setting: the moderating impacts of ethnicity and legitimacy beliefs on test performance. *Social Psychology of Education*, *18*, 667-684.

O'Connor, M. C., & Michaels, S. (1996). Shifting participant frameworks: Orchestrating thinking practices in group discussion. *Discourse, Learning, and Schooling*, 63-103.

Ogden, L., Kearns, J., & Bowman, N. (2018). Prerequisite knowledge of mathematics and success in Calculus I. In (Eds.) A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, and S. Brown, *Proceedings of the 21st Annual Conference on Research in Undergraduate Mathematics Education* (pp. 1325-1330), San Diego, California.

Ohland, M. W., Sheppard, S. D., Lichtenstein, G., Eris, O., Chachra, D., & Layton, R. A. (2008). Persistence, engagement, and migration in engineering programs. *Journal of Engineering Education*, *97*(3), 259-278.

Osen, L. M. (1974). *Women in mathematics*. Mit Press. Cambridge, Massachusetts

Osterman, K. F. (2000). Students' need for belonging in the school community. *Review of Educational Research, 70*(3), 323-367.

Paavola, S., & Hakkarainen, K. (2005). The knowledge creation metaphor–An emergent epistemological approach to learning. *Science & Education*, *14*(6), 535-557.

Palmiter, J.R. (1991). Effects of computer algebra systems on concept and skill acquisition in calculus. *Journal for Research in Mathematics Education. 22*(2), 151-156.

Pandey, C., & Kapitanoff, S. (2011). The influence of anxiety and quality of interaction on collaborative test performance. *Active Learning in Higher Education*, *12*(3), 163-174.

Paterson, J., Sheryn, L., & Sneddon, J. (2013). Student responses to team-based learning in tertiary mathematics courses. In (Eds.) S. Brown, G. Karakok, K. H. Roh, M. Oehrtman, *Proceedings of 15tht Annual Conference on Research in Undergraduate Mathematics Education* (Vol. 2, pp. 619-626), Denver, Colorado.

Patton, M. Q. (2015). *Qualitative research and evaluation methods*. Newbury Park, CA: Sage.

Pedersen, S., & Liu, M. (2003). Teachers' beliefs about issues in the implementation of a student-centered learning environment. *Educational Technology Research and Development*, *51*(2), 57-76.

Penn, H. (1994). Comparisons of test scores in Calculus I at the Naval Academy. Focus on Calculus, *a Newsletter for the Calculus Consortium Based at Harvard University*. Wiley (6), 6-7.

Perry, Ayanna Dara–Franklin. (2013) *Equitable spaces in early career teachers' mathematics classrooms*. (Doctoral dissertation, North Carolina State University, North Carolina).

Perry, A. (2018). 7 Features of equitable classroom spaces. *Mathematics Teacher*, *112*, 186-191.

Pettersson, K., & Scheja, M. (2008). Algorithmic contexts and learning potentiality: A case study of students' understanding of calculus. *International Journal of Mathematical Education In Science & Technology*, *39*(6), 767-784.

Piaget, J. (1953). How children form mathematical concepts. *Scientific American*, *189*(5), 74-79.

President's Council of Advisors on Science and Technology (2012). *Engage to Excel: Producing One Million Additional College Graduates with Degrees in Science, Technology, Engineering, and Mathematics*. Report to the President. Retrieved from: https://files.eric.ed.gov/fulltext/ED541511.pdf

Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. *Journal of Experimental Social Psychology*, *40*(2), 152-168.

Quinn, C. M. (2018, May). *Connecting calculus to the real world through a fast, fun, and furious problem.* Poster presented at the Inquiry-Based Learning and Teaching Conference. Austin, TX.

Quinn, C. M., Reid, J. W., & Gardner, G. E. (2020). S+ T+ M= E as a convergent model for the nature of STEM. *Science & Education, 29*, 881-898.

Rasmussen, C., & Ellis, J. (2013). Who is switching out of calculus and why? In A. M. Lindmeier & A. Heinze (Eds.), *Proceedings of the 37th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 73–80). Kiel, Germany: PME

Rasmussen, C., Kwon, O. N., Allen, K., Marrongelle, K., & Burtch, M. (2006). Capitalizing on advances in mathematics and K-12 mathematics education in undergraduate mathematics: An inquiry-oriented approach to differential equations. *Asia Pacific Education Review*, *7*(1), 85-93.

Rao, S. P., Collins, H. L., & DiCarlo, S. E. (2002). Collaborative testing enhances student learning. *Advances in Physiology Education*, *26*(1), 37-41.

Reinholz, D. L. (2017). Not-so-Critical Friends: Graduate Student Instructors and Peer Feedback. *International Journal for the Scholarship of Teaching and Learning*, *11*(2), 1-9.

Revere, L., Elden, M., & Bartsch, R. (2008). Designing group examinations to decrease social loafing and increase learning. *International Journal for the Scholarship of Teaching and Learning*, *2(*17), 1-16. DOI: https://doi.org/10.20429/ijsotl.2008.020117

Roberts, T., Jackson, C., Mohr-Schroeder, M. J., Bush, S. B., Maiorca, C., Cavalcanti, M., ... & Cremeans, C. (2018). Students' perceptions of STEM learning after participating in a summer informal learning experience. *International journal of STEM education*, *5*(1), 1-14.

Roddick, C. (2003). Calculus reform and traditional students' use of calculus in an engineering mechanics course. In Selden, A., Dubinsky, E., Harel, G., & Hitt, F.

(Eds.), *Research in Collegiate Mathematics Education V*, (pp 56-78). Providence, RI: American Mathematical Society.

Roediger III, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: Applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition*, *1*(4), 242-248.

Roh, K. H. (2008). Students' images and their understanding of definitions of the limit of a sequence. *Educational Studies in Mathematics*, *69*(3), 217-233.

Sanabria, T., & Penner, A. (2017). Weeded out? Gendered responses to failing calculus. *Social Sciences*, *6*(2), 1-14.

Sandahl, S. S. (2010). Collaborative testing as a learning strategy in nursing education. *Nursing Education Perspectives*, *31*(3), 142-147.

Saxe, K., Braddy, L., Bailer, J., Farinelli, R., Holm, T., Mesa, V., ... & Turner, P. (2015). *A common vision for undergraduate mathematical sciences programs in 2025*. Washington, DC: Mathematical Association of America. Retrieved from: http://www.maa.org/sites/default/files/pdf/ingenious/INGenIOuS-report.pdf

Scheja, M., & Pettersson, K. (2010). Transformation and contextualisation: Conceptualising students' conceptual understandings of threshold concepts in calculus. *Higher Education*, *59*, 221-241.

