DISSERTATION

SUCCESS IN CALCULUS I: IMPLICATIONS OF STUDENTS' PRECALCULUS CONTENT KNOWLEDGE AND THEIR AWARENESS OF THAT KNOWLEDGE

Submitted by

Benjamin D. Sencindiver

Department of Mathematics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2020

Doctoral Committee:

Advisor: Jess E. Hagman
Co-Advisor: Mary E. Pilgrim

Henry Adams
Gene Gloeckner
Jill Zaretsky

ABSTRACT

SUCCESS IN CALCULUS I: IMPLICATIONS OF STUDENTS' PRECALCULUS CONTENT KNOWLEDGE AND THEIR AWARENESS OF THAT KNOWLEDGE

High failure rates in Calculus I contribute to the course acting a filter, rather than a pump, for STEM disciplines. One often cited source of difficulty for students in Calculus I is their weak precalculus content knowledge. In this three-paper dissertation, I investigate Calculus I students' precalculus content knowledge and their awareness of that knowledge. In the first paper, I describe a methodology for collecting data about Calculus I students' tendency to regulate their precalculus content knowledge and analyze the utility of quantifying self-regulated learning as a means for identifying at-risk students. In the second paper, I focus on two factors (calibration and help-seeking) to investigate the how they correlate with Calculus I students' first exam performance. Results highlight the importance of calibration of precalculus content knowledge both directly on student success and how calibration accuracy mediates the benefits of help-seeking. Quantitative analyses of students' precalculus content knowledge highlight Calculus I students' difficulty with the concept of graph, despite students' high confidence in questions related to graph. In the third paper, I conduct interviews with Calculus I students to examine their conceptions of outputs and differences of outputs of a function in the graphical context to understand nuance in how students understand and reason with graphs. Results highlight that students' understandings of quantities and frames of references in graphs of functions can be varied and stable. Students' understanding of quantities also impacts their understanding of other concepts such as differences of outputs and difference quotient. Results of this dissertation have

implications for educators, tutor center leaders, and researchers interested in students' understanding of graph, calibration, and help-seeking.

ACKNOWLEDGEMENTS

I would like to begin first thanking my three advisors, Jess Hagman, Mary Pilgrim, and Cameron Byerley. I was truly raised by a village, and your collective mentorship has shaped my perspective in many different ways. Thank you to Jess Hagman for continuing to push me in my thinking about research and writing. Every conversation I've had with you has helped me become more confident in my work. You've helped frame all of the difficulties I've had in a positive light and helped me move forward in a productive way. Thank you to Mary Pilgrim for working with me since the beginning, despite being in a different state. You introduced me to mathematics education research and instantly began shaping me in a researcher. Your endless support, enthusiasm, and positivity has helped me stay curious and excited about research. Thank you to Cameron Byerley for working with me over the past two years. I've grown so much as a researcher and as a constructivist under your guidance. You met with me with weekly to discuss literature, study design, interview data, and my writing, both in Colorado and Georgia. Thank you for helping integrate me into the UGA community and acting as an advisor to me, regardless of what's on paper. In my eyes (and others' eyes, too), you are undoubtedly one of my three advisors.

Thank you to my past and present committee members, Henry Adams, Gene Gloeckner, Jill Zaretsky, and James Folkestad, for your insightful comments and thoughtful feedback on various iterations of my doctoral work. Thank you to Ken McLaughlin, Steve Benoit, and the CSU Mathematics Department for supporting me while I finish my doctoral work remotely. Thank you to the many friends I've made at CSU for your support, friendship, and many walks to Mugs to get coffee. Thank you also to my new friends and colleagues in Athens and at UGA. I was hesitant how I would cope with working in a new town, but you who have welcomed me in the community with open arms.

Thank you to my parents, Mike and Laura, for their endless encouragement and support. Thanks for reminding me of my goals and helping me think critically about my choices. I'd never be where I am without you. Thank you to my brother Andy, sister Sarah, sister-in-law Kate, and future brother-in-law Ryan for your encouragement and support. Lastly, I want to thank my wife, Jessica. You've been so supportive of me. Thank you for talking about the many ideas I've had, helping wordsmith awkward phrases, cooking dinner while telling me to go back to work, and being an amazing partner. I'm so happy to continue spending the rest of my life with you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

This dissertation was motivated from my own experiences as an instructor of undergraduate calculus at Colorado State University (CSU), particularly Math 160: Calculus I for Physical Scientists. While teaching, I noticed that students in my courses were having difficulty with content that I viewed as a prerequisite for the course. Students were generally successful with symbolically computing derivatives and integrals. However, students struggled with interpreting these concepts, using the algebraic and symbolic language to identify important features, such as critical points, or using graphs to evaluate limits. Given my mathematical training, I was confused by how students had difficulty with what I interpreted to be the easy part (e.g., identifying the critical values of the function after having found the algebraic expression for the derivative of a function). I believed that these impoverished understandings of precalculus concepts were why Calculus I students had difficulty with the course. Further, these difficulties were contributing to the high rates at which students failed Math 160 at CSU, where *failing the course* involves either withdrawing from the course or receiving a grade of D or F. This is because often students taking Math 160 at CSU need to earn a grade of 'C' or better in the course for their program of study or to take the next course in the calculus sequence, Math 161: Calculus II for Physical Scientists. Thus, I wanted to better understand the relationship between students' success in Calculus I and their knowledge of precalculus content, specifically as a way to identify and support students who are at-risk of failing Math 160.

High rates of failures in Calculus I are not isolated to CSU; they are prevalent across the nation, with failure rates between 25% and 40% at research universities (Bressoud et al., 2013). Calculus I is a foundation for disciplines in Science, Technology, Engineering, and Mathematics (STEM) disciplines. Thus, high failure rates in Calculus I can act as a filter to STEM degrees because students often need an A, B, or C in order to progress through a STEM program of study (Steen, 1987). As the nation calls for increased numbers of STEM graduates (President's Council of Advisors on Science and Technology (PCAST), 2012), it is critical that students are successful in Calculus I.

Study Design

The goal of this dissertation study is to understand (1) how Calculus students' awareness of their mastery of precalculus content knowledge and regulatory actions are related to course success, and (2) how Calculus I students can be used to identify and support students who are at-risk of failing the course. Focusing on precalculus content knowledge affords early interventions with students at may be at-risk of failing the course. As early as the first day of class, Calculus I students can begin reviewing and supplementing their precalculus content knowledge as a way to prepare for the course. Hence, data regarding students' precalculus content knowledge can be collected early in the semester, and potentially used to identify factors that are beneficial or counterproductive for learning and success in Calculus I.

This study followed a mixed methods explanatory sequential design (Creswell & Plano-Clark, 2011). During the initial quantitative phase of this study (described Papers 1 and 2), I examined trends across all students in Calculus I to get a broad look at relationships between student success in the course and ways students regulate their precalculus content knowledge. During these analyses, I noticed that many Calculus I students answered questions related to graphs of functions incorrectly. Consequently, I introduced an emergent research goal to understand how students' conceptions for precalculus content support or hinder their conceptual development in Calculus I. During the second phase of this study (described in Paper 3), I conducted clinical interviews with students in enrolled in a different Calculus I at the institution. The purpose of these interviews was to gain insight into how their conceptions of graph related to their conceptual development. Together the quantitative and qualitative analyses provided insight into students' awareness of precalculus content, particularly regarding graphs of functions. These results are discussed in the conclusion chapter of this dissertation.

Study Background

This dissertation study took place at a predominantly white western research university. Quantitative data for this study were collected in the fall semester of 2016 from students in a first-semester calculus course intended for engineering students. Students seeking most STEM degrees are required to complete this course, except for students in the biological sciences (e.g., biology, zoology). Quantitative data collection was a part of an ongoing study investigating high failure rates in the course. With support from the course coordinator, I developed tools in the students' learning management system that were intended to collect data about the students' tendency to regulate their precalculus content knowledge. These tools were implemented across all sections of the course.

Qualitative data was collected in the summer semester of 2019 from students in a calculus course intended for students in biological sciences. This course focused on applications of calculus, and concepts such as exponential growth and rate of change. Additionally, the calculus course for biological scientists had comparably high rates of failure to the calculus course for engineering students. Data collection for this phase of the study was intended for this dissertation study.

At CSU, there is evidence that students' precalculus content knowledge, study habits, motivation, and Exam 1 scores are correlated with course success. Researchers have looked at factors correlated with success, such as precalculus content knowledge, study skills, motivational competencies (Reinholz, 2009; Worthley, 2013), while other researchers have looked at the impact of various institutional efforts (e.g., offering preparatory courses and two-semester Calculus I courses) to support students in being more successful (Pilgrim, 2010; Pilgrim & Gehrtz, 2016). The relationship between Exam 1 scores and course success suggests that any intervention for Math 160 students would need to occur before Exam 1, which occurs in the fourth week of classes, in order to have the greatest impact. Although precalculus content knowledge is important for success in the course, resources for improving students' precalculus content knowledge should not be required of students because there is evidence that students' activity with prerequisite content may detract students from Calculus I course work (Reinholz, 2009). Students should be able to leverage these resources as necessary, particularly if they self-identify as someone who needs to improve their precalculus content knowledge. In other words, students need to *self-regulate* their precalculus content knowledge to prepare for Calculus I (Zimmerman, 2000). It is my assumption that students who act in a self-regulated way to improve their precalculus content knowledge will also act in a self-regulated way as they learn calculus content.

Overview of Papers

The goal of this dissertation was to explore the ways that precalculus content knowledge impacts students' success in Calculus I. My research goals were to understand (1) how Calculus I students' awareness of their mastery in precalculus and help-seeking actions correlate with their success, and (2) how Calculus I students' conceptions for precalculus content impacted their conceptual development in Calculus I. My dissertation comes in the form of three papers, where the first two papers address the first research goal. The second research goal was emergent from my analyses conducted related to the first and second papers, which I will also describe. The third paper address this second emergent research goal.

Summary of Paper 1

In Paper 1, I describe self-regulated learning (SRL) and how acting in regulatory ways can be beneficial for students. To this end, I describe the creation of three online tools that can provide evidence of students regulating their precalculus content knowledge, specifically as a means to prepare for Calculus I. The design of these tools aligns with Zimmerman's (2000) model of SRL, where students regulating their learning (1) prepare for engaging in a task (forethought phase), (2) engage in the task, and (3) reflect of the performance of the task. I aim to answer the research question: How does the degree to which Calculus I students self-regulate their precalculus content knowledge relate to course success?

With help from two experts (one in a mathematics department, and one in a school of education), I created a coding scheme that measured the extent to which students regulate their precalculus content knowledge in preparation for the course. This SRL score utilized data from the three tools as well as data regarding students' access to precalculus help resources. This measure of SRL leverages SRL theory and related constructs, such as calibration, to create the score. I examined correlations between student success and the SRL score. Results highlight students' SRL practices around reviewing precalculus content knowledge as a means to identify at-risk students in Calculus I.

Summary of Paper 2

The goal of the second paper of this dissertation is to better understand how one factor related to the SRL score, calibration, relates to success in Calculus I. Calibration is the alignment of one's mastery with their perceptions of mastery, and is often measured in two ways: calibration accuracy and calibration bias. Calibration accuracy refers to the degree to which a student's perceptions of mastery aligns with their mastery, while calibration bias refers to the degree to which a student's perceptions of mastery exceeds their mastery. I hypothesize that calibration of precalculus knowledge is correlated with success in the first exam of Calculus I, and that help-seeking is one mechanism in which calibration supports student success. My research questions for this paper are:

1. How do Calculus I students' levels of calibration bias and accuracy correlate to performance on Exam 1, after controlling for incoming content mastery? and
2. How do Calculus I students' levels of calibration mediate the benefits of help-seeking on Exam 1 performance, after controlling for incoming mastery?

To answer these questions, I used data from two tools in students learning management system to create measures of calibration accuracy and calibration bias. I conducted a hierarchical linear regression analysis to examine how much variance in Exam 1 scores was captured by successive models that include additional combinations of covariates, such as number of visits to a university mathematics tutoring center. Results highlight the impact of students' calibration on Exam 1 scores and the role of calibration in effective help-seeking.

Subsequent Analyses and the Emergence of Paper 3

The third paper in this dissertation was influenced by the analysis conducted on a content quiz used in Papers 1 and 2. In the fall of 2016, students enrolled in a calculus course for engineering students had access to a precalculus content quiz to gauge students' mastery of precalculus content. Of the nearly 450 students enrolled, 224 students consented to the study and completed the optional quiz. Of these students, only 31.3% received full points on at least one of the two multiple answer items asking about what *f(2)* refers to on a graph (Item 1) or what solutions to *f(x) = 2* represent (Item 2). These items are provided in Figure 1.1 and Figure 1.2.

Q1: Consider the function $\mathrm{g(x)} = -3\mathrm{x} + 2$.

What does g(2) represent? (Mark all that apply)

a. The function gets multiplied by 2.
b. The function evaluated at 2.
c. The y-value on the graph of the function with x-coordinate 2.
d. The x-value on the graph of the function with y-coordinate 2.
e. The height of the graph of the function at $x = 2$.
f. The distance between the graph of the function at $x = 2$ and the x-axis.
g. The distance between the graph of the function at $x = 2$ and the y-axis.
h. The slope of the graph of the function at $x = 2$.
i. $-3(2) + 2$.

Figure 1.1: Item 1 on the precalculus content quiz

Q2: Consider the function $g(x) = -3\text{x} + 2$.

What does the solution to $g(x) = 2$ represent? (Mark all that apply)

a. The function gets multiplied by 2.
b. The function evaluated at 2.
c. The y-value on the graph of the function with *x*-coordinate 2.
d. The x-value on the graph of the function with *y*-coordinate 2.
e. The height of the graph of the function at $x = 2$.
f. The distance between the graph of the function at $x = 2$ and the *x*-axis.
g. The distance between the graph of the function at $x = 2$ and the *y*-axis.
h. The slope of the graph of the function at $x = 2$.
i. $-3(2) + 2$.

Figure 1.2: Item 2 on the precalculus content quiz

Of the students that answered the items, 88.8% and 42.3% answered Item 1 and Item 2 at least partially correct (selecting only correct responses), respectively. Further 25.4% and 17.0% answered Item 1 and Item 2 correctly (selecting only correct answers and selecting all correct answers). Table 1.1 provides data regarding the accuracy of student’s responses to these two items.

Table 1.1: Calculus I student's responses to two items about graph, n = 224.