Schoenfeld, A. H. (1996). In fostering communities of inquiry, must it matter that the teacher knows “the answer"?. *For the Learning of Mathematics*, *16*(3), 11-16.

Schwingendorf, K. E., McCabe, G. P., & Kuhn, J. (2000). A longitudinal study of the C4 L calculus reform program: Comparisons of C4 L and traditional students. In Dubinsky, E., Schoenfeld, A., & Kaut, K. (Eds.), R*esearch in Collegiate*

Mathematics Education IV (pp 63- 102). Providence, RI: American Mathematical Society.

Seymour, E., & Hewitt, N. M. (1997) *Talking about leaving: Why undergraduates leave the sciences.* Boulder, CO: Westview.

Seymour, E., Wiese, D., Hunter, A., & Daffinrud, S. M. (2000, March). Creating a better mousetrap: On-line student assessment of their learning gains. In *National Meeting of the American Chemical Society* (pp. 1-40).

Sfard, A. (1998). On two metaphors for learning and the dangers of choosing just one. *Educational Researcher*, *27*(2), 4-13.

Sfard, A. (2003). Balancing the unbalanceable: The NCTM Standards in light of theories of learning. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A Research Companion to Principles and Standards for School Mathematics* (pp. 353–392). Reston, VA: National Council of Teachers of Mathematics.

Shapiro, J. R., & Williams, A. M. (2012). The role of stereotype threats in undermining girls' and women's performance and interest in STEM fields. *Sex Roles, 66*(3-4), 175-183.

Shapiro, M., Grossman, D., Carter, S., Martin, K., Deyton, P., & Hammer, D. (2015). Middle school girls and the "Leaky Pipeline" to leadership: An examination of how socialized gendered roles influences the college and career aspirations of girls is shared as well as the role of middle level professionals in disrupting the influence of social gendered messages and stigmas. *Middle School Journal*, *46*(5), 3-13.

Slavin, R. E., Lake, C., & Groff, C. (2009). Effective programs in middle and high school mathematics: A best-evidence synthesis. *Review of Educational Research*, *79*(2), 839-911.

Smith, T.J., McKenna, C.M., & Hines, E. (2014). Association of group learning with mathematics achievement and mathematics attitude among eighth-grade students in the US. *Learning Environment Research, 17*, 229-241.

Smith III, J., & Star, J. (2007). Expanding the notion of impact of K-12 standards-based mathematics and reform calculus programs. *Journal for Research in Mathematics Education*, *38*(1), 3-34.

Solomon, Y. (2007). Not belonging? What makes a functional learner identity in undergraduate mathematics?. *Studies in Higher Education*, *32*(1), 79-96.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*, 4-28.

Springer, L, Stanne, M.E., & Donovan, S.S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology: A meta-analysis. *Review of Educational Research*, *69*(1), 21-51.

Srougi, M. C., Miller, H. B., Witherow, D. S., & Carson, S. (2013). Assessment of a novel group-centered testing schema in an upper-level undergraduate molecular biotechnology course. *Biochemistry and Molecular Biology Education*, *41*(4), 232-241.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*(6), 613-629.

Stewart, S., Reeder S., Raymond, K., Troup, J. (2018). Could algebra be the root of problems in calculus courses?. In (Eds.) A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, and S. Brown, *Proceedings of the 21st Annual Conference on Research in Undergraduate Mathematics Education* (pp. 1325-1330), San Diego, California.

Stout, J. G., & Blaney, J. M. (2017). "But it doesn't come naturally": how effort expenditure shapes the benefit of growth mindset on women's sense of intellectual belonging in computing. *Computer Science Education, 27*(3-4), 215-228.

Strayhorn, T. L. (2019). *College students' sense of belonging: A key to educational success for all students.* Routledge. New York, NY.

Tall, D. O. (1986). *Building and testing a cognitive approach to the calculus using interactive computer graphics* (Doctoral dissertation, University of Warwick, Coventry, England.) Retrieved from: http://wrap.warwick.ac.uk/2409/

Thorson, K. R., Forbes, C. E., Magerman, A. B., & West, T. V. (2019). Under threat but engaged: Stereotype threat leads women to engage with female but not male partners in math. *Contemporary Educational Psychology*, *58*, 243-259.

U.S. Department of Education, Office of Innovation and Improvement. (2016). *STEM 2026: A Vision for Innovation in STEM Education.* Washington, DC: Author. Retrieved from: https://innovation.ed.gov/what-we-do/stem/

U.S. Department of the Interior STEM Education and Employment Pathways. (2013). *Strategic Plan Fiscal Years 2013-2018*. Retrieved from:

https://nctc.fws.gov/programs/education-outreach/DOI-STEM-Strategic-Plan-2013-2018.pdf

VandenBos, G. R. (2007). *APA dictionary of psychology*. American Psychological Association.

Van Sickle, J., Schuler, K. R., Holcomb, J. P., Carver, C. D., Resnick, A., Quinn, C., ..., & Sridhar, N. (2020). Closing the achievement gap for underrepresented minority students in STEM: A deep look at a comprehensive intervention. *Journal of STEM Education Innovation and Research*, *21*(2), 5-18.

Vogler, J. S., & Robinson, D. H. (2016). Team-based testing improves individual learning. *The Journal of Experimental Education*, *84*(4), 787-803.

Vogt, C. M., Hocevar, D., & Hagedorn, L. S. (2007). A social cognitive construct validation: Determining women's and men's success in engineering programs. *The Journal of Higher Education*, *78*, 337-364.

Vygotsky, L. (1962). *Thought and language*. Cambridge, MA: MIT Press.

Walczyk, J. J., Ramsey, L. L., & Zha, P. (2007). Obstacles to instructional innovation according to college science and mathematics faculty. *Journal of Research in Science Teaching*, *44*(1), 85-106.

Walshaw, M., & Anthony, G. (2008). The teacher’s role in classroom discourse: A review of recent research into mathematics classrooms. *Review of Educational Research*, *78*(3), 516-551.

Walton, G. M., & Carr, P. B. (2012). Social belonging and the motivation and intellectual achievement of negatively stereotyped students. In M. Inzlicht & T. Schmader

(Eds.), *Stereotype threat: Theory, processes, and application* (pp. 89 –106). New York, NY: Oxford University Press.