		Accuracy of Responses to Item 1			
		Incorrect	Partially Correct	Correct	
Accuracy of Responses to Item 2	Incorrect	18 (8.3%)	87 (39.9%)	23 (10.6%)	128 (57.1%)
	Partially Correct	3 (1.4%)	46 (21.1%)	9 (1.4%)	58 (25.9%)
	Correct	4 (1.8%)	9 (4.1%)	25 (11.5%)	38 (17%)
		25 (11.2%)	142 (63.4%)	57 (25.4%)	

The large percentage of incorrect responses to these items suggests that many students in the calculus course have not yet mastered content related to graphs. This analysis lead to the research goal of understanding how students' conceptions for precalculus content support or hinder their conceptual development in Calculus I. The third paper in this dissertation addresses this goal.

Summary of Paper 3

In the data from the precalculus content quiz designed in the first paper of this dissertation, I found that many students in Calculus I answered items related to graphs of functions incorrectly (see Table 1.1). Since graphs are commonplace in the Calculus I curricula, it is vital that students understand graphs in the normative way. The goal of this paper is to explore what conceptions of graph students have and how those conceptions impact students' conceptual development. As such, I adopt the radical constructivist perspective (von Glasersfeld, 1995) to examine what students are thinking about graphs, rather than students' responses are correct.

In this paper, I aim to answer the research question: How do Calculus I students' conceptions of output impact students' conceptions of the difference of outputs in a graphical context? To answer this question, I conducted clinical interviews with five Calculus I students in the summer of 2019. I used theoretical thematic analysis (Braun & Clarke, 2006) to categorize students' conceptions of output and differences of outputs. I used David, Roh, and Sellers' (2019) constructs location-thinking and value-thinking describe students' conceptions of output of a function and Thompson's (1993) distinction of subtraction and quantitative difference describe their differences of outputs. I used inductive thematic analysis (Braun & Clarke, 2006) to code the conceptions that did not fit into these categories, identifying the mathematical object that students were referring to as they engaged with the task. This analysis showed that students have a variety of conceptions for output and they were related to their conceptions of difference of output.

Chapter 2

Paper 1: Self-Regulated Learning: A Framework for Contextualizing Student Activity in Calculus I

Since the 1980s, Calculus I has commonly been labeled as a "filter" in the STEM pipeline, "blocking access to professional careers for the vast majority of those who enroll" (Steen, 1987, p. xi). As failure rates in the course (grade of D, grade of F, or course withdrawal) continue to remain high (around 25% at research universities) efforts to address the issue persist (Bressoud et al., 2013). Unfortunately, Calculus I at research institutions has been documented to have a negative impact on the student experience and, in fact, has led to students leaving programs of study requiring additional courses in mathematics (Bressoud & Rasmussen, 2015; Seymour et al., 2019; Seymour & Hewitt, 1997). Many are attempting to tackle this problem through changes in instructional practices, however, isolated instructional changes can be particularly challenging to sustain (Henderson et al., 2011). More recent efforts, such as the *Student Engagement in Mathematics through an Institutional Network for Active Learning* project, have sought to sustain instructional practices such as active learning and study what mechanisms at the institutional, departmental, and classroom level support lasting instructional change (Smith et al., 2017)

In conjunction with instructional change, it is important that students are active participants in their own learning and success, especially at the undergraduate level where students are more autonomous in their learning. As students need to become more independent, I believe that students must enact and sustain a self-regulated approach to their own success, including engaging in a process that enables them to assess, reflect on, and then modify their own learning behaviors. In this paper, I present a methodology to collect evidence of students behaving in a self-regulated way. I discuss what self-regulated learning (SRL) is, how it has been measured in the past, how I developed a set of tools that can measure SRL, and its relationship with academic performance.

What We Can Learn About At-risk Students from Data

Large amounts of data are collected on students. In addition to demographic data and course performance data, universities are capable of collecting data that describe engagement with courses via online and in-person interactions. If educational tools are designed with foresight, the data gathered by these tools can provide instructors with an ongoing assessment of student's engagement and provide insight into the ways students engage with the course over time (Baker et al., 2004; Ma et al., 2015; Winne & Baker, 2013; Zhou & Winne, 2012). Additionally, the ways students use educational tools has been related to student's academic achievement (Jo et al., 2015; Morris et al., 2005). Hence, data around students' use of these tools could be leveraged to detect if students may potentially be at-risk of failing a course, affording early interventions to help students better prepare for the course. While fine-grained data about students' interactions with digital resources can provide a rich set of data about individual students, those who do not engage with resources can easily be misrepresented by their digital footprint.

Consider two Calculus I students, Matilda and Derek, both of whom are enrolled in the same Calculus I course. The course requires completion of online homework, which is due at the end of the semester, and all students have free access to a mathematics tutoring center (MTC). Despite the availability, neither student begins working on their online homework until the end of the semester. Derek has been successful with cramming in his high school math classes, so he does not work on the online homework until the last week of class. Matilda, on the other hand, keeps up with the course content well, but has a heavy STEM course load and is struggling in her chemistry course. Having assessed her calculus and chemistry abilities, Matilda determines that she needs to spend more time on chemistry and prolongs working on her online homework until the last couple of weeks of the semester. Thus, by the time both students take their first exams (after four weeks of class), neither has made progress on any online homework.

Before the last few weeks of the course, both Derek and Matilda are indistinguishable with respect to their digital footprints with the Calculus I online homework system because neither of the students have engaged with the online homework, albeit for different reasons. Each students' lack of engagement does not imply either student will fail the course but may suggest that both are acting in non-productive ways that will hinder their success in the course. By combining the students' online homework interactions with other data points, such as help-seeking strategies, performance on assessments, and responses to feedback, a more detailed picture of the student can be painted. Anyone in a position to support student learning (e.g. educators, course coordinators, and administrators) can then use this data to build nuanced models of students that better represent and contextualize students' actions in a course in terms of self-regulated learning theory.

In this paper, I will discuss how I used a self-regulated learning framework to contextualize students' out-of-the-classroom interactions in terms of students' self-regulation through an SRL score. Further, I present findings that relate the SRL score to course performance data. In addition to looking at these relationships for the entire student population, I will also focus on students like Derek and Matilda, who have not used available calculus resources early in the course. This subpopulation will be referred to as *disengaged students* because regarding their actions early in the course, these students have not yet accessed the course's online homework system nor visited the university MTC for help with calculus content.

Self-Regulated Learning

Identification of mistakes can be challenging for mathematics students. Students struggle with recognizing when a mistake occurs and how to make changes in study habits in order to minimize the occurrence of mistakes and address potential knowledge gaps (Zimmerman et al., 2011). While there are many models of SRL (Boekaerts & Corno, 2005; Pintrich, 2000; Winne & Hadwin, 1998), I draw upon Zimmerman's (2000) SRL model, which is based in Social Cognitive Theory. According to Zimmerman, SRL is "the self-directed process by which learners transform their mental abilities into academic skills" (p. 65) by actively monitoring and regulating their mental processes. Students that are described as self-regulated learners are continually deepening their understanding of the process by which they learn, and modifying and adapting their study habits and strategies to become more successful learners. Zimmerman (2000) describes SRL in terms of a three-phase model, including forethought, performance, and self-reflection.

Forethought

When a self-regulated learner plans to engage in a task, the student first engages in the pre-task phase *forethought*, which consists of two primarily processes: task analysis and self-motivation. During task analysis, a self-regulated learner will set goals and identify strategies to employ so that those goals can be achieved (Zimmerman, 2000, 2002). Self-regulated learners have the skills to assess a task and determine appropriate strategies to apply to reach their decided goals. A student's ability to analyze the task alone, however, is not enough for a student to regulate their learning; their motivation will influence the level to which students enact these practices.

Student's self-motivation comes from their beliefs about themselves as learners such as their self-efficacy to complete the task, and the student's expected outcomes from learning. For example, a student who expects to use the knowledge gained from the learning task may attempt to overcome potential obstacles and difficulties during the task than a student who does not expect to use the knowledge gained. Similarly a student who is interested in calculus content and feels efficacious about learning differentiation of trigonometric functions will more likely learn the content in a self-regulated way than a student who does not expect to use differentiation in their future profession and does not believe that they can learn the content area. A students' beliefs such as their confidence or self-efficacy to complete the task about learning will further influence whether a student prepares for the learning task, as well as how they execute the task during the performance phase (Zimmerman, 2002).

Performance

Upon engaging with the learning task (i.e. during performance phase of SRL), self-regulated learners begin implementing the strategies they identified in the forethought phase. Self-control and self-observation are primary elements of performance that require focus and adaptation in order to optimize effort (Zimmerman 2000). Through self-control and self-observation competencies such as time management and attention focusing, a self-regulated learner is able to make adjustments to strategies that are not in the moment proving successful (Zimmerman 2000, 2002). By monitoring and maintaining a record (mental or physical) of task details during this phase, a self-regulated learner can consider adjustments to their strategies that may need to occur in the moment or in the future. The accuracy of the learner's record of the performance phase and their content knowledge are vital for a learner to make effective decisions about what strategies are proving successful, and which need to be adjusted to attain the intended learning goal.

Self-Reflection

Self-reflection involves learners being deliberately introspective after performing the learning task by drawing on their knowledge of the performance phase to assess what strategies were successful and how they could be improved. A self-regulated learner assesses their performance and then reacts to that self-judgment. Such a learner evaluates their performance by "comparing self-monitored information with a standard or goal" and "attributing causal significance to the results" (Zimmerman 2000, p. 21). For example, when a student receives a lower grade than expected on an exam (e.g., a 'C' instead of an 'A'), their reaction may be to change their study habits and strategies or to seek help from an instructor or tutor. Each of these responses could be based on differing causal attributions. A student who deems a study tactic, such as creating flash cards, as unhelpful for an exam may change their study habits in the future to incorporate more completion and review of practice exams. A student who identifies the content of focus during studying may seek help with the content that they do not know sufficiently. For a self-regulated learner, results from the self-reflection phase will play a significant role in engagement with future tasks, such as the forethought and performance phases of future SRL cycles.

Relationships between SRL and Calibration, Self-Efficacy, and Academic Performance

In much of the literature, self-regulation has been closely tied to self-efficacy (Pintrich, 2004; Pintrich et al., 1993; Pintrich & De Groot, 1990). Self-efficacy can be described as judgements of one's competence to perform a task (Bandura, 1981; Schunk & Pajares, 2004).

The Motivated Strategies for Learning Questionnaire (MSLQ) (Pintrich et al., 1991) explicitly uses self-efficacy as one of its constructs, and significant positive correlations between self-efficacy and academic performance have extensively been reported in the literature (Honicke & Broadbent, 2016; Pintrich, 2004; Pintrich et al., 1993). Bandura and Schunk (1981) found that for students who set attainable sub-goals in an academic setting had higher progress in self-directed learning (related to SRL), academic performance, and perceived mathematical self-efficacy. Self-efficacy in academic settings can be closely tied to student self-confidence, as researchers have used confidence measures on particular tasks as a measure of self-efficacy (Zimmerman et al., 2011).

Related to SRL, students' calibration of their self-efficacy and self-reflective judgements have been reported to be vital to success of implementing self-reflection in the SRL cycle (Schunk & Pajares, 2004; Zimmerman et al., 2011). By aligning one's self-confidence to complete a task (both prior to completion and after completion) with their task abilities, a student will be better positioned to leverage feedback from a performance than if a student's confidence and abilities are misaligned. For example, Zimmerman and colleagues (2011) summarized that overly confident self-beliefs about one's abilities can "hinder the adaptive use of feedback" (p. 110). Given influence of students' calibration on their reactions to feedback, it is no surprise that calibration of student's pre- and post-task confidence have been found to be positively correlated to academic performance (Zimmerman et al., 2011).

Pajares and Graham (1999) developed a methodology for quantifying calibration – analyzing how much a student underestimated or overestimated their confidence in performing the task. Using this methodology, Zimmerman et al. (2011) found that students who underestimated their abilities during pre- and post-task measures of confidence tended to have higher academic performance than students who overestimated their abilities. Additionally, they saw that students who were conditioned with SRL instruction reported less overconfidence in both pre- and post-task confidence than students with standard instruction and were more accurate in their assessments of confidence. Students with SRL instruction, instruction focused on self-reflection, self-efficacy, and correcting mistakes, reported higher academic achievement results than students with the standard instruction. While this may be a result of the testing effect (i.e., increased retention due to more frequent testing; Roediger & Karpicke, 2006), analysis suggests that *high reflectors* - students who complete reflection forms more often on their quizzes – performed better on exams than *low reflectors*, after controlling for a pretest (Zimmerman et al., 2011). These results suggest that students who self-reflect more have better accuracy in confidence measure and these students tend to perform better on exams. Further, these self-regulatory skills can be taught to students.

Measuring SRL

The most commonly used instrument for measuring SRL is the MSLQ. The MSLQ is an 81-item self-report questionnaire that evaluates undergraduate students' motivation and learning strategies (Pintrich et al., 1991). Through its 15 subscales, the MSLQ has been used primarily to study components of SRL as well as examine the relationship to academic performance (Pardo et al., 2016; Pintrich et al., 1993; Pintrich & De Groot, 1990; Zimmerman & Kitsantas, 2014). There have also been some novel cases in which the MSLQ has been used to look at how SRL relates to interactions with learning management systems (Dabbagh & Kitsantas, 2005; J.-E. Lee & Recker, 2017). However, although extensively validated, the MSLQ is not without criticism. Two major criticisms are due to the MSLQ being a self-report, which give highly subjective responses, and that students need to recall back on previous events to accurately determine levels of engagement in different processes, which can be skewed due to memory (Winne & Perry, 2000). In addition, finding evidence of SRL outside of such self-reports is non-trivial (Winne & Baker, 2013; Winne & Jamieson-Noel, 2002).

Research suggests that students' self-reports on study behaviors may reflect how they think they should study (Worthley et al., 2016), and that students' perceptions of what they should do does not align with how they actually studied (Winne & Jamieson-Noel, 2002). To overcome these issues, Cleary and colleagues (2012) argue that microanalytic protocols (questions used to assess self-regulation behaviors during a task) can provide better insight into a learner's self-regulation. They argue that "assessment tools that examine regulatory thought and action as they occur in real time during a particular task [have] the potential to provide more useful information that can lead to contextualized, individualized interventions for youth who struggle in school" (p. 16). Microanalytic protocols alleviate some inaccuracies of self-reports due to memory by having subjects reply to questions around SRL during the task, though the subjectivity of the respondent can still be problematic. To make aspects of microanalytic protocols more objective, I designed online tools to capture observable, measurable events that align with SRL competencies while the student engages in preparing for a calculus course.

Methods

Data for this study came from a large, predominantly Caucasian western research university in the United States during the fall semester of 2016. The course (Calculus I) is large multi-section coordinated course with common syllabus, homework, and exams that primarily serves students seeking STEM majors, such as engineering, physics, mathematics, statistics, computer sciences, and chemistry. Course content includes limits, differentiation, and integration for functions of one variable, much of which is considered foundational content for many STEM disciplines.

Calculus I at the institution has historically had some of the highest rates of failure (finishing the course with a grade of D or F, or withdrawing from the course), with failure rates between 25% and 43% in the past 10 years. Research conducted at the institution suggests that poor study skills and impoverished content knowledge of precalculus may be factors contributing to the high failure rates in the course (Reinholz, 2009; Worthley, 2013). Nationally, similar findings hold, particularly that weakness with prerequisites for calculus (e.g., algebra, functions, and trigonometry) are source of difficulty for many calculus students (Agustin & Agustin, 2009; Breidenbach et al., 1992; Carlson, 1998; Carlson et al., 2015; Moore & Carlson, 2012). For example, Agustin & Agustin (2009) found that students struggling with first-semester calculus tend to have more errors due to precalculus content rather than calculus content.

For these reasons, the importance of addressing prerequisite deficiencies was stressed to the students as being an integral part of the course, as calculus concepts rely heavily on the application of precalculus knowledge. I designed online tools around the self-regulation of prerequisite skills for Calculus I, particularly around readiness and remediation of prerequisite skills. These online precalculus tools were made available in the first two weeks of the course for three primary reasons: (1) to encourage students to revisit precalculus material early in the course, (2) to support students in improving vital precalculus skills for the course, and (3) to collect data about students' activity prior to the Exam 1. Reinholz (2009) identified Exam 1 at the institution as highly correlated with students' success in the course, so any attempts at early intervention with students would likely need to occur before the fourth week of the course (i.e. before Exam 1). By collecting data about students by the second week of the course, analysis of the data presented in this paper could be used to inform future interventions for students that may be at-risk of receiving a grade of D of F in the course. I hypothesize that self-regulation practice around precalculus material will translate into self-regulation around calculus content and, hence, success in Calculus I (Labuhn et al., 2010; Zimmerman et al., 2011; Zimmerman & Schunk, 2001).

Data Sources

To study self-regulation around precalculus remediation, I designed three tools aligned with Zimmerman's three phases of SRL. These tools included a self-assessment (forethought), a content quiz (performance), and a post-quiz reflection (self-reflection), all of which were optional for the students and made available through the students' learning management system (LMS).

Precalculus Self-Assessment

The Precalculus Self-Assessment (PSA) is a 16-item survey asking students to rate their confidence in correctly answering precalculus questions on a 5-point Likert scale from 1 to 5 (*no confidence, little confidence, average, confident,* and *very confident*). By assessing their confidence in precalculus topics, the PSA fits into the task analysis component of forethought, allowing students to ask themselves how well they think they know precalculus material.

Precalculus Content Quiz

Students' participation in performance phase of SRL was determined by whether or not the student took the Precalculus Content Quiz (PCQ). The PCQ is comprised of 12 multiple-choice and multiple-answer questions about precalculus material essential for Calculus I. Precalculus topics for the content quiz and self-assessment included function notation, graphs of functions, simplifying algebraic expressions, solving algebraic equations, basic trigonometry, and solving trigonometric equations. A PCQ score was created from students' responses. Multiple choice questions were scored either with zero points (incorrect response) or one point (correct response), while multiple answer questions were scored either with zero points (at least one incorrect response was selected), 0.5 points (only correct responses were selected, though not all responses), or one point (only correct responses were selected, and all correct responses were selected). Therefore if a student completed the PCQ, their score could vary between 0 and 12.

Upon finishing the PCQ, students received information on what questions they answered correctly and incorrectly. When a question was answered incorrectly, immediate feedback was provided, though the correct answer was not given. Feedback included the relevant topic(s) associated to the incorrectly answered question that the student could review. In addition, links to available resources regarding the topic were also provided (see Precalculus Help Resources). Data collected from the PCQ by the researchers included student responses to each question. This information can provide insight to better understanding students' self-regulation in the performance phase relating to self-control and strategy implementation.

Precalculus Reflection Tool

Students were given a five-item survey called the Precalculus Reflection Tool (PRT) which was intended to be used after the PCQ. The PRT asked students questions such as 'What topics from the prerequisite content quiz do you plan to study?' and 'How do you plan to study/practice problems from the prerequisite content quiz material?'. Responding to the prompts of the PRT provides evidence that a student is reflecting on his or her performance on the PCQ. This behavior indicates that a student may be planning to address possible content weaknesses, though it does not provide evidence of subsequent follow through (i.e. additional forethought and performance of the intended task). Such evidence would need to be obtained by tracking student access to digital and physical resources related to precalculus.

Precalculus Help Resources

To understand whether students were completing the cycle of SRL around preparing for Calculus I, data was collected about students' usage of precalculus help resources (PHR). PHR were available both online and in-person formats to support students in improving their precalculus knowledge and skills. Online, students had access to a repository of activities and videos targeted at precalculus concepts, including the topics in the PSA and PCQ (e.g., trigonometric functions, manipulating algebraic expressions, solving algebraic equations, etc). To make the use of these digital resources as easy as possible for the students, the repository of the precalculus activities and videos was housed in the same system as the students' online Calculus I homework. This online resource was made available to students through their LMS in two ways: either through a direct link, or through the links in the generated feedback of the PCQ. If a student navigated through the LMS to the online precalculus resources, data was collected that identified the student, thus providing information as to whether or not they used the precalculus online resources and when they used them.

In addition to online resources, students could also seek help with precalculus content by participating in weekly precalculus workshops. These sessions were free to all students and publicized during Calculus I class periods. Students participating in the precalculus workshops were asked to work in groups to complete weekly worksheets and activities. The precalculus workshops aligned the content begin covered each week with important precalculus concepts that were either currently being used in the class or would be needed in the following week. For example, one week before students in Calculus I were scheduled to cover optimization, the precalculus workshop focused on methods for finding zeros of functions (e.g., factoring binomials, solving trigonometric equations) because identifying zeros of derivative functions is one essential technique to identifying critical values of functions. The precalculus workshops took place in the same space as the university MTC, which afforded collecting data about what students were going to the precalculus workshops.

Formulating the SRL Score

Using the data from the online tools, I categorized responses (or lack thereof) within each tool as well as resource usage (see Table 2.1). Categories for students' use of the PSA was based on their average responses across the 16 items on the PSA (between one and five), and coded to have either high confidence (mean confidence rating greater than or equal to three), low confidence mean confidence rating less than three), or did not use the PSA. The threshold of three was chosen to distinguish students with high and low confidence because three represented average confidence on the PSA, and anything above three would represent above average confidence for the particular item. Categories for students' use of the PCQ was based on the number of correct responses they made (i.e. PCQ score) and coded as either high performance (PCQ score of eight points or higher), low performance (PCQ score of less than 8 points), or did not use the PCQ. The threshold of eight was chosen to distinguish students with high and low performance because the median PCQ score was eight. Categories for student's use of the PRT was coded as either used or did not, based on whether or not the student used the PRT. Similarly, categories for students' use of PHR were based on whether or not the student used either attended any of the precalculus workshops or whether they accessed the online precalculus resources in the first two weeks of the semester. Students' usage of the PHR was hence coded as either used or did not use.

Table 2.1: Data Sources Leveraged to Create the SRL Score and How Each Data Source was Coded.

Data Source	Categories for Data Sources
PSA	High Confidence (Mean Response ≥ 3), Low Confidence (Mean Response < 3), or Did Not Use
PCQ	High Performance (PCQ Score ≥ 8), Low Performance (PCQ Score < 8), or Did Not Use
PRT	Used or Did Not Use
PHR	Accessed/Attended or Did Not Access/Attend in the First Two Weeks of the Semester

After coding student usage of each data source, I considered combinations of tool usage/response (e.g., student had high confidence, low performance, but not use PRT) and resource categories (e.g., accessed PHR). As a team, two educational researchers (one from the mathematics department, and one from the education department) and I then coded each combination of usage categories into a 6-point scale (between zero and five) representing the strength of evidence that a student was engaging in SRL. It is worth noting that SRL is a process that can stretch weeks, semesters, and years, so the SRL score is an early snapshot of the SRL process around the task of assessing one's own precalculus knowledge and potential remediation of those skills within the first two weeks of class.

Through this coarse coding, students' interactions with the four data sources in Table 2.1 were coded as one of 36 different combinations of student behaviors, though only 27 combinations were observed. Each of the observed outcomes were analyzed using Zimmerman's three phase SRL model. Using the guiding question 'How appropriately is the student responding to the feedback about their impressions on their precalculus abilities?', students were assigned a SRL score of zero, one, two, three, four, or five, with a score of zero indicating no evidence of self-regulation and a score of five indicating evidence of a high level of self-regulation. Further, the SRL score incorporated the alignment of students' precalculus confidence (PSA) and performance (PCQ), where students' whom were more accurate in their precalculus mastery levels were coded to have higher SRL scores than students who were inaccurate alignment. This choice was made in alignment with prior research that suggests that the alignment of student's perceptions and mastery has been found to be a product of SRL training (Zimmerman et al., 2011). Additionally, students' PSA was also taken into account in this score, as various studies have found self-efficacy to be strongly correlated with academic performance (Honicke & Broadbent, 2016; Pajares & Graham, 1999; Pajares & Miller, 1994). A full description of the coding scheme is provided in Appendix A.

Consider Derek. Derek had high confidence (average PSA confidence greater than or equal to three) on the PSA and low precalculus performance on the PCQ (less than eight points on the PCQ). Further Derek did not use the PRT to reflect on his precalculus mastery, nor did he access any of the precalculus resources. Though Derek used both the PSA and the PCQ, he is not responding to the prompt from the PCQ suggesting precalculus deficiencies. Using the coding methodology for SRL, Derek would receive an SRL score of two, since he is not using the low PCQ score as a prompt to revisit precalculus material (See Table 2.2). Additionally, Derek is miscalibrated with his precalculus confidence and performance, as evidenced by his perception of high ability (i.e. high confidence in precalculus) and low precalculus mastery. If Derek instead used the PRT and sought precalculus remediation through either the digital or physical resources, Derek would have received a four on the SRL score, since he would be tending to the feedback that his precalculus abilities were insufficient (see “Alternative” Derek in Table 2.2).

Table 2.2: Resource Usage and SRL Score for Derek and “Alternative” Derek

Student	Coding of for Each Data Source				SRL Score
	PSA coding	PCQ coding	PRT coding	PHR coding	
Derek	High Confidence	Low Performance	Did Not Use	Did Not Use	2
“Alternative” Derek	High Confidence	Low Performance	Did Use	Did Use	4

Now consider Matilda who used both the PSA and PCQ. She had low confidence on the PSA (average PSA confidence less than three), but had high precalculus performance on the PCQ (eight or more points on the PCQ). Since Matilda proved sufficiently strong with precalculus material, she decided not to reflect on her score nor seek precalculus remediation so that she could spend more time on courses in which she was having more difficulty such as chemistry. The way Matilda engaged with the tools shows appropriate actions given her background, and she therefore received an SRL score of four (see Table 2.3). Matilda did not receive a five because she was miscalibrated in her precalculus abilities and she had low confidence. Even if Matilda were to seek help or reflect on her PCQ score, Matilda would still receive a four since her abilities and confidence regarding precalculus material are still miscalibrated (see “Alternative” Matilda in Table 2.3). Only by having higher confidence in her abilities as well could she receive an SRL Score of five.

Table 2.3: Resource Usage and SRL Score for Matilda and "Alternative" Matilda

Student	Coding of for Each Data Source				SRL Score
	PSA coding	PCQ coding	PRT coding	PHR coding	
Matilda	Low Confidence	High Performance	Did Not Use	Did Not Use	4
"Alternative" Matilda	Low Confidence	High Performance	Did Use	Did Use	4

Lastly consider the different student, Jay. Jay engaged with none of these optional tools and accessed none of the precalculus resources (did not use the PSA, PCQ, PRT, nor PHR). Though he may have been remediating his precalculus knowledge via tools and resources that were not being monitored, there was no evidence to indicate this due to the lack of engagement with the provided tools and resources. Therefore, Jay would have an SRL score of zero (see Jay in Table 2.4).

Table 2.4: Resource Usage and SRL Score for Jay

Student	Coding of for Each Data Source				SRL Score
	PSA coding	PCQ coding	PRT coding	PHR coding	
Jay	Did Not Use	Did Not Use	Did Not Use	Did Not Use	0

Data Collection and Analysis

In addition to data gathered to compute the SRL score for each student, I collected data about students' exam scores, online Calculus I homework access, MTC attendance, and final course grades. Many students enrolled in the course need to obtain a grade of 'C' or better in the course to satisfy their program of study. Hence, 'success' in Calculus I was defined by a final letter grade of A, B, or C, and grades of D and F were classified as 'failure'.

Table 2.5 presents the distribution of SRL scores across the 376 consenting students with the failure rate for each group, that is the percentage of students who completed the course with a grade of D or F. The same data is also shown for the subset of students that were considered disengaged in the course, that is students who did not access the Calculus I online homework system nor visit the university MTC for help in the Calculus I course as of week four (prior to the first exam). Note that for this particular semester, the online homework for the entire course was due at the end of the semester, so students who had not yet accessed their online homework did not lose any points in the course. Data from this analysis is provided in Table 2.5. For example, of the 64 students who received an SRL score of three, 31.2% of those students received a grade of D or F. Of these 64 students, 23 of these students were identified as being disengaged with the course, and 52.2% of these disengaged students received a grade of D or F.

Table 2.5: Summary Statistics for SRL Score Within Two Group of Consenting Students: All Students and All Disengaged Students.

	All Students Enrolled in Calculus I		Disengaged Students in Calculus I	
SRL Score	Number of Students	Failure Rate	Number of Disengaged Students	Failure Rate
0	32	46.9%	20	60%
1	50	32%	16	56.2%
2	18	27.8%	6	16.7%
3	64	31.2%	23	52.2%
4	104	24%	29	31%
5	108	17%	23	21.7%
Total	376	26.3%	117	41%

Within both of these groups, failure rates tend to decrease as students SRL score increases (with the exception of those with an SRL score of two), and the disengaged students are of higher risk of failing than for all students.

Table 2.6 provides data comparing the various academic performance measures of two groups of students: those who had 'high' SRL scores (three or larger) and low SRL scores (less than three). This comparative analysis was conducted for all consenting students (n = 376), as well as with the subpopulation of consenting students that were considered disengaged (n = 117). When considering all consenting students, spearman's non-parametric correlation reported statistically significant differences between students with high and low SRL scores on all course exams and the final course grade, with $r(374) > 0.14$ and $p < 0.005$ for all correlations. Positive correlation coefficients suggest that students with high SRL scores have higher mean rank on all exams and the final course grade compared to students with low SRL scores. According to Cohen (1988), the effect sizes of this difference are small.