Walton, G. M., & Cohen, G. L. (2007). A question of belonging: Race, social fit, and achievement. *Journal of Personality and Social Psychology*, *92,* 82–96. doi:10.1037/0022-3514.92.1.82

Walton, G. M., Cohen, G. L., Cwir, D., & Spencer, S. J. (2012). Mere belonging: The power of social connections. *Journal of Personality and Social Psychology, 102*, 513-532.

Weaver, C. L., Duran, E. C., & Nikles, J. A. (2014). An integrated approach for development of scientific writing skills in undergraduate organic lab. *Addressing the Millennial Student in Undergraduate Chemistry*, *1180*, 105-123.

Webb, N. M. (1995). Group collaboration in assessment: Multiple objectives, processes, and outcomes. *Educational Evaluation and Policy Analysis, 17*, 239-261.

White, P., & Mitchelmore, M. (1996). Conceptual knowledge in introductory calculus. *Journal for Research in Mathematics Education, 27*(1), 79–95.

Wiedmann, M., Leach, R. C., Rummel, N., & Wiley, J. (2012). Does group composition affect learning by invention?. *Instructional Science*, *40*, 711-730.

Williams, C. G. (1998). Using concept maps to assess conceptual knowledge of function. *Journal for Research in Mathematics Educatio*n, 29(4), 414-421

Wood, M. B. (2013). Mathematical micro-identities: Moment-to-moment positioning and learning in a fourth-grade classroom. *Journal for Research in Mathematics Education, 44*(5), 775–808. doi:10.5951/jresematheduc.44.5.0775

Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, *330*, 686-688.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, *27*(4), 458-477.

Young, C. Y., Georgiopoulos, M., Hagen, S. C., Geiger, C. L., Dagley-Falls, M. A., Islas, ... & Bradbury, E. E. (2011). Improving student learning in Calculus through applications. *International Journal of Mathematical Education in Science & Technology*, *42*, 591-604.

APPENDICES

Appendix A: Math Sense of Belonging Scale

Today I have some questions I would like you to answer about your experience with mathematics courses and in the mathematics community. When I mention the mathematics community, I am referring to the students in a mathematics course and those you are paired with when working in groups. I would like you to consider your membership in the mathematics community. By considering your experiences in your mathematics courses, both in high school and at Cleveland State University. Please respond to the following statements based on how you feel about these groups and your membership in it. There are no right or wrong answers to any of these statements; I am interested in your honest reactions and opinions. Please read each statement carefully and indicate the number that reflects your degree of agreement.

Strongly Disagree Strongly Agree

1 2 3 4 5 6 7 8

When I am in a math setting:

1. I feel that I belong to the math community.
2. I consider myself a member of the math world.
3. I feel like I am part of the math community.
4. I feel a connection with the math community.
5. I feel like an outsider.
6. I feel accepted.
7. I feel respected.
8. I feel disregarded.
9. I feel valued.
10. I feel neglected.
11. I feel appreciated.
12. I feel excluded.
13. I feel like I fit in.
14. I feel insignificant.
15. I feel at ease.
16. I feel anxious.
17. I feel comfortable.
18. I feel tense.
19. I feel nervous.
20. I feel content.
21. I feel calm.
22. I feel inadequate.
23. I wish I could fade into the background and not be noticed.
24. I try to say as little as possible.
25. I enjoy being an active participant.
26. I wish I were invisible.
27. I trust the testing materials to be unbiased.
28. I have trust that I do not have to constantly prove myself.
29. I trust my instructors to be committed to helping me learn.
30. Even when I do poorly, I trust my instructors to have faith in my potential.

Appendix B: Working Group Survey

I would like to gather some information from each of you about how you feel when working in a mathematics class. This survey will help determine the best learning environment we can create for everyone to succeed and set the permanent groups for the remainder of the semester. Please fill out the form below to help me determine the groups you will be placed in for the group tests and the group work atmosphere for the summer. I may consider allowing you to take the second group test by yourself, however, I reserve the right to place you into groups that I feel are most beneficial to your learning and your performance. The survey consists of 12 Likert-Scale questions and 4 open-response questions. Please answer each question honest using the given scale for the question. There are no right or wrong answers to any of these statements; I am interested in your honest reactions and opinions. Please read each statement carefully, and indicate the number that reflects your degree of agreement.

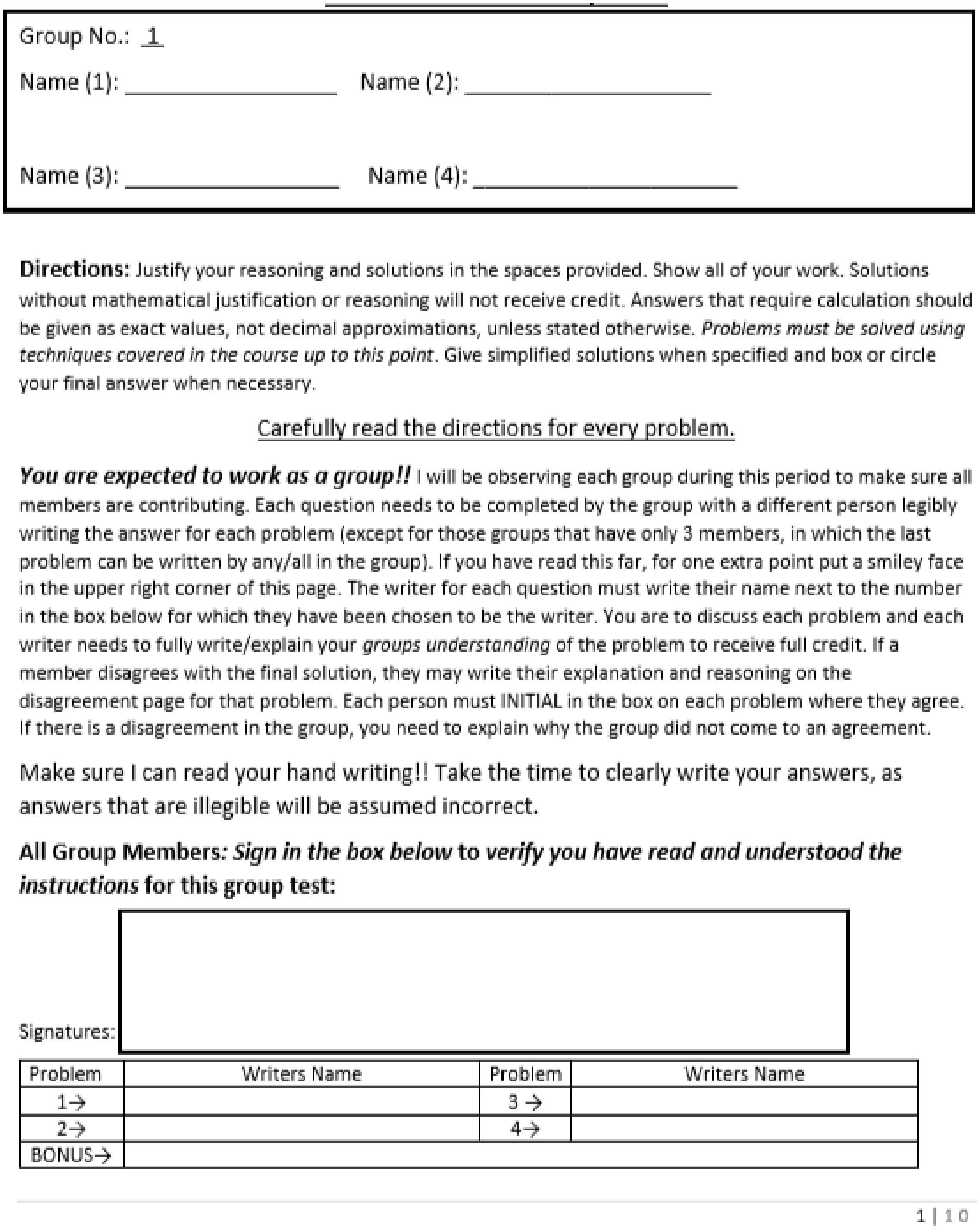
Strongly Disagree					Strongly Agree
1	2	3	4	5	6