Table 2.6: Spearman's Correlation Between Students with High and Low SRL Scores in Academic Performance Within Two Group of Consenting Students: All Students and All Disengaged Students.

Performance Variables	Spearman's correlation for all students (n = 376)		Spearman's correlation for disengaged students (n = 117)	
Exam 1	$r(374) = 0.14$	$p = 0.005$	$r(115) = 0.29$	$p = 0.002$
Exam 2	$r(374) = 0.16$	$p = 0.002$	$r(115) = 0.36$	$p < 0.0001$
Exam 3	$r(374) = 0.17$	$p = 0.001$	$r(115) = 0.36$	$p < 0.0001$
Final Exam	$r(374) = 0.19$	$p = 0.0001$	$r(115) = 0.38$	$p < 0.0001$
Course Grade	$r(374) = 0.23$	$p < 0.0001$	$r(115) = 0.35$	$p < 0.0001$
Total	376	26.3%	117	41%

When considering only students that were considered disengaged with the Calculus I course early in the semester, Spearman's non-parametric correlation reported statistically significant differences between students with high and low SRL scores on all course exams and the final course grade as well, with $r(115) \geq 0.29$ and $p \leq 0.002$ for all correlations. Positive correlation coefficients suggest that disengaged students with high SRL scores have higher mean rank on all exams and the final course grade compared to disengaged students with low SRL scores. According to Cohen (1988), the effect sizes of this difference are medium.

Lastly, I compared mean SRL scores with across two measures: students' behavior in the Calculus I course with (1) online homework and (2) help seeking in the MTC. Data across these groups is provided in Table 2.7, below.

Table 2.7: Mean SRL scores of Students According to Online Homework System Usage and Mathematics Tutoring Center Usage.

		Online Homework System Usage		
		Used	Did not use	Total
MTC Usage	Visited	3.60 (n = 50)	3.16 (n = 25)	3.45 (n = 75)
	Did not Visit	3.52 (n = 184)	2.81 (n = 117)	3.24 (n = 301)
	Total	3.54 (n = 234)	2.87 (n = 142)	3.28 (n = 376)

Of these four groups, the disengaged students (i.e., those who had not been to the MTC or worked on their Calculus I course online homework as of week four) had the lowest mean SRL score (2.81), while students who both sought help and used the online homework had the highest mean score (3.6). Those who only worked on online homework had slightly higher mean SRL score (3.52) than those who only sought help (3.16). A Kruskal-Wallis non-parametric test verified that these four behavioral groups differ in mean rank SRL score, $\chi^2(3) = 15.0625$, p =

0.013. Post Hoc Dunn's test with FDR correction revealed that the mean rank of disengaged students is statistically lower than students who only engage in the Calculus I online homework before Exam 1, $z = -3.67$, $p = 0.0018$, $r = 0.21$. This provides some evidence that the SRL score is capturing evidence of students' self-regulation.

Discussion

These findings show promise for being able to use an SRL framework to develop tools that measure the degree to which students' behaviors are suggestive of SRL. Using these tools, I discussed a method for generating SRL scores for students by analyzing their behaviors, specifically those around prerequisite remediation and readiness for Calculus I. This was achieved by leveraging data from students' interactions with online tools aimed to support students in preparing for Calculus I through assessing their precalculus content knowledge. In addition to capturing data about students' SRL around precalculus material, I hypothesize that this score would be indicative of students' SRL around calculus material in the course.

In addition to measuring evidence of students' engaging with SRL, the SRL score particularly seems to benefit identification of which disengaged students may be at-risk of failing Calculus I. The relationship between the SRL score and academic performance metrics suggests that more self-regulatory behaviors around prerequisite material correlate with higher course performance, which aligns with what is seen in the literature (Labuhn et al., 2010; Zimmerman et al., 2011; Zimmerman & Schunk, 2001). Assuming that students who act in a self-regulated way around prerequisite material will act in a self-regulated way in the course, one would expect the positive effects from self-regulating one's learning in the beginning of the semester to compound as the semester progresses, which would agree with these stronger statistical relationships with later exams found in the data.

Additionally, the relationship between the SRL score and achievement in Calculus I grows stronger when considering only those students who are disengaged with Calculus I before their first exam. Students who are disengaged but have a higher SRL score tend to have higher success rates in the course than those who are disengaged with lower SRL scores. Similarly, when looking across all students (not just those who are disengaged), students who have higher SRL scores tended to be successful in Calculus I (i.e. receiving a grade of A, B, or C) on average, providing additional evidence of benefit of self-regulating one's own learning (Broadbent & Poon, 2015).

The SRL score provides a metric for those in positions to support students (e.g., instructors, course coordinators, administrators) to begin identifying students who may be at-risk of failing a course. For example, this study used a threshold of three on the SRL score to categorize students who provided evidence of self-regulating to a 'high' or 'low' degree. This work found differences in Calculus I course success can potentially be detected early. Such analysis can be done with all students but seems to be even more telling for students that are disengaged. With information about students' self-regulation around precalculus remediation, instructors can potentially intervene with students before their behaviors and poor study habits negatively impact their mathematical understanding and course performance. Such interventions could involve focusing on students' calibration of content knowledge, promoting available academic supports such academic help centers, or enrolling in a single credit course where students learn about SRL and apply this process throughout their courses.

Limitations and Future Direction

In the present work, the presence of the different phases of SRL was determined by whether or not the students used the designed online tools. This method relies on students' understanding the purpose for each tool and makes the assumption that lack of use is a conscious effort to avoid the tool and the associated SRL phase. Students were informed of the reasoning for the online resources in place during recruitment for the study, so it is unlikely that students used the online resources for other reasons, however, I recognize students whom are engaging in the SRL process may not be using the provided tools, and hence would have lower SRL scores that are accurate. The SRL score only measures the degree to which students engage in behaviors suggestive of engaging with SRL for which evidence is available. To begin providing evidence for other means of remediating, open-ended surveys could be implemented asking students what resources they currently utilize, and what resources they plan to use throughout the semester. This would provide an avenue for hypothesizing whether or not students use resources whose usage are not monitored.

To further improve the reliability of these resources, I plan to adjust the tools in three ways: (1) merge multiple tools into one, (2) provide options where students can skip to particular aspects of the tools, (3) utilize data in the PRT to inform the SRL score, and (4) include questions about students strategy usage when engaging with the PCQ. The first adaptation would both encourage students to engage in the SRL cyclic process, so that fewer tools can be used to capture an entire SRL cycle. By providing options whereby students can bypass questions, tool usage can begin to provide evidence of students actively avoiding a phase of self-regulation. Incorporating more data from students' responses (such as from the PRT) in the SRL score would afford making a more nuanced metric, where, for example, data can be used to determine if a student is following through with their plans. Another means to improve the reliability of the SRL score would be to pose questions regarding how students navigated through the PCQ to better understand how they are self-regulating during the performance phase of SRL, such as by implementing strategies.

Another future direction for the present work is to continue SRL analysis throughout the semester around different tasks in Calculus I, such as around exam preparation. The SRL score described in this paper is a snapshot of the degree to which students are engaging in SRL during the Calculus I course. By including additional resources throughout the semester, the SRL score presented in this paper can act as an initial SRL score, which can be updated as students engage with other online tools in self-regulated ways. By measuring SRL throughout the semester, temporal aspects of students' self-regulation may vary throughout the semester, and these changes can be investigated and correlated with students' success.

Based on student interactions with online tools and external resources, I am gaining insight in the role SRL plays in students' resource usage. While this data combined with the SRL framework is informing modification, enhancements, and additional of online tools, the method does not take into account the student perspective. I believe that this evidentiary approach is a strength of this work, however conducting interviews with students (e.g., using Zimmerman and Martinez-Pons' (1986) structured interview protocol) affords triangulating what I am seeing in the data with other qualitative methods for measuring SRL. Student interviews would provide an opportunity to better understand other ways students engage in the course as well as validate our quantitative findings. Additionally, SRL scores in conjunction with qualitative interview data can then inform intervention support to improve student success in Calculus I. For example, educators and departmental leaders could present students with the data that identified them as at-risk. Students could provide their opinion on their own actions indicated in the data, and the interviewers and students could discuss productive ways for the student to proceed in the course and be successful. With the quantitative and qualitative data, departments can suggest to students an array of support mechanisms for each student that has been identified as at-risk.

Chapter 3

Paper 2: Improving Effectiveness of Help-Seeking: A Study Supporting Student Success in Calculus I

Calculus I is a vital mathematics course for many STEM disciplines, providing foundational tools for studying the discipline and being a requirement for graduation. Low pass rates in college calculus courses have been documented in the literature for decades (Blair et al., 2013; President's Council of Advisors on Science and Technology (PCAST), 2012; Sonnert & Sadler, 2014; Steen, 1987), giving the Calculus I course the title 'gatekeeper' or 'filter' for college STEM disciplines. Many efforts have been made to improve course content (Thompson & Ashbrook, 2019) and to identifying particular aspects of successful calculus programs (Bressoud et al., 2013, 2015; Bressoud & Rasmussen, 2015; Hagman, 2019). While efforts for improvement vary in aim and scale, the importance of improving undergraduate calculus and students' success in the course is clear.

Factors Related to Success in Undergraduate Mathematics

Frequently cited reasons for student difficulty in Calculus I include prior mathematical success, prerequisite knowledge, and student preparedness for college calculus courses (Agustin & Agustin, 2009; Carlson et al., 2015; Carlson, Oehrtman, et al., 2010; Champion & Mesa, 2018; Murray, 1931; Sadler & Sonnert, 2018; Sonnert & Sadler, 2014). Almost a century ago, Murray (1931) identified that many students enrolled in calculus have difficulty with fractions, algebra, radicals, and exponentials. Additional research suggests that students' difficulties with precalculus content continue to hinder students' success in Calculus I. Agustin and Agustin

(2009) provide evidence that students struggling on exams in Calculus I course may be struggling with precalculus content to a higher degree than calculus content. They found that students struggling on Calculus I exams tended to lose more points on computational problems from mistakes on precalculus content than from mistakes on calculus content. Sonnert and Sadler (2014) suggest that while there is doubt as to whether taking a precalculus course improves the success of students enrolled in subsequent calculus courses, students' mathematical preparation for college calculus courses is correlated with success. Sadler and Sonnert (2018) further found that a high level of mastery of prior mathematical content can improve a student's predicted success in calculus by double what can be expected from having taken calculus in high school. All of these results highlight the importance of a high level of mastery in precalculus content when entering undergraduate calculus.

Mathematics content knowledge is not the only reason students may have difficulty in undergraduate calculus, however. Students must be motivated to learn course content and, if necessary, re-evaluate their content knowledge to determine whether they need to enrich impoverished understandings of present and past content. Studies have found that students' motivation and dispositions towards learning a content area can be related to success in mathematics (Champion, 2009; Ironsmith et al., 2003; McKenzie et al., 2004; Pajares & Graham, 1999; Pajares & Kranzler, 1995; Pajares & Miller, 1994; Pintrich & De Groot, 1990; Thanheiser et al., 2013; Worthley et al., 2016). Ironsmith and colleagues (2003) found that students who were more confident tended to perform better in developmental mathematics courses, and that students whose goals for the course were learning based performed better than students whose goals were performance based. Conversely, overconfidence in mathematical judgments can hinder students' performance through avoidance of beneficial practices such as help-seeking

(Ferla et al., 2010). Even with prospective elementary teachers, Thanhesier and colleagues (2013, 2014) found that preservice teachers believed that their procedural understanding of mathematics was sufficient for teaching elementary mathematics students, and thus viewed mathematics content courses as superfluous and "annoying prerequisites" (2014, p.234).

Students who are entering Calculus I may view reviewing ideas from precalculus and algebra similarly to those prospective teachers. Having already passed a precalculus course, students have some evidence that they are sufficiently prepared for Calculus I. However, students' perceived preparation may not be as accurate as they believe in the beginning of the course. The *Insights and Recommendations from the MAA National Study of College Calculus* suggests a similar story with students enrolled in calculus: Bressoud (2015) reports that while 81% of the 7440 university students believed they were ready for college calculus, only 56% of the 3664 students who passed the course reported (after completing the course) that they had been prepared for calculus at the beginning of the semester. These findings point to a discrepancy between students' preparedness for calculus and students' perceived preparedness for calculus. The alignment of one's perceived mastery and observed mastery of a content area is called *calibration*.

This paper aims to understand the role of calibration of Calculus I students' prerequisite content knowledge with success in Calculus I, as well as calibration's role in the effective use of a calculus help center (CHC). Specifically, this paper aims to answer the following questions:

1. How do students' calibration of precalculus content relate to their first exam scores in a Calculus I course?
2. How does students' calibration of precalculus content and in-person help-seeking at a mathematics learning centering together relate to their first exam scores in a Calculus I course?

Literature Review

To refine the research questions of this study, I will first discuss metacognition to highlight how calibration fits into the larger theory of metacognition. I then describe calibration, highlighting some key findings in the literature, as well as how calibration has been measured in the past. To connect calibration to help-seeking, I then discuss tutoring centers in mathematics to connect help-seeking with student success, concluding with the specific research sub-questions addressed in this study.

Metacognition

Metacognition is often characterized as knowledge of what one does and does not know. While used by many researchers (e.g., Flavell, 1979; Garofalo & Lester, 1985; Pintrich, 2002; Schoenfeld, 1992; Schraw, 1998), metacognition is often described in terms of two interrelated components: knowledge of cognition and regulation of cognition. Knowledge of cognition is described as what one knows, while regulation of cognition describes in what ways one acts or cognitively processes information (Garofalo & Lester, 1985). Metacognitive knowledge (a component of knowledge of cognition) is described in terms of three variables: the person, the task the person is engaging in, and the strategies the person can use during the task (Flavell, 1979; Pintrich, 2002). Self-knowledge is a particularly important aspect to metacognitive knowledge with regard to academic success (Pintrich, 2002). For example, a student may understand that their content knowledge on solving quadratic equations is sufficient for an upcoming exam, while their knowledge of graphing quadratic functions is inadequate. This assessment of one's knowledge may cue the student to enact strategies (such as reading a textbook or self-testing) to improve their knowledge about graphing quadratic equations prior to the upcoming exam. For this reason, metacognitive knowledge is important for learning and student achievement. Without accurate self-knowledge however, a student may not recognize what content they have not yet mastered, thereby limiting cues to improve one's content knowledge.

Calibration

When considering the student who judged their determined that their knowledge of graphing quadratic functions as insufficient for an upcoming exam, they are drawing on their metacognitive knowledge to determine when and how to regulate their cognition, and it is vital that the student's self-knowledge/judgment is accurate (Pintrich, 2002). *Calibration* can be thought of as a component of metacognitive knowledge, as calibration is the degree of alignment between a person's judgment of their performance on a task and their actual performance on the task (Bol & Hacker, 2001; Labuhn et al., 2010; Winne & Jamieson-Noel, 2002). Research suggests that calibration is correlated with student success (Bol & Hacker, 2012; Champion, 2009; Kline & Dibbs, 2018; Labuhn et al., 2010; Pajares & Graham, 1999; Pajares & Miller, 1994; Schraw et al., 1993; Sheldrake et al., 2014; Tian et al., 2018). Recent work by Wakefield and colleagues (2018) suggest that data about students' prior mathematical ability, such as high school rank and standardized mathematics exam scores, in conjunction with a local course readiness activity for a given course can be used to help identify students' risk of failing the course. The local readiness activity described involved a precalculus content quiz, which students could then retake to earn a better score. By looking at students' success with the readiness activity, the researchers were also gathering information related to students' motivation, current mastery of prior content, and productive reactions to cues about insufficient content knowledge (as determined by scores on the course readiness quiz). By retaking the quiz, students are likely adjusting their inflated perceptions of mastery for particular material based on their feedback from prior quizzes in order to identify content areas that warrant review

When looking at undergraduate students enrolled in developmental and introductory mathematics courses, Zimmerman and colleagues (2011) found that students who underestimate their abilities tended to perform better than students who overestimate. They found that while an instructional intervention did not seem to decrease students' tendency to overestimate their abilities, they did find evidence that an intervention focusing on self-reflection may help students become more accurate in their self-judgments. This work provides evidence that students' calibration (and metacognitive monitoring) is teachable (Desoete & De Craene, 2019; Ramdass & Zimmerman, 2008; Shilo & Kramarski, 2019). Research suggests that calibration is correlated with student success (Bol & Hacker, 2012; Champion, 2009; Kline & Dibbs, 2018; Labuhn et al., 2010; Pajares & Graham, 1999; Pajares & Miller, 1994; Schraw et al., 1993; Sheldrake et al., 2014; Tian et al., 2018). In addition, research suggests the relationship between student achievement and metacognitive knowledge (e.g., calibration) may be mediated by self-efficacy (Tian et al., 2018).

Measurement of Calibration

The calibration literature has generally found that students who are either more accurate in their perceptions or tend to underestimate their abilities have higher achievement on average than those who are less accurate or tend to overestimate (Bol & Hacker, 2012; Champion, 2009; Kline & Dibbs, 2018; Labuhn et al., 2010; Pajares & Graham, 1999; Pajares & Miller, 1994; Schraw et al., 1993; Sheldrake et al., 2014; Tian et al., 2018). This is found by measuring calibration in terms of two measures: calibration accuracy and calibration bias. *Calibration accuracy* describes magnitude of the difference between the perception of mastery and the observed mastery, while calibration bias describes the degree to which the perceptions of mastery are higher than their actual mastery or the direction of the judgement errors (Bol & Hacker, 2012; Labuhn et al., 2010; Pajares & Graham, 1999). *Calibration bias* is often used to quantify the degree to which students overestimate (positive calibration bias) or underestimate (negative calibration bias) their abilities, while calibration accuracy describes the magnitude of these underestimations and overestimations. Methods for collecting data around calibration have varied through the literature (Lingel et al., 2019), though they all involve (a) a measure of students' self-efficacy beliefs (or judgments of confidence to successfully perform a task; Bandura, 1997) for correctly answering a type of question and (b) observations of students' performance on a task . Further variance in measurement of calibration can come from the means of data collection regarding students' self-efficacy judgements (e.g., question format, predictions vs. postdictions), as well as how calibration is computed from this information.

As an example, we describe a method for computing these measures of calibration bias and accuracy described by Pajares and Graham (1999), and implemented many times in the literature (e.g., Labuhn et al., 2010), using measures of self-efficacy to correctly answer items on a quiz (a prediction), and whether or not the student correctly answers the items. This methodology informs the methods used in the study presented in this paper.

In this method, students are given multiple-choice mathematics questions, where prior to solving each mathematics questions, students' must rate their self-efficacy for answering the preceding mathematics questions correctly. Measures of students' self-efficacy (called their prediction scores) are then averaged across all items to create an average self-efficacy score. Students' performance on each item is then scored along the same scale as the self-efficacy score. For example, if self-efficacy could range from 1 to 7 on each item, then the student students' performance on each item would be scored from 1 (incorrect) to 7 (correct). Calibration bias is then computed by subtracting the performance score from the self-efficacy score, giving a scale centered at 0 (e.g., a scale from -6 to 6). Calibration accuracy is then computed by computing the absolute value of calibration bias and reversing the resulting scale. For example, if calibration bias ranges from -6 to 6, then subtracting the absolute value calibration bias from 6 would produce a scale from 0 to 6. Using this method, high values in calibration accuracy correspond to values of calibration bias near 0 and low values of calibration accuracy correspond to values of calibration bias that are far farther from 0.

This method of calculating calibration bias and accuracy can be summarized as follows: (a) an item-wise calibration bias by subtracting the students true score from the students predicted score, (b) the calibration bias by averaging the item-wise calibration biases, and (c) the calibration accuracy by computing the absolute value of the averages of the item-wise calibration biases as described in Table 3.1.

Table 3.1: A Description of Measures of Calibration Bias and Accuracy Used in the Literature

Measure	Computation
Calibration Bias (Item-wise)	Observed Score – Predicted Score
Calibration Bias (General)	Average of (Item-wise) Calibration Bias
Calibration Accuracy (General)	Absolute Value of (General) Calibration Bias

This perspective recognizes the individual biases that students have on each question (here, called *item-wise calibration bias)*. By averaging students' item-wise calibration bias, one loses variance in one's judgments that may be present across various content. For example, a student who has maximal calibration bias on half of the content items and minimal calibration bias on the remaining half would have general calibration bias of 0, hence giving a maximal general calibration accuracy. This measure of calibration accuracy would seem to suggest that the student would show perfect accuracy on their self-efficacy judgments, however this measure is misrepresentative as the student has been wrong across all self-efficacy judgments. The accuracy of a student's self-judgments is important for their learning as these perceptions likely indicate to the student whether their mastery is sufficient for their coursework or inadequate, thereby cueing the student to improve their content knowledge through studying or seeking help with an instructor or at an institutional tutoring center.

Tutoring Centers

Tutoring centers have been described as an important resource for students enrolled in undergraduate mathematics course. The *Characteristics of Successful Programs in College Calculus* project identified tutoring centers and support mechanisms for students as one of seven characteristics of successful programs common among the studied programs (Bressoud & Rasmussen, 2015). The researchers found tutoring centers provides support for students struggling with course work and can provide a community for students. Additional evidence suggests that students who visit tutoring centers are expected to have higher exam and course outcomes (Byerley et al., 2018; Rickard & Mills, 2018). Byerley and colleagues (2018) found that after controlling for incoming ability and other factors, university students enrolled in Calculus II who visit a tutoring center more frequently are more likely to pass the course than students who visit the center less. Rickard and Mills (2018) reported a similar finding that university students enrolled in Calculus I who visited the testing center more frequently tended to have a higher course grade score than students who had visited the tutoring center less frequently. Further, Rickard and Mills provide some evidence for differential effects for students visiting the tutoring center. When predicting final grades using ACT math scores, visits to tutoring center, high school math GPA and an interaction effect between visits to tutoring center and high school GPA, the resulting multiple regression model accounted for 35% of the variance of final scores. The interaction effect between visits and high school GPA alone accounted for 3.5% of the variance in the final scores in Calculus I, which would suggest that students with lesser high school math GPA are expected to benefit from help-seeking at a tutoring center more than higher achieving students.

Like Byerley et al. and Rickard and Mills, this study also examines differential effects of visiting tutoring center, but instead looks at how students' calibration may mediate the impacts of visiting the tutoring centers. From a metacognition perspective, visiting a tutoring center can support students as they attempt to regulate their cognition. Metacognition theory (Flavell, 1979; Pintrich, 2002) anticipates the possibility of interactions between visiting a tutoring center and one's calibration, and the present study aims to better understand the relationship between calibration, visiting a help center (or help-seeking more generally), and academic achievement.

Research Questions

In light of the research on metacognition, calibration and help-seeking, the overarching research questions were operationalized by using calibration bias and calibration accuracy to measure students' calibration. The following research questions and sub-questions are hence addressed:

1. How do students’ calibration of precalculus content relate to their first exam in a Calculus I course?
    a. How do students’ calibration bias and calibration accuracy of precalculus content at the beginning of a Calculus I course relate to students’ first exam scores?
    b. How do students’ calibration bias and calibration accuracy of precalculus content at the beginning of a Calculus I course relate to students’ first exam scores after accounting for incoming mathematical mastery?
2. How does students’ calibration bias and calibration accuracy of precalculus content at the beginning of a Calculus I course and in-person help seeking at a calculus help center in a Calculus I course relate to their first exam score in the Calculus I course?
    a. How do measures of calibration bias and calibration accuracy mediate the relationship between students’ performance on the first exam and students’ number of attendances to the calculus help center?

Methods

Data for this study came from a first-semester calculus course (Calculus I) at a large, predominantly white, western university in the United States during the fall semester of 2016. The course reported in this study is one of three courses that could be considered a first-semester calculus courses available to students at the institution. The particular course primarily serves students seeking STEM majors, such as engineering, physics, mathematics, statistics, computer sciences, and chemistry. The other Calculus I courses offered at the university are specific for biology students and business students. The course is a large multi-section coordinated course with common homework, exams, and final exam. For brevity, this course will be referred to as *Calculus I*. Data for this study were collected as a part of a larger study investigating relationships between students’ activity during the course and pass rates. Data regarding students’ use of particular resources and prior mathematical knowledge were used to conduct a hierarchical regression analysis to both control for students’ incoming ability and examine the relationships of calibration and help-seeking with Exam 1 performance.

Data Sources

In the first two weeks of the Calculus I course, all students had the opportunity to engage with a variety of online resources built and housed in their learning management system. All resources were optional and only available in the first two weeks of the course. Availability for the resources was staggered to suggest a particular sequence of engagement with the resources. Each resource focused on precalculus concepts that were deemed important for success in Calculus I, as determined by the local course-coordinator and experienced instructors of the course. Students who engaged with all of the resources (1) responded to a questionnaire about their perception of mastery of precalculus content, (2) completed a quiz on precalculus content, and (3) responded to a questionnaire regarding their plans to potentially remediate their mathematical skills and understandings of precalculus content. During the recruitment for this study, students were informed of the online resources made available to them through their learning management system, as well as the purpose of them. While students were encouraged to use these online resources, students were explicitly told that the resources were optional and that students' use or non-use of the resources did not directly impact their course grade. To address the research questions for this study, data from two of the online resources were used to measure students' calibration bias and accuracy of precalculus content, namely the precalculus self-assessment and the precalculus content quiz.

Precalculus Self-Assessment

The first tool available to the students was a survey asking students to rate their confidence in precalculus self-assessment (PSA). This was an optional 16-item assessment where students were asked to rate their ability to answer an item about particular precalculus concepts correctly on a 5-point Likert scale from 1 to 5 (*no confidence, little confidence, average, confident,* and *very confident*).

Precalculus Content Quiz

The next resource that became available to students was a precalculus content quiz (PCQ). This 12-item optional quiz asked students a variety of questions vital to precalculus. Question types on the precalculus content quiz were either multiple choice (nine questions) or multiple answer (three questions). Multiple-choice questions were either marked as correct or incorrect, while multiple-answer questions were marked as either incorrect (at least one incorrect response was selected), partially correct (only correct answers were selected, though not all of them), or correct (all correct answers were selected, and no incorrect responses were selected).

The PCQ and PSA were developed in the previous academic year and was refined through the fall semester of 2016. Content for the PCQ was based on past instructors' experiences with students, common mistakes from students' work on past exams, and the course coordinator's expertise in the area. Precalculus topics for the PCQ and PSA included function notation, graphs of functions, simplifying algebraic expressions, trigonometry, and solving algebraic and trigonometric equations. Unlike the PCQ, the PSA was intended to ask students to make judgments about their mastery of content related to the PCQ but asked more generally.

Incoming Mathematics Mastery

Student's American College Test (ACT) Math scores and Scholastic Aptitude Test (SAT) Math scores were requested and collected from the university's institutional research group. For students without an SAT Math score, their ACT Math score was converted to the SAT Math equivalent score (Dorans, 1999). For those students who had both SAT Math and

ACT Math scores, ACT Math scores were converted to SAT Math equivalent scores and the maximum of the SAT Math and the SAT Math equivalent scores were used as the student's SAT/ACT Math score in the analysis. Students whose SAT Math and ACT Math scores were not available were omitted from the study.

Calibration Bias and Calibration Accuracy Scores

Items on PSA and PCQ were not intended to match identically. One reason for this was to limit the number of mathematical questions on the PCQ. It was hypothesized that adding more questions on the PCQ would reduce the number of students that would voluntarily complete it. Another reason the PSA and PCQ items did not match identically was due to similarities in ideas on PSA items. For example, two items on the PSA asked students to rate their confidence in recognizing incorrect polynomial expansion (PSA item 4, see Figure 3.1) and incorrect factoring (PSA item 5, see Figure 3.2) in someone's work. However when examining whether such an error has been made, in for example the statement $(x-3)^2 = x^2 - 9$ (PCQ item 3, see Figure 3.3), the error made could be viewed as either incorrect polynomial expansion or factoring. To reduce redundancy and the number of items in the PCQ, only one item (PCQ item 3, see Figure 3.3) was used.

Q4: How confident are you that you can recognize the following errors in someone's work?

Incorrect polynomial expansion

(1) No Confidence, (2) Little Confidence, (3) Average Confidence, (4) Confident, (5) Very Confidence

Figure 3.1: Item 4 on the PSA

Q5: How confident are you that you can recognize the following errors in someone's work?

Incorrect factoring

(1) No Confidence, (2) Little Confidence, (3) Average Confidence, (4) Confident, (5) Very Confidence

Figure 3.2: Item 5 on the PSA

Q3: The statement $(x-3)^2 = x^2 - 9$ is:

(a) Correct, (b) Incorrect, or (c) I don't know

Figure 3.3: Item 3 on the PCQ

When using items on the PCQ and the PCA to measure calibration bias and calibration accuracy, a process of mapping PSA items to each PCQ item was developed. Items on the PSA were assessed as to how relevant the content area of the PSA item aligned with each item on the PCQ. For each PCQ item, PSA items were coded as either not related to the content quiz item (coded as 0), partially related to the content quiz item (coded as 1), or strongly related to the content quiz item (coded as 2). As many topics in mathematics are related, this coding method was intended to capture additional relationships between each PCQ item and items on the PSA that were not initially intended during the creation of the assessments. The author and the course coordinator coded the relationships between each PSA item and each PCQ item independently and discussed each code until agreement was reached.