1. I feel comfortable asking my peers for help.
2. I feel comfortable asking my instructor for help.
3. I feel comfortable talking about my ideas in class.
4. I feel confident in my algebra skills.
5. I feel confident in my mathematical understanding.
6. I feel confident that I have the right answer when I solve a mathematics problem.
7. I enjoy working in groups more than working alone in class.
8. I enjoy hearing the thoughts and ideas of my fellow students in class.
9. I prefer working alone rather than in groups when doing math.
10. I feel comfortable speaking in front of my peers.
11. I feel comfortable speaking in front of the class.
12. I feel my peers carefully listen to what I have to say.

For the following five open response questions please provide as much detail in your response as you can.

1. Suppose you got one answer to a problem and your partner(s) got a different answer, how do you determine who has the correct answer?
2. What do you perceive your job as when solving math problems in a group?
3. Have you been working with someone already that you think you learn well with and would like to keep working with? If so, please list their name(s) and explain why you work well together.
4. If you would prefer to work alone, please describe the strategies you use how to learn mathematics, and how working in a group would prevent you from using these strategies. Additionally, describe how completing the work by yourself will benefit you in your future studies and career?
5. If you would prefer to work in a group, please describe the strategies you use how to learn mathematics, and how working in alone would prevent you from using these strategies. Additionally, describe how completing the work in a group will benefit you in your future studies and career?

Appendix C: Group Test 1



Group No.: 1

Name (1): ________________ Name (2): __________________

Name (3): ________________ Name (4): ___________________

Directions: Justify your reasoning and solutions in the spaces provided. Show all of your work. Solutions without mathematical justification or reasoning will not receive credit. Answers that require calculation should be given as exact values, not decimal approximations, unless stated otherwise. *Problems must be solved using techniques covered in the course up to this point*. Give simplified solutions when specified and box or circle your final answer when necessary.

Carefully read the directions for every problem.

You are expected to work as a group!! I will be observing each group during this period to make sure all members are contributing. Each question needs to be completed by the group with a different person legibly writing the answer for each problem (except for those groups that have only 3 members, in which the last problem can be written by any/all in the group). If you have read this far, for one extra point put a smiley face in the upper right corner of this page. The writer for each question must write their name next to the number in the box below for which they have been chosen to be the writer. You are to discuss each problem and each writer needs to fully write/explain your *groups understanding* of the problem to receive full credit. If a member disagrees with the final solution, they may write their explanation and reasoning on the disagreement page for that problem. Each person must INITIAL in the box on each problem where they agree. If there is a disagreement in the group, you need to explain why the group did not come to an agreement.