For example, item 1 on the PCQ is provided in Figure 3.4. This item focuses on graphing and function notation, and hence was related to items 1, 2, 3, 15, and 16 on the PSA. Of those items, question 15 on the PSA was determined to have a strong relationship with item 1 on the PCQ. Hence the relationship between item 1 on the PCQ and item 15 on the PSA was coded as '2', while the relationship between item 1 on the PCQ and items 1, 2, 3, and 16 were coded as '1'. The relationship between item 1 on the PCQ and other items on the PSA were coded as '0'. The relationships between the PSA items and PCQ items are summarized in Appendix B.

Q1: Consider the function $g(x) = -3x + 2$.

What does g(2) represent? (Mark all that apply)

a. The function gets multiplied by 2.
b. The function evaluated at 2.
c. The y-value on the graph of the function with x-coordinate 2.
d. The x-value on the graph of the function with y-coordinate 2.
e. The height of the graph of the function at $x = 2$.
f. The distance between the graph of the function at $x = 2$ and the x-axis.
g. The distance between the graph of the function at $x = 2$ and the y-axis.
h. The slope of the graph of the function at $x = 2$.
i. $-3(2) + 2$.

Figure 3.4: Item 1 on the PCQ

For each PCQ item, a weighted average of each student's responses to the PSA items was computed to create a confidence score. This weighted average was based on the coded relationships between the PSA and PCQ items and was used to create a confidence score for each PCQ item. For example, the confidence score for item 1 on the PCQ was computed by adding students' responses to items 1, 2, 3, 15, and 16 on the PSA together after first multiplying the response to item 15 by 2. This sum was then divided by 6 to produce the weighted average. The formula for computing the confidence score for item 1 on the PCQ is provided below, where SA_k represents the student's response to item *k* on the PSA.

$$\text{Confidence score for item 1 of the PCQ} = \frac{SA_1+SA_2+SA_3+2\cdot SA_{15}+SA_{16}}{1+1+1+2+1}$$

Similarly to each PSA item, each PCQ item's confidence score can vary from one to five, where five represents high confidence of success in the content area of the PCQ item, one represents a low confidence of success in the content area of the content quiz item, and three represents a mid-level confidence in the content area.

Measuring Calibration Bias and Accuracy. The method for computing average calibration bias score used in this study is similar to those used by Schraw et al. (1993) and Labuhn et al. (2010) to compute (general) calibration bias. Students' performance on each PCQ item was first scaled to match the scale of the PSA, where students' responses to the PCQ were assigned either a score of 5 (for a correct response), 3 (a partially correct response), or 1 (an incorrect response). To create the item-wise calibration bias score, the performance score (1 through 5) was subtracted from the self-efficacy score (1 through 5), resulting in a score potentially ranging from - 4 to 4. The sign of the calibration bias score for a given item represents the directionality of a student's judgment errors, where a negative calibration bias score suggests the student underestimated their abilities, a positive calibration bias score suggests the student overestimated their abilities, and a calibration bias scores of 0 suggests that the student is accurate in their judgement abilities.

The methods employed in this study differ from those of Schraw et al. (1993) and Labuhn et al. (2010) in the computation of calibration accuracy. In this study, an item-wise calibration accuracy score is assigned to each PCQ item by first computing the absolute value of the item-wise calibration bias scores (from 0 to 4) for each content quiz item and then reversing the scale to create a scale from 1 to 5. Item-wise calibration accuracy scores were then averaged to obtain the average calibration accuracy scores. This choice was made to preserve the variance in students' biases. In the previous methods for computing calibration accuracy (e.g., Labuhn et al., 2010; Pajares & Graham, 1999), a student who over predicts their performance for half of the items and under predicts the remaining half would have an average calibration bias score near 0, hence having a relatively high calibration accuracy score. The approach taken in this study would have the same calibration bias score but would have a low calibration accuracy since each prediction on each item was inaccurate.

For example, a student (Student A) who answered a content quiz item correctly (performance score of 5) and who responded with the highest self-efficacy on all self-assessment items that were related to that particular content quiz item (self-efficacy score of 5) would have a calibration bias score of zero for the content quiz item. If another student (Student B) correctly answered the item (performance score of 5) yet responded with no self-efficacy for all self-assessment items related to the content quiz item (self-efficacy score of 1), the student would have a calibration bias score of negative four for that the content quiz item. Similarly, a student who answered a content quiz item incorrectly (performance score of 1) and exhibited the highest self-efficacy on self-assessment items related to the content quiz item (self-efficacy score of 5) would have a calibration bias score of 4 on the particular content quiz item. The computation for the item-wise calibration bias and accuracy scores are described in Table 3.2. These item-wise calibration bias scores were then averaged together to obtain an average calibration bias score for each student's precalculus content mastery. For ease of language, the average calibration bias score will be referred to as the 'Calibration Bias score' or 'Calibration Bias.'

Table 3.2: Item-wise Calibration Bias and Calibration Accuracy Computations for Hypothetical Students

	Self-Efficacy Score for PCQ Item	Performance Score for PCQ Item	Item-wise Calibration Bias	Item-wise Calibration Accuracy
Student A	5	5	5 - 5 = 0	5 – 0 = 5
Student B	5	1	5 – 1 = 4	5 - 4 = 1
Range	1 to 5	1 to 5	-4 to 4	1 to 5

Students' calibration accuracy on each content quiz item was also created by subtracting the absolute value of the calibration bias score of each content quiz item from 6, resulting in a score ranging from 1 to 5. For each item, the calibration accuracy score represents how accurate the student's self-efficacy was to the student's performance on the content quiz item, where 5 represents perfect accuracy between their self-efficacy score and their ability, and 1 represents complete inaccuracy between their self-assessment and their abilities. For example, students who received a calibration bias score of 0 would have a calibration accuracy score of 5, and students with calibration bias score of - 4 or 4 would have a calibration accuracy score of 1 for the particular content quiz item. For each student, the item-wise calibration accuracy scores were averaged to obtain an average calibration accuracy score for each student's precalculus content mastery. For ease of language, the average calibration accuracy score will be referred to as the 'Calibration Accuracy score' or 'Calibration Accuracy.'

Calibration bias and calibration accuracy variables will further be normalized in the subsequent regression models for more easier interpret the regression models. Hence in the subsequent models, a calibration bias score of 0 will refer to the mean calibration bias score for the sample, and a calibration accuracy score of 0 will refer to the mean calibration accuracy score for the sample.

Calculus Help Center

In addition to the calibration bias, calibration accuracy, and SAT/ACT scores, students' attendance to the university's Calculus Help Center (CHC) was recorded. Office hours of instructors and teaching assistants were held in the CHC. Students could freely visit the CHC anytime during the workday for scheduled appointments with their instructor or for drop-in appointments with an instructor or teaching assistant of a different section of the course. During this semester, 57.5% of the students enrolled in Calculus I visited the CHC at least once. On average, a student who visited the CHC visited 7.4 times throughout the semester. Of the students involved in this analysis, 65.5% of the students visited the center at least once, and students visiting the CHC on average visited 6.8 times throughout of the semester.

Study Sample

Of the 426 students enrolled in the course, 401 students consented to all aspects of the study. This study intended to only investigate students who were engaged throughout the entire course, and therefore included only students who completed all exams in the course. Students included in the sample for this analysis included students who (1) completed both the PSA and PCQ, and (2) had SAT Math or ACT Math scores accessible through institutional research. Two additional students were identified as outliers with respect to exam scores or number of visits to the CHC. The final sample for analysis included 194 students. The process of determining the final sample can be found in more detail in Appendix C.

Results

Before investigating the relationship of student calibration of precalculus material with Exam 1 performance, student competency on precalculus and prior mathematics content (both prior ability and current ability) was examined with regards to Exam 1 performance. Through bivariate linear regression, it was observed that SAT/ACT Math scores were positively correlated with Exam 1 scores, $r = 0.4792$, $r^2 = 0.2296$, $p < 0.0001$, suggesting that students with higher SAT/ACT Math scores tend to score higher on the first exam in Calculus I. Likewise PCQ scores were positively correlated with Exam 1 scores, $r = 0.3962$, $r^2 = 0.157$, $p < 0.001$, suggesting that students scoring higher on relevant precalculus questions in the first two weeks of the Calculus I course tend to have higher Exam 1 scores. This may not be a surprise for the reader familiar with precalculus content and limits of functions. The first exam primarily covers topics related to limits, and the evaluation of limits rely heavily on manipulation of algebraic expressions, understandings graphical representations of functions, and other past mathematical concepts. Likewise, many researchers have seen that both SAT and ACT scores are correlated with success in Calculus (e.g., Ellis et al., 2016; Sadler & Sonnert, 2018; Sonnert & Sadler, 2014), as are entrance exams and readiness exams (e.g., Carlson, Madison, & West, 2010, 2015; Carlson, Oehrtman, & Engelke, 2010; Wakefield et al., 2018).

When using multiple linear regression to understand how scores on the precalculus content and the SAT/ACT together relate to Exam 1 performance, both measures of past mathematical knowledge predict Exam 1 scores after accounting for the other variable. Even after accounting for SAT/ACT scores, hierarchical regression analysis revealed that student's scores on the PCQ added predictive power to the model for predicting Exam 1 scores, $r^2 = 0.2948$, $\Delta r^2 = 0.0652$, $p < 0.0001$. This suggests that student's SAT/ACT scores do not solely predict students' precalculus knowledge as of the Calculus I course, particularly that the students' performance on the PCQ accounts for variance in Exam 1 scores that wasn't accounted for by their SAT/ACT scores. This suggests that students' past performance in mathematics alone (as measured by the SAT/ACT math scores) does not tell the whole story of how precalculus and prior mathematics knowledge impacts success in Calculus I. One potential explanation for this is proximity of mastery of content relative to taking the calculus course. While many students in this study were freshman and sophomores, students may not have taken a mathematics course recently, so their SAT/ACT math scores may not be representative of their mathematics ability when enrolled in Calculus I. Students enrolled in a calculus course are drawing on their precalculus and prior mathematics content knowledge frequently when engaging in calculus content, such as reasoning about functions graphically, reasoning and solving algebraic equations and expressions, and making sense of new mathematical ideas using mathematical symbolizations such as function notation. Note that the SAT/ACT Math scores explains 10% of the variance in PCQ scores.

Calibration Measures Predicting Exam 1 Scores

To investigate the relationship between student's calibration bias and calibration accuracy of precalculus content knowledge and Exam 1 scores, a multiple linear regression was conducted. Descriptive statistics for calibration bias and calibration accuracy can be found in Table 3.3. Note that students in the sample on average tended to overestimate their abilities in precalculus content knowledge, though not to a significant degree.

Table 3.3: Descriptive Statistics of non-normalized calibration bias and accuracy

Variable	Minimum	Maximum	Mean	Median	Standard Deviation
Calibration Accuracy	2.861	4.882	3.722	3.718	0.394
Calibration Bias	-1.862	1.500	0.085	0.101	0.578

Note. $N = 194$.

Both calibration bias and calibration accuracy were correlated with students' Exam 1 scores. Calibration accuracy was found to be positively correlated with Exam 1 performance, $r = 0.2694$, $r^2 = 0.0726$, $p < 0.001$, which is considered to be a medium effect size. This suggests that students with one point higher average calibration accuracy are expected to score 8.3 points higher on Exam 1 than those students with a lesser average calibration accuracy. Note that the mean and standard deviation in calibration accuracy are 3.718 and 0.394, respectively, so students that are one standard deviation higher in calibration accuracy are expected to score 3.28 points higher on Exam 1 from the model. Calibration bias was negatively associated with Exam 1 performance, $r = 0.2313$, $r^2 = 0.0535$, $p < 0.001$, which is considered to be a small-medium effect size. This suggests that students with one point lesser calibration bias are expected to score

4.8725 points higher on Exam 1 on average. Note that the mean and standard deviation in calibration bias are 0.0849 and 0.5778, respectively, so students with one standard deviation lesser of calibration bias are expected to score on average 2.815 points higher on Exam 1. This suggests that students who tend to be more accurate in their perceptions of their mastery of precalculus content knowledge tend to perform better on exams, and that students who tend to underestimate their content mastery tend to perform better on Exam 1 than students that tend to overestimate.

In a multiple linear regression using both calibration bias and calibration accuracy as covariates to predict Exam 1 scores, the two measures of calibration account for 11.2% of the variance of Exam 1 scores (with 10.8% Adjusted R^2). For ease of interpretation of the multiple regression model (to follow), the covariates (calibration bias and calibration accuracy) were normalized. Further in the combined model, the directionality of the relationship between both calibration measures and Exam 1 performance persists. In the multiple linear regression model for predicting Exam 1 performance, the model suggests that Exam 1 performance can be predicted with the following model: Exam 1 Score = 77.1045 + 3.0805(Calibration Accuracy[1]) -2.5773(Calibration Bias[2]). All coefficients in the model are statistically significantly non-zero at the $p = 0.01$ level.

The resulting model revealed a negative correlation coefficient of Exam 1 score with calibration bias. This suggests that students who have lesser calibration bias tend to have higher Exam 1 scores, when calibration accuracy is fixed. Since negative bias corresponds to underestimations in one's precalculus knowledge and positive calibration bias corresponds to

[1] The variable 'calibration accuracy' has been normalized in this model, to have mean 0 and standard deviation 1.
[2] The variable 'calibration bias' has been normalized in this model, to have mean 0 and standard deviation 1.

overestimations, the model suggests that students who underestimate their abilities more tend to have higher Exam 1 scores than students who overestimate their abilities more. According to the model, students with one standard deviation lesser calibration bias score, on average, 3.0805 points higher on Exam 1 when levels of calibration accuracy are the same.

The model also has a positive correlation coefficient for calibration accuracy, which suggests that students who are more accurate in the PSA tend to perform better on Exam 1 than students whom are less accurate. Specifically, students who are one standard deviation higher in calibration accuracy tend to score 2.5773 points higher on Exam 1 scores when levels of calibration bias are the same.

Calibration Measures when Controlling for Incoming Content Knowledge

To investigate how calibration bias and calibration accuracy relate to Exam 1 scores after controlling for incoming ability, SAT/ACT Math scores were used to account for student's incoming ability. This choice was made to avoid multicollinearity given the high correlations between PCQ scores and calibration measures. PCQ scores alone account for 35% and 50% of the variance in calibration bias and calibration accuracy, respectively[3]. The SAT/ACT Math scores, however, showed minimal collinearity with calibration measures, calibration bias: $F(1,192) = 2.19$, $p = 0.14$; calibration accuracy: $F(1,192) = 16.79$, $p < 0.001$, $r = 0.28$, $r^2 = 0.08$. To quantify the potential impact of multicollinearity on the model, the variance inflation factor (VIF) was computed for each covariate in the model[4]. Variance inflation factors for calibration

[3] High correlations between these variables is not highly surprising as both calibration measures use students' item-wise scores on the PCQ.

[4] The variance inflation factor (VIF) quantifies the change in standard error in a regression model based on correlations between one covariate and the other covariates. The square root of the VIF corresponds to the multiplicative factor by which the standard error is expected to change. For example, a VIF of 4 suggests that the standard error in the regression model is at most $\sqrt{4} = 2$ times as large as if there were no multicollinearity in the

bias, calibration accuracy, and PCQ scores were 1.1, 1.0, and 1.1, respectively. This suggests that for each covariate, the standard error in the regression model is at most 1.049 times as large as if there were no multicollinearity in the model. This is significantly less than the variance inflation factor when PCQ scores are used to account for incoming mathematical ability in the model, $VIF_{\text{PCQ scores}} = 4.752$. While 4.752 is a still considered sufficiently low measure of VIF (Kutner et al., 2004; Sheather, 2009), SAT/ACT Math scores were chosen to account for incoming ability in the model due to the lower variance inflation factor[5].

Using hierarchical regression, SAT/ACT Math scores, average calibration bias, and average calibration accuracy were used to predict exam 1 scores. SAT/ACT Math scores were entered first into this model to control for incoming ability when predicting exam 1 scores. This initial model was significant, indicating that SAT/ACT Math scores were correlated with exam 1 scores, $F(1, 192) = 57.21$, $p < 0.0001$, $r^2 = 0.2296$, and that SAT/ACT Math score was a significant predictor of exam 1 scores, $\beta = 5.8332$, $t(192) = 7.562$, $p < 0.001$. This model suggests that students with a one standard deviation increase in SAT/ACT Math score on average score 5.8332 points higher on exam 1 than students with lesser SAT/ACT Math scores. Note that this is no surprise, as many sources have found that SAT and ACT scores are positively correlated with success in undergraduate course work (Sadler & Sonnert, 2018; Sonnert & Sadler, 2014).

---

model. A VIF of 1 corresponds to the same in the standard error in the regression model as if there were no multicollinearity, one times as much error corresponds to no increase in error.

[5] Note that some researchers suggest that there is not cutoff for 'high' and 'low' variance inflation factors (David et al., 2019). For this reason, the author chose to use the variable corresponding to the lesser variance inflation factor, in this case SAT/ACT Math scores.

In addition to SAT/ACT Math scores, calibration bias and calibration accuracy were then used as covariates in a multivariate linear regression model to predict Exam 1 performance in the step of the hierarchical regression analysis. Comparing the new model (with measures of calibration and SAT/ACT Math scores as covariates) with the prior model (only SAT/ACT Math scores as covariates) affords examining the predictive value of calibration measures after controlling for incoming precalculus ability. The model including the two measures of calibration into the model improved the prior model, $\Delta F(2, 190) = 6.5806$, $p = 0.0018$, $\Delta r^2 = 0.05$. In the new model, calibration bias was statistically significant, $\beta = -3.707$, $t(190) = -2.837$, $p < 0.01$, suggesting that, on average, students who are one standard deviation lesser in calibration bias are expected to score 3.707 points higher on Exam 1 per change in standard deviation of calibration, when calibration accuracy and incoming ability are fixed. Calibration accuracy was also statistically significant in the new model, $\beta = 4.204$, $t(190) = 2.118$, $p < 0.05$, suggesting that, on average, students who are one standard deviation higher in calibration accuracy are expected to score 4.204 points higher on Exam 1. Together, these three variables (SAT/ACT Math scores, average calibration bias, and average calibration accuracy) account for 29% of the variance in Exam 1 scores.

Hierarchical Regression Analysis

To investigate how measures of calibration and visits to the CHC impact Exam 1 scores, a four-step hierarchical multiple regression analysis was conducted where Exam 1 performance was the dependent variable. SAT/ACT Math score was entered in the first step of the model (Model 1) to control for variance in Exam 1 performance due to differences in SAT/ACT Math scores for later steps in the hierarchical multiple regression. The calibration variables (calibration bias and calibration accuracy) were entered at the second step (Model 2), the number of visits to the CHC before Exam 1 (CHC Visits) variable was entered at the third step (Model 3), and the interaction effects between CHC visits and each measure of calibration were entered at the fourth step of the model (Model 4). The regression statistics are included in Figure 3.4.

Table 3.4: Summary of Hierarchical Regression Analysis Predicting Exam One Performance

	Model 1		Model 2		Model 3		Model 4	
Variables	**B (SE of B)**	β	**B (SE of B)**	β	**B (SE of B)**	β	**B (SE of B)**	β
SAT/ACT Math	0.0854 (0.0113)	5.8322 ***	0.0752 (0.0115)	5.136 ***	0.0777 (0.0115)	5.305 ***	0.0779 (0.0114)	5.322 ***
Calibration Accuracy			4.2039 (1.9852)	1.658 *	4.4081 (1.9780)	1.738 *	2.9024 (2.1311)	1.145
Calibration Bias			-3.707 (1.3065)	-2.142 **	-3.3568 (1.3147)	-1.940 *	-2.7776 (1.4570)	-1.605 .
# of CHC Visits					1.5333 (0.8753)	1.533 .	1.4317 (0.9343)	1.432
(CHC Visits) × (Calibration Accuracy)							6.0728 (2.6426)	2.3946 *
(CHC Visits) × (Calibration Bias)							-2.5564 (1.6071)	-1.4772
r^2 (Δr^2)	0.2296		0.2795 (0.0499)		0.2910 (0.0115)		0.3123 (0.0328)	
F-test			$F(2) = 6.5806$ **		$F(1) = 3.0688$.		$F(3) = 2.9759$ *[(a)]	

Note: . $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$; [a] This *F*-test compares Models 4 and Model 2; SAT/ACT Math score, Calibration Bias, and Calibration Accuracy were centered at the mean.

The first step of the hierarchical regression analysis highlighted that students' incoming mathematical mastery (as measured by SAT/ACT Math scores) significantly contributed to the model, accounting for almost 23% of the variance in Exam 1 performance. Introducing measures of calibration into the model (Model 2) accounted for an additional 5% of variance in Exam 1 performance compared to the model only using incoming mathematical mastery (Model 1). This change in variance was further found to be statistically significant, $F(2) = 6.5806$, $p < 0.05$.

When including number of visits to the CHC before Exam 1 to the hierarchical model (Model 3), the additional variance that CHC visits contributed was not statistically significant. Hence, the number of visits students make to the CHC before the first exam is not associated with performance in on the exam after controlling for Calibration Accuracy, Calibration Bias, and SAT/ACT Math scores. Given how close the *p*-value ($p = 0.08$) is to the significance threshold ($\alpha = 0.05$) however, there is potential for type II error. Further investigation into the relationship between CHC Visits and Exam 1 scores (after controlling for measures of calibration and incoming ability) may be warranted in future studies.

As statistical significance was not achieved with the inclusion of the CHC Visits variable to Model 2, the final step of the hierarchical regression (Model 4) was compared to Model 2, thereby examining the additional variance contributed in predicting Exam 1 scores by including three predictors into the model: (1) CHC Visits, (2) CHC Visits × Calibration Accuracy, and (3) CHC Visits × Calibration Bias, in effort to examine potential interaction effects between CHC Visits and measures of calibration on Exam 1 scores. Comparison of these models revealed that these three variables (CHC visits, and interaction effects between CHC visits and calibration measures) improved Model 2 by accounting for an additional 3.28% of the variance in Exam 1 performance, $F(3) = 2.9759$, p=0.03289. Of the three included variables, only the interaction term between CHC visits and calibration accuracy was statistically significant, $\beta = 2.395$, $t(187) = 2.6426$, $p < 0.05$, suggesting that the relationship between the number of CHC visits and predicted exam one performance is mediated by students' measures of calibration accuracy.

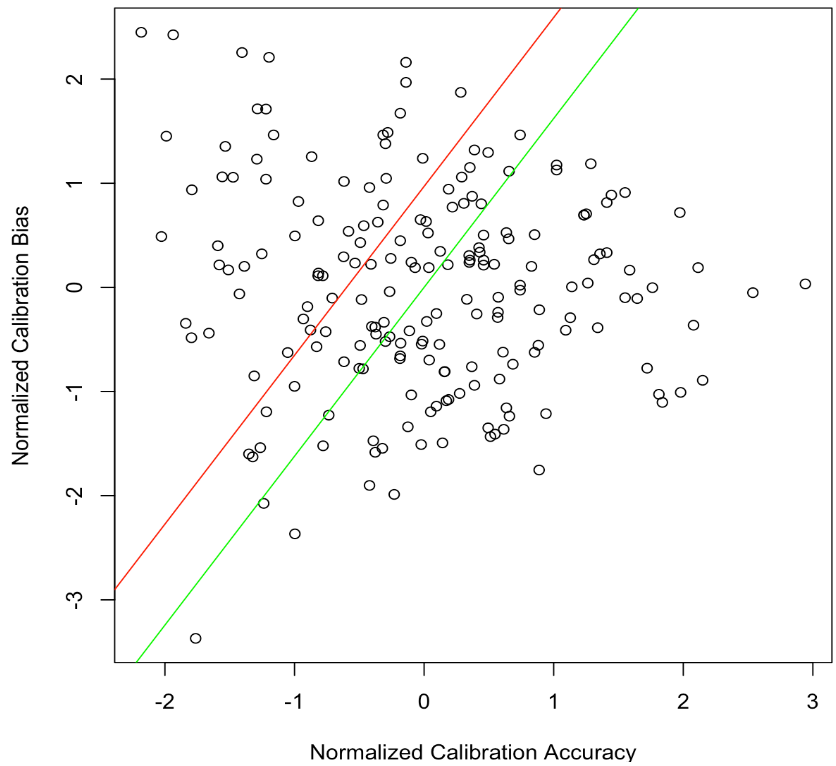
Model 4 suggests that the rate of change in Exam 1 scores per visit to the CHC differs between students with differing level of calibration accuracy, where the expected gain in predicted Exam 1 score per visit to the CHC is higher for students with higher calibration accuracy than students with lesser scores. Considering students with equal SAT/ACT Math scores and equal (and average) average calibration bias, students with the mean calibration accuracy are predicted, on average, to gain 1.43 points on their Exam 1 score per CHC visit, while students who are one standard deviation above the mean in calibration accuracy are predicted, on average, to gain 3.83 points per CHC visit, which is a rate 2.39 larger than that of students with average calibration accuracy.

Further, Model 4 predicts those who are one standard deviation below the mean in calibration accuracy performed 0.96 lesser on Exam 1 per CHC visit compared to those whose average calibration accuracy is average within the sample. Model 4 is not suggesting that each visit to the CHC is negatively impacted the students' grade. Rather, it suggests that students whose calibration accuracy is 1 standard deviation below the mean and who tend to go to the CHC tend to perform, on average, lesser than those students that do not visit the CHC prior to the exam. For instance, this trend may suggest that CHC visits are instead an indicator of a student struggling in the course. This will be further discussed in the Discussion section.

In this model, the impact of number of CHC visits for predicting Exam 1 scores depends on both calibration accuracy and bias. The full model for predicting Exam 1 score (using normalized variables) is: Exam 1 Score = 76.61 + 5.32 (SAT/ACT Math Scores) + 1.43 (CHC

Visits) + 1.14(Calibration Accuracy) - 1.61(Calibration Bias) + 2.39 (Calibration Accuracy) × (CHC Visits) - 1.48*(Calibration Bias) × (CHC Visits).

According to the model, the expected gains in Exam 1 scores per visit to the CHC is dependent on students' measures of calibration accuracy and bias. To help interpret the interaction terms between measures of calibration and CHC visits, consider the scatterplot of students' normalized calibration bias and calibration accuracy in Figure 3.5. The green line represents all students in the sample that are expected to gain 1.43 points on Exam 1 for each visit to the CHC before Exam 1, including the 'average student' who has average levels of calibration bias and calibration accuracy (i.e. normalized calibration accuracy and bias of zero). According to the model, students below and to the right of the red line (i.e., have greater normalized calibration accuracy and lesser normalized calibration bias) are predicted to have higher Exam 1 gain for every visit to the CHC than students along the green line, while students who are above and to the left of the green line (i.e., have lesser normalized calibration accuracy and greater normalized calibration bias) are predicted to have lesser Exam 1 gains for every visit to the CHC.



Note. The green line represents all students who are expected to gain 1.43 points on Exam 1 per visit to the CHC, and the red line represents all students who are expected to gain no points on Exam 1 per visit to CHC.

Figure 3.5: Scatterplot of Normalized Calibration Bias with respect to Normalized Calibration Accuracy

The red line in Figure 3.5 represents all students whose predicted Exam 1 score is neither greater nor lesser regardless of the number of CHC visits before Exam 1. According to the model, students below and to the right of the red line (i.e., have greater normalized calibration accuracy and lesser normalized calibration bias, such as students on the green line) are predicted to have higher Exam 1 scores for every visit to the CHC, while students who are above and to the left of the red line (i.e., have lesser normalized calibration accuracy and greater normalized calibration bias) are predicted to have lesser Exam 1 scores for every visit to the CHC.

Discussion

In addition to being correlated with student success, the PCQ accounts for additional variance in students' exam scores beyond what was expected from SAT/ACT math scores. This suggests that students' past performance in mathematics alone (as measured by the SAT/ACT math scores) does not tell the complete story of how precalculus and prior mathematics knowledge impacts success in Calculus. As previously mentioned, a potential explanation for this is proximity of mastery of content relative to taking the calculus course. While many students in this study were freshman and sophomores, students may not have completed a mathematics course in the last year, so their SAT/ACT math scores may not be representative of their mathematics ability when enrolled in Calculus I.

This work highlights the affordances of giving students an assessment (graded or ungraded) at the beginning of a course. This information communicates to the instructor material that may be difficult for students. While students have met the appropriate requirements to enroll in a course, students may not have mastered all of the prior content. An entry assessment can communicate to the instructor what content students are fluent with, which may or may not align with the instructor's expectations. For example, an assessment may indicate the importance for an instructor to use precise language when discussing concepts such as graphs and function notation, where students may have non-normative meanings for the content (David, Roh, & Sellers, 2018, 2019; Moore & Thompson, 2015; Sencindiver, 2020; Thompson & Milner, 2018; Van Vliet & Mirin, 2020). Students enrolled in a calculus course are drawing on their precalculus and prior mathematics content knowledge frequently when engaging in calculus content, such as reasoning about functions graphically, reasoning and solving algebraic equations and expressions, and making sense of new mathematical ideas using mathematical symbolizations such as function notation.