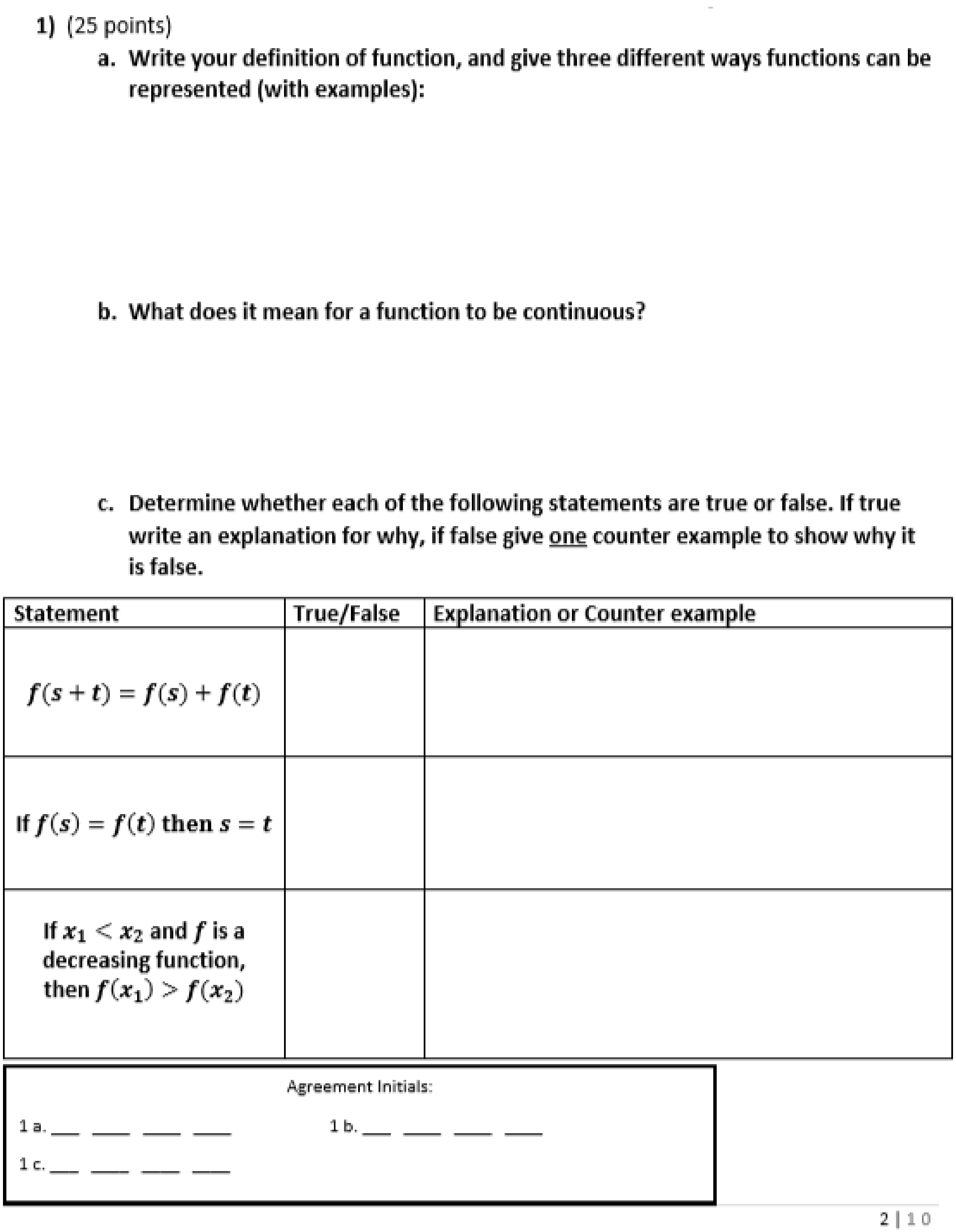
Make sure I can read your hand writing!! Take the time to clearly write your answers, as answers that are illegible will be assumed incorrect.

All Group Members*: Sign in the box below* to *verify you have read and understood the instructions* for this group test:

Signatures:

Problem	Writers Name	Problem	Writers Name
1→		3 →	
2→		4→	
BONUS→			

1 | 10



1) (25 points)

- **a. Write your definition of function, and give three different ways functions can be represented (with examples):**

- **b. What does it mean for a function to be continuous?**

- **c. Determine whether each of the following statements are true or false. If true write an explanation for why, if false give <u>one</u> counter example to show why it is false.**

Statement	**True/False**	**Explanation or Counter example**
$f(s+t) = f(s) + f(t)$		
If $f(s) = f(t)$ then $s = t$		
If $x_1 < x_2$ and f is a decreasing function, then $f(x_1) > f(x_2)$		

Agreement Initials:

1 a. ___ ____ ____ ____ 1 b. ___ ____ ____ ____

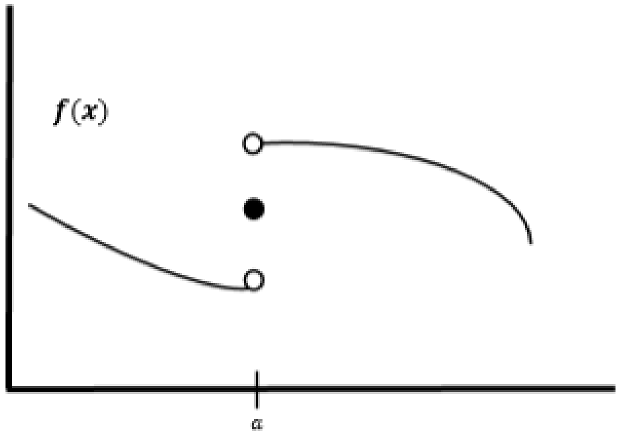
1 c. ___ ____ ____ ____

Problem 1. Disagreements and alternative answers:

2) (25 points)

a. Write your definition of a limit, and give examples for when: (1) a limit exists, (2) does not exist, (3) or is ambiguous (may or may not exist):

b. Does the limit exist as $x \to a$ for the function $f(x)$ shown below? Justify your answer using the epsilon-delta definition of a limit and mark the terms you use in your explanation on the given graph.



Agreement Initials:

2 a. ___ ___ ___ ___ 2 b. ___ ___ ___ ___

Problem 2. Disagreements and alternative answers:

3) (25 points)

a. **State the definition of a derivative using mathematical notation and your own words (i.e., what does it mean/tell you?)**

For each of the following explain the concept, give two examples for each, and explain how each connects limits to the idea of derivatives:

b. **The slope of a tangent line of a function.**

c. **The average velocity of an object.**

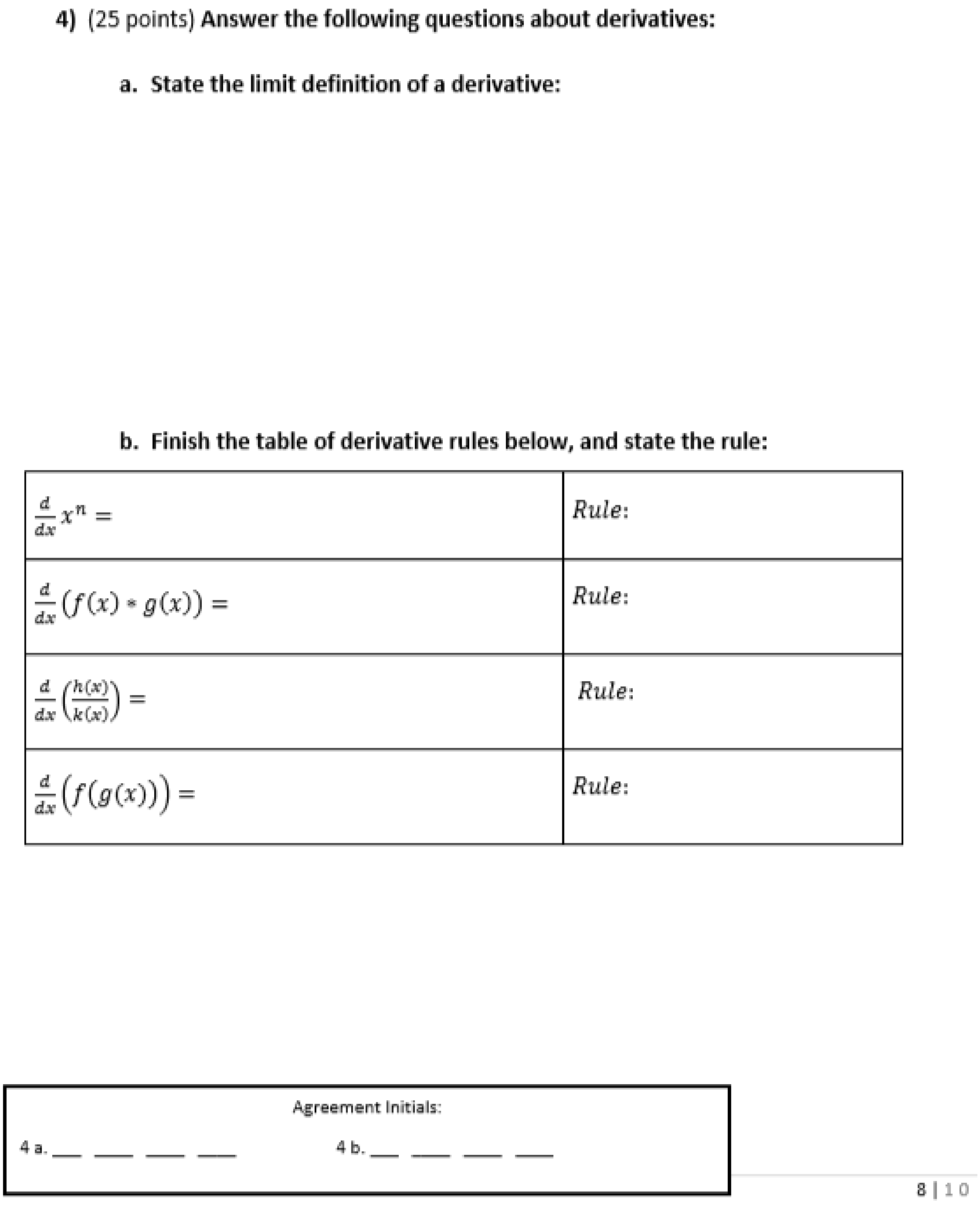
d. **The instantaneous rate of change of a function at a specific point.**

Agreement Initials:

3 a. ___ ____ ____ ____ 3 b. ___ ____ ____ ____

3 c. ___ ____ ____ ____ 3 d. ___ ____ ____ ____

Problem 3. Disagreements and alternative answers:



4) (25 points) **Answer the following questions about derivatives:**

a. State the limit definition of a derivative:

b. Finish the table of derivative rules below, and state the rule:

$\frac{d}{dx}x^n =$	$Rule$:
$\frac{d}{dx}(f(x) * g(x)) =$	$Rule$:
$\frac{d}{dx}\left(\frac{h(x)}{k(x)}\right) =$	$Rule$:
$\frac{d}{dx}\left(f(g(x))\right) =$	$Rule$:

Agreement Initials:

4 a. ___ ____ ____ ____ 4 b. ___ ____ ____ ____

Problem 4. Disagreements and alternative answers:

BONUS (for up to 10 points):

Is the following a function? Using mathematical terms explain why the machine is or is not a function. Thoroughly explain your choice, list the domain and range of the function based off the given machine, and create a graph to support your answer. You must give enough logical information to *convince the grader* of your choice (function or non-function). Since you are trying to convince me, you must all agree on one approach to explain this problem.

Appendix D: Post-Test Reflections Code Book

Deductive codes

Codes were applied to statements in red, note multiple codes could apply to one student's response.

Code	Definition of Code	Examples from Post-Test Reflections
Membership (+)	Comments regarding "team" or "team work" in which the student felt a part of their group community.	*"I truly enjoy my group. I think this is the best possible group I could be in, they help [me] understand material."*
Membership (-)	Comments in which the student did not feel as part of the group, or left out of group, or left out of the group discussion. Note feeling as a member of the group community.	*"My ideas and thoughts were not listened to and disregarded by all of my group members. Even if my answers were wrong, I would have liked it if someone could have told me that I was wrong and explained why."*
Affect (+)	Comments in which the student felt listened to and comfortable in their group to speak their ideas.	*"I felt that I was able to share anything I needed to say and got great feedback."*
Affect (-)	Comments in which the student feels stressed, inadequate, and not listened too by their group.	*"I don't think that I was listened to at all."*
Acceptance (+)	Comments in which the student felt as a valued and respected contributing member of their group.	*"I think my contributions were received well, when I added to the conversation, I felt like my comment was being considered."*
Acceptance (-)	Comments in which the student feels disregarded or excluded by the other group members.	*"[…] even when I did speak or try to contribute my ideas to the group, I felt like they just assumed I was wrong and disregarded what I said."*
Trust (+)	Positive comments about the testing materials or instructor's commitment to the students' learning.	*"I think it was a nice way to test."* *"I could express more knowledge of things I learned on the group test rather than the individual one."*

Trust (-)	Negative comments about the testing materials or instructor’s commitment to the students’ learning.	*“I thought it was much more difficult than everything we did so far (classwork, homework, practice test/game)”*
Desire to Fade (+)	Comments in which the student describes their active participation in the group.	*“We bounced off more ideas and felt more open about voicing opinions”*
Desire to Fade (-)	Comments in which the student describes being less of an active participant in the group.	*“I was more quiet during this group test and I didn’t want to speak in fear of confusing people or holding the group back.”*
Inductive codes ***Codes were applied to statements in red, note multiple codes could apply to one student’s response.***		
Collaboration	*Comments about having others feedback, or bouncing ideas among group members*	*I felt like we had good conversations going over the concepts in a group test is a warm up for the individual test*
Need for Collaboration	*Comments that expressed a want for more discussion, feedback, or group members to participate more in the discussion.*	*My ideas and thoughts were not listened to and disregarded by all of my group members. Even if my answers were wrong I would have liked it if someone could have told me that I was wrong and explained why.*
Communication/Discussion	*Having others to agree or correct one’s ideas (merged with collaboration)*	*Yes, everyone was able to come to a confirmation or a disagreement and explaining why they either agreed or disagreed.*
Answer/math specific	*Specific comments about problems on the test (code removed from themes as in general the impact was neutral to SOB).*	*The one question that had to do with the particle moving on the ellipse was confusing. We didn’t really know what it was asking for, and there weren’t any key/indicating phrases.*
Self-confidence (+)	*Feeling more confident to take the individual test. Encouraged to share*	*I felt like we had good conversations going over the concepts in a group test is a warm up for the individual test*

	ideas and increased confidence	
Self-confidence (-)	*Comments such as "if I knew" or only speaking if they felt sure about their answer/ideas, or feeling as if they could not help the group with the answers*	*I just felt bad because I wasn't contributing very much to the group because I couldn't connect my ideas and apply them to the problem. I just felt really helpless and bad for my group because I couldn't help more.*
Loss of control	*Not being able to write on the paper. Telling others what to do (specific to writing on the test paper). Not being in control of how to represent the group answer*	*I personally have never felt more stressed and nervous before a test, not because I don't know the material but when am only allowed to write one problem from a test some group members are struggling on basic algebra concepts during the test I get very frustrated and this test today caused a lot of tension in our group.*
Frustration	*Comments about feeling frustrated, added stress, or problems with a specific group member.*	*When we are working in groups during class and are all able to do some of our separate work then come together we get more accomplished. It is stressful and angering to use one pen.*
Time-management	*General comments about managing time in the group during the test*	*We did seem to spend a lot of time on one question.* And *We were spending to much time on some of the problems*

Appendix E: Group Problem Solving Code Book

Dimension 1 - Evaluation of Previous Actions		
Codes:	Meaning	When to code
(+)	Supportive	Supportive actions reinforce the direction of the current approach, can included responses like "yeah" "yep" "yes". Encourages mathematical ideas from the line(s) before. *Current approach* as in the way they are mathematically solving the problem.
(-)	Critical	Critical actions *alter the current approach*, can be simple "no" or alterations such as "yes, but" if the current utterance is altered. *Current approach*
(o)	Unresponsive	Does NOT hear speaker before them, actions that do not evaluate or acknowledge previous speaker, can be off task utterances "what time is it" -- OR -- logistics stuff (signing paper, critiquing handwriting, or deciding how to read/pass paper).
(o*)	Unresponsive-Neutral	Does hear but does not evaluate previous speaker *but does consider previous* response "let me think about that" or "I don't know", "hmm?", "Wait, what?": A pause in though without evaluation -- OR -- introduces new topics based on mathematics or initial solution path for problem. --OR-- Proposing a new way to think about a problem to the group
Dimension 2- Knowledge Content (Directives about writers or readers will be coded as N)		
Codes:	Meaning	When to code
(C)	Contributions	New *Mathematical* ideas *or actions* introduced into the collaborations. Can include *proposals* (this is false), justifications, critiques, *alternative ideas*.
(R)	Repetitions	Repeats knowledge of previous utterance "it's just epsilon" "yeah, it's just epsilon" the second utterance would be a repetition --OR-- Finishes, continues idea of previous speaker (keeping the group on the same path).
(N)	Null	Do not include any mathematical content, includes acknowledgements "mm-hmm" and "yeah", simple evaluations "no" and general questions, "wait, why?"
Dimension 3 - Invitational Form		
Codes:	Meaning	When to code
(_.)	Statements	Declares information without eliciting participation from others "The epsilon goes with Y"
(?)	Questions	Invites participation either directly "What did you say" or implicitly "the epsilon goes with Y right?"
(!!)	Commands	Demand participation, either physically "Write that down" or in the negotiation "Tell me what you are thinking"

Appendix F: Interview Protocol

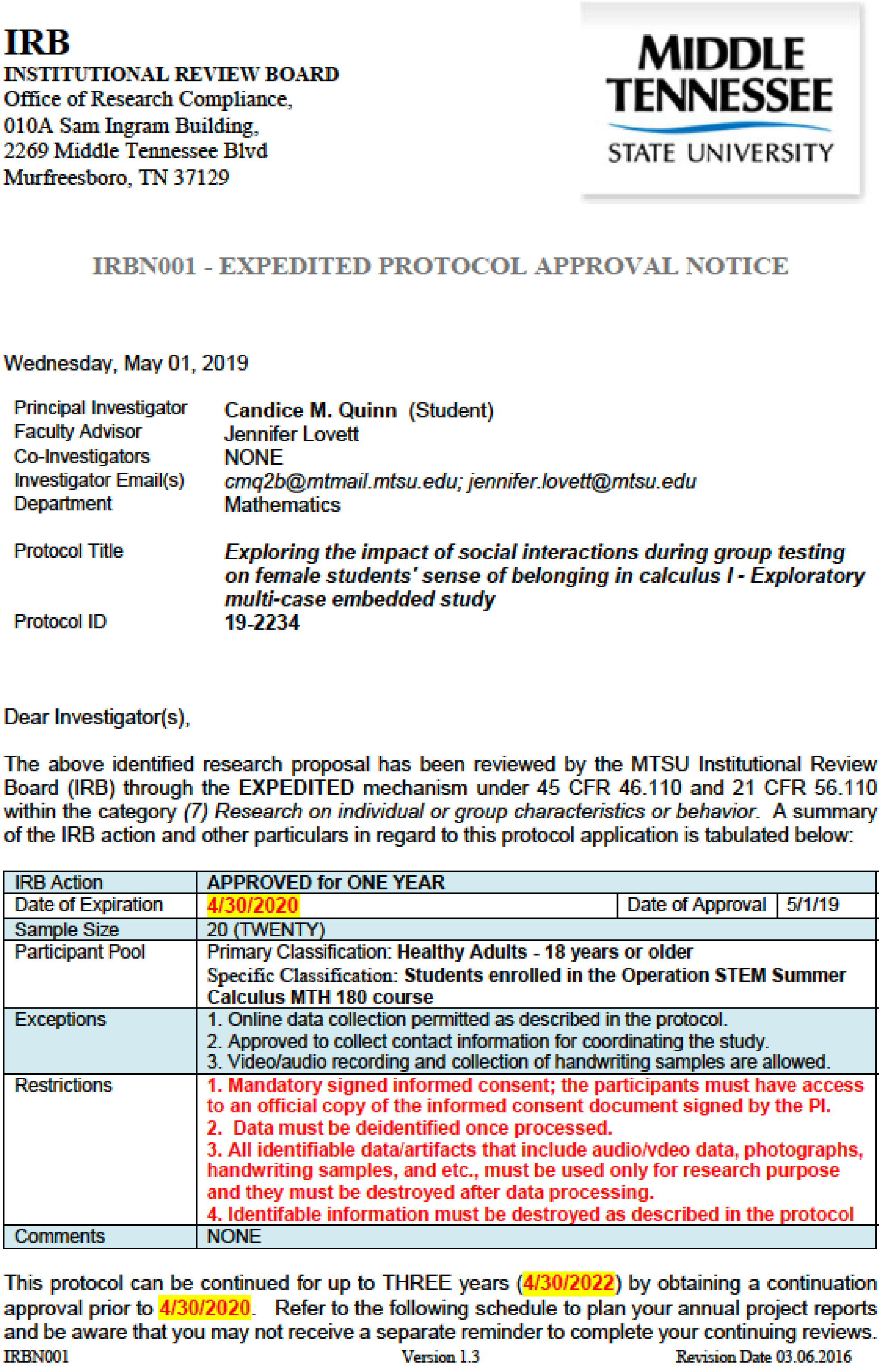
General Questions for all interviews:

1. What is your major?
2. Why did you choose your major?
3. How many more math classes do you have to take past Calculus I?
4. How would you describe your previous experiences in mathematics classrooms?
5. How would you describe your current experiences in our mathematics classrooms?
6. What is your definition of the word “community”?
7. Do you feel like you are part of a community when in our classroom?
8. How about in your group?
9. Do you feel like your group listens to your ideas?
10. In general, how did it feel to take a test in a group?
11. Did you feel confident in your ability to answer the question assigned for you?
12. Do you feel your group members believe you could represent the group answer for that question?

Questions from the post-test reflection. The first question was asked for each response the student provided on the two post-test reflections.

1. In your reflection, you noted that you did (or did not) actively contribute to the ideas discussed during the group test. Can you explain what you meant by your comment (insert students comment here).
2. Can you explain why you think the test was fair or unfair with assessing you and your group’s understanding of the calculus concepts?

Appendix G: Internal Review Board Approval



IRB
INSTITUTIONAL REVIEW BOARD
Office of Research Compliance,
010A Sam Ingram Building,
2269 Middle Tennessee Blvd
Murfreesboro, TN 37129

MIDDLE TENNESSEE
STATE UNIVERSITY

IRBN001 - EXPEDITED PROTOCOL APPROVAL NOTICE

Wednesday, May 01, 2019

Principal Investigator	**Candice M. Quinn** (Student)
Faculty Advisor	Jennifer Lovett
Co-Investigators	NONE
Investigator Email(s)	*cmq2b@mtmail.mtsu.edu; jennifer.lovett@mtsu.edu*
Department	Mathematics
Protocol Title	***Exploring the impact of social interactions during group testing on female students' sense of belonging in calculus I - Exploratory multi-case embedded study***
Protocol ID	**19-2234**

Dear Investigator(s),

The above identified research proposal has been reviewed by the MTSU Institutional Review Board (IRB) through the **EXPEDITED** mechanism under 45 CFR 46.110 and 21 CFR 56.110 within the category *(7) Research on individual or group characteristics or behavior*. A summary of the IRB action and other particulars in regard to this protocol application is tabulated below:

IRB Action	**APPROVED for ONE YEAR**		
Date of Expiration	4/30/2020	Date of Approval	5/1/19
Sample Size	20 (TWENTY)		
Participant Pool	Primary Classification: **Healthy Adults - 18 years or older** Specific Classification: **Students enrolled in the Operation STEM Summer Calculus MTH 180 course**		
Exceptions	1. Online data collection permitted as described in the protocol. 2. Approved to collect contact information for coordinating the study. 3. Video/audio recording and collection of handwriting samples are allowed.		
Restrictions	**1. Mandatory signed informed consent; the participants must have access to an official copy of the informed consent document signed by the PI.** **2. Data must be deidentified once processed.** **3. All identifiable data/artifacts that include audio/vdeo data, photographs, handwriting samples, and etc., must be used only for research purpose and they must be destroyed after data processing.** **4. Identifable information must be destroyed as described in the protocol**		
Comments	NONE		

This protocol can be continued for up to THREE years (4/30/2022) by obtaining a continuation approval prior to 4/30/2020. Refer to the following schedule to plan your annual project reports and be aware that you may not receive a separate reminder to complete your continuing reviews.

Failure in obtaining an approval for continuation will automatically result in cancellation of this protocol. Moreover, the completion of this study MUST be notified to the Office of Compliance by filing a final report in order to close-out the protocol.

Post-approval Actions

The investigator(s) indicated in this notification should read and abide by all of the post-approval conditions imposed with this approval. Refer to the post-approval guidelines posted in the MTSU IRB's website. Any unanticipated harms to participants or adverse events must be reported to the Office of Compliance at (615) 494-8918 within 48 hours of the incident. Amendments to this protocol must be approved by the IRB. Inclusion of new researchers must also be approved by the Office of Compliance before they begin to work on the project.

Continuing Review **(Follow the Schedule Below:)**

Submit an annual report to request continuing review by the deadline indicated below and please be aware that **REMINDERS WILL NOT BE SENT.**

Reporting Period	Requisition Deadline	IRB Comments
First year report	3/31/2020	NOT COMPLETED
Second year report	3/31/2021	NOT COMPLETED
Final report	3/31/2022	NOT COMPLETED

Post-approval Protocol Amendments:

Only two procedural amendment requests will be entertained per year. *In addition, the researchers can request amendments during continuing review. This amendment restriction does not apply to minor changes such as language usage and addition/removal of research personnel. .*

Date	**Amendment(s)**	**IRB Comments**
NONE	NONE.	NONE

Other Post-approval Actions:

Date	**IRB Action(s)**	**IRB Comments**
NONE	NONE.	NONE

Mandatory Data Storage Requirement: All of the research-related records, which include signed consent forms, investigator information and other documents related to the study, must be retained by the PI or the faculty advisor (if the PI is a student) at the secure location mentioned in the protocol application. The data storage must be maintained for at least three (3) years after study has been closed. Subsequent to closing the protocol, the researcher may destroy the data in a manner that maintains confidentiality and anonymity.

IRB reserves the right to modify, change or cancel the terms of this letter without prior notice. Be advised that IRB also reserves the right to inspect or audit your records if needed.

Sincerely,

Institutional Review Board
Middle Tennessee State University

Quick Links:

Click here for a detailed list of the post-approval responsibilities.
More information on expedited procedures can be found here.