While this study shows correlations of student performance with in-house assessments (e.g., the PCQ), success in accounting for students incoming ability could also be measured using validated instruments, such as the Precalculus Concept Assessment or Calculus Concept Readiness assessment (Carlson, Madison, et al., 2010; Carlson, Oehrtman, et al., 2010). These assessments utilize research on student learning and extensive testing and refinements to produce psychometrically desirable assessments for diagnosing student difficulties as well as determining one’s potential readiness for a course. One would expect similar trends in the data to be found if the same study were conducted using such assessments.

Impacts of Calibration on Students’ Motivation and Success

This study expands on the calibration research by examining the effects of measures of calibration after accounting for variance from students incoming mathematics ability. Additionally, this work draws on a new way of measuring of calibration accuracy and bias through an average of item-wise measures of calibration accuracy and bias. I argue that this method better accounts for the subtitles of calibration accuracy and bias, as past methods for calibration accuracy may code students who are overconfident on some content and underconfident on other content as accurate across the entire content.

The findings of this study align with other research to suggest that measures of calibration are correlated with student’s success on exams (Bol & Hacker, 2012; Desoete & De Craene, 2019; Kline & Dibbs, 2018; Lingel et al., 2019), particularly that calibration bias is negatively correlated and calibration accuracy is positively correlated. While students who are underconfident may tend to have higher exam scores than students who are overconfident, students' accuracy of their perceptions of mastery account for more variance, as evidenced by the larger standardized coefficient of calibration accuracy than calibration bias in the model. Together these results highlight the importance of students' metacognitive monitoring in the learning process. In addition to a direct correlation on exam scores, measures of calibration and metacognitive monitoring likely have indirect effects on students learning as well. Ferla and colleagues (2010) concluded that while high self-efficacy is correlated with high academic performance, students with high self-perceived competence can lead to overconfidence, thus impacting lower persistence and poor study habits. Similarly, calibration inaccuracy may impact students' motivation to study and seek help. In order for help seeking to be deemed as necessary, a student must recognize a lack of mastery or a need to practice in a content area. Inaccuracies in self-perceptions of mastery, particularly overestimates, lie contrary to recognizing this need, thereby constraining this call to action. The current data further seem to provide evidence to this claim, particularly via the significant differences in the rates by which students are expected to benefit from help-seeking per visit.

Impacts of Calibration on Help-Seeking

Tutoring centers provide an avenue for students to get help and additional practice with difficult content and find a source of community with others seeking to improve their content knowledge (Bressoud & Rasmussen, 2015). For students struggling with course content, help seeking provides a mechanism for a student to potentially increase their course success through improving content knowledge. When considering Model 3, the analysis of these data suggests that the number of visits a student takes to help-center prior to an exam is not necessarily correlated with student performance on an exam, after controlling for measures of calibration and incoming mathematics mastery. From one perspective, this may seem as if the tutoring center is not improving student success, however this does not seem to be the case. The analysis regarding Model 4 suggests that the rate at which students benefit on exam scores per visit to the help center is dependent on students' calibration accuracy. Particularly students with more accurate perceptions of their mastery of precalculus (i.e. have higher measures of average calibration accuracy) are expected to score more points on Exam 1 per visit than students' whose perceptions are less accurate. This work provides further evidence that visiting tutoring centers can improve students' success and that these benefits may be stronger for some students than others (Rickard & Mills, 2018).

One potential explanation for this relationship could be of the type of help that students ask for (and receive) when at a help center. Students who are more accurate in their perceptions of their mastery are likely more attuned to what content areas they need support in. A student who is more accurately calibrated in their perceptions of their mastery would therefore be better positioned to identify content areas that are particularly difficult for them compared to a student who are less accurate in their perceptions of their own content mastery. This accuracy in perceptions of mastery would enable such a student to ask a tutor for help with those particular content areas to optimize the impact of the help received and the student's time in the center. From this perspective, the measure of calibration accuracy regarding precalculus content may be related to a students' metacognitive awareness of calculus content mastery or general accuracy in perceptions of mathematical mastery. Future studies would need to investigate in more detail (a) how measures of one's accuracy about precalculus content correlate with measures of one's accuracy about calculus content (or more general content that students have learned with content that students are currently learning), and (b) how measures of students' accuracy about mathematical content relates to students' planned and in the moment actions when seeking help in a help center.

Students who tend to be inaccurate in their perceived mastery of content may not have as accurate a sense of what content they are weakest in and should be the focus of their studies and help seeking. In this regard, students who tend to be inaccurate in their mastery perceptions may be experiencing the Dunning-Kruger effect, that is, when students are generally unaware of what they know and not know (Dunning, 2011; Kruger & Dunning, 1999). While the Dunning-Kruger effect at times used to describe the phenomena of people who have little experience or skill in some task to describe themselves as highly competent in the skill, that is not how this term is being used here; it is intended to refer to those whose mastery is low (according to some standard) yet who believe that there are more knowledgeable or exhibit more mastery than they do. In this study, students whom are both inaccurate in their perceived mastery and exhibit minimal mastery of precalculus content would be experiencing the Dunning-Kruger effect. Calibration inaccuracy (or high calibration bias) coupled with low mastery would suggest perceptions of mastery of content which are not evidenced in data. For such students, help-seeking would likely be less helpful because the content being reviewed with the tutor could potentially be determined by the student (who is not aware of what content they have not yet mastered). From this perspective, students who are experiencing the Dunning-Kruger effect may not be in the best position to determine what their strengths and weaknesses are, and hence may not ask for help with the content that would optimize their learning. When speaking with tutors in a tutoring center, students who are inaccurate in their own content mastery may tell a tutor that (a) they have mastered content that they need more practice with, (b) they have difficulty with content that they may be proficient with, or (c) they need general help with everything. These sorts of requests may result in generally inefficient or unproductive discussions during a help session, thereby reducing the impact of visits to the help center, causing lesser benefits for students whose perceptions of their mastery are inaccurate than students whose perceptions of mastery are accurate.

It is important to point out that this model may suggest that students who are inaccurate in their perceptions of mastery of precalculus content may be expected to have lesser scores on exam scores per visit to a help center. While this may be what Model 4 suggests, the analysis does not suggest that visiting to the help center is negatively impacting a student's mastery of calculus content (thereby decreasing their Exam 1 score). While the relationship between tutors serving in help centers and the courses they serve can be quite varied generally (Byerley et al., 2019), all tutors at the CHC were either instructors or learning assistance for the course. Given the expertise of those offering assistance during this study, it is unlikely that student interactions with a tutor would hinder one's mastery of the content area. The potential negative correlation with visits and Exam 1 score instead could be indicative of a student struggling with the calculus content. In this sense, the negative correlation of exam performance with visiting the CHC is not a product of the interaction at the CHC; rather, it would be the fact the student needed to seek help.

To illustrate how seeking help may be an indication of a student struggling with a course, consider a Calculus I student, Jerry. On the PSA for Calculus I, Jerry rates his confidence in precalculus as high, and on the PCQ, he has difficulty answering the questions, ultimately performing poorly. Jerry is not cued to visit the help center, however, given his confidence in his abilities, his past success in a proceduralized precalculus course, and his recognition of the symbols being used in class. After multiple weeks of difficulty and poor grades on homework and quizzes, Jerry visits the CHC the week of the Exam 1. Jerry has difficulty identifying what he is having difficulty with in this Calculus I course and given the number of students in the CHC the week before the exam, the tutor does not have sufficient time to identify and address Jerry's specific needs.

Notice that Jerry prolongs his help-seeking based on his assumption that his precalculus skills are strong (based on his past mathematical experience) but does not realize that his proceduralized notions of precalculus are constraining him from making sense of what is happening during the class period or what is being asked in the homework. Jerry is unable to point to the source of his difficulties given his belief in the strength of his precalculus knowledge, but his overconfidence delays his recognition of needing to seek help. While his help-seeking is a step in the right direction, the time needed to identify his sources of difficulties is not enough, especially given the number of students all seeking help prior to the exam. In this sense Jerry visiting the CHC is an indication of struggle, but the time allotted to help the student and the frequency of visits are insufficient to help him in time.

Suggestions for Practitioners: Supporting Students in Classrooms and Tutoring Centers

This work highlights students' thinking, perceptions, and actions that can impact their academic success. Given the importance of measures of calibration of perceptions of content knowledge both directly on exam performance and through improving the benefits of help-seeking visits, it is important to support students enrolled in calculus (and likely other undergraduate courses) align their perceptions of their content knowledge with what is generally accepted in the mathematics community. Based in this analysis and literature, we provide suggestions that practitioners can implement in their classrooms and in tutoring centers to help support students enrolled in Calculus I.

In the classroom

Instructors often have the most interactions with students, and directly influence what mathematical practices and activities students engage in, whether in the context of large lecture or small classrooms. This role affords instructors the opportunity to draw on literature and research findings to create a classroom environment that (a) supports students in learning mathematical content and (b) fosters the development of metacognitive skills related to student success (Labuhn et al., 2010; Zimmerman et al., 2011). To help students become more accurately calibrated, instructors may consider including more opportunities for students to assess their mastery of content knowledge, such as by providing students ungraded assessments where students can rate their confidence on content matter and test their hypotheses. By providing questions after the confidence assessment, students can be confronted with content that had not yet mastered, and hence adjust their mastery perceptions. Additionally, providing prompt (if not instantaneous) feedback about whether or not content items are correct would support students in recognizing that their solutions for mathematical questions may be non-normative. These practices would likely cause perturbations in a student's self-perceptions of their content mastery and whether their methods of solutions align with the mathematical community. These perturbations can cue students to further practice and study mathematical content independently, with peers, or via help-seeking with an instructor or in a tutoring center.

As with the other mentioned practices, instructors can further support students in aligning their perceptions of content mastery with normative views mathematical correctness through reflective practices. Instructors could provide students time in class to reflect on what content areas they can mathematically grow and a list of resources and ways for students to further practice, study, and get help with mathematical content areas both for the current course as well as other content necessary for course. For example, an instructor may give students a list of mathematical topics covered each week, a place to rate their perceived mastery of each content area, as well as list of actionable ways that students can assess and learn particular mathematical content, such as using online resources for calculus and precalculus material, reviewing notes, and visiting a mathematics help-center and special content-focus workshops.

Lastly, the instructor has the opportunity to establish sociomathematical and social norms in the classroom to foster alignment of perceived mastery with a normative mathematical standard (Cobb & Yackel, 1996; Yackel et al., 2000; Yackel & Cobb, 1996). By regularly implementing these practices, instructors can begin negotiating a sociomathematical norm where students are evaluating their perceived content mastery to a normative mathematical standard (Labuhn et al., 2010). Through the prominence of resources, help seeking (such as at a tutoring center) can become a calculated and productive means of (a) learning mathematical content that has been identified as difficult or not-yet mastered, and (b) adjusting one's self-perceptions of mastery. By establishing help seeking as a social norm, students may be more likely to visit help centers earlier and more frequently, potentially decreasing cases where students seek help right before the exam (like Jerry) and increasing the effectiveness of help centers.

In Tutoring Centers

As measures of students' calibration accuracy were correlated with the added benefits from in-person help-seeking, there are several potential ways that people involved in tutoring centers can leverage this data to potentially improve the effectiveness of help offered at help centers and student success.

One potential reason for why students who are inaccurate in their perceptions of content mastery may not benefitting from help-seeking as much as students with accurate perceptions of mastery could be potential inefficiencies from students' selected content areas during help-seeking. From this perspective, students may not be the best judges of what content they should focus their studies when seeking help. An instructor, teaching assistant, or tutor in a help center may be better suited to determine where a students' may be having difficulties. Rather than simply discussing a topic that a student suggests, members of help centers may want to seek evidentiary means for students' mastery on suggested topics, such as through quizzes, exams, or even conversations about the mathematical topics or problems. Likewise tutor questioning would play a vital role in probing student thinking and determining strengths and weaknesses of students' content knowledge, assuming the tutor's notions of the content are normative. Tutors questioning and attention to students' cognitive biases would help avoid potential issues of students misrepresenting their cognitive ability. For this reason, it would be beneficial for tutors to become aware of students' potential inaccuracies in their content knowledge, as well as tutoring center directors to potentially train tutors in probing student thinking and seeking evidentiary accounts of students' claims of mastery. With information about students' thinking and mastery, tutors can suggest problems and content areas for students to practice both in and out of the help center.

Additionally, it would likely benefit tutors to have a set of materials, such as assessments and content questions, that they can easily draw upon when determining whether a student has mastered a content area or needs additional practice and support. These materials could be leveraged by both tutors and instructors to make student thinking accessible and to begin addressing content that is difficult for individual students. The use of such materials would afford tutors to organize students in tutoring centers with similar mathematical difficulties to the same area where (a) the tutor can be more efficient by helping multiple students at once (such during times of high volume, such as prior to an exam) and (b) allow students whom have now mastered the material to help other students whom are currently learning that material. This later practice would both provide quicker response time when students enter the tutoring center, as well as helping the students who now has mastered the material to better her understanding by explaining to and helping others. Leaders of tutoring center and mathematics departments may want to establish a relationship and collaborate to create or collect such materials. Together, they can draw on research literature as well as the experiences of course coordinators, instructors, and tutors to inform items based on difficult concepts and common un-productive ways of thinking for the student population.

To further support students, course coordinators could provide the tutoring center (or students) a list of topics for the course. One benefit would be to provide students with specific areas where they can focus their studies. By conferring with such a list of specific content areas, students would be less likely to overgeneralize their mastery of content areas, potentially improving their accuracy of self-knowledge.

Another benefit for a list of content would be for tutors to assess student's growing mastery of the content in alignment with the particular course. Given the varied levels of familiarity of tutor centers and the courses they serve (Byerley et al., 2019), tutors may or not know be familiar with what content and ways of reasoning are valued in a particular course. A list of content areas would help the tutors support students in their particular course by better assessing students' content knowledge for the course and identifying particular content areas that a student may need additional support. Students could keep a copy of such a list and continually track their progress on course content to further guide their studies at home and at a tutoring center. This document could further provide tutors more evidence of what content a student has and has yet to master.

In addition to having tutors better understand what content a student may want to focus their time on, prompt feedback on a student's understanding of content would likely further improve the student's metacognitive abilities, particular their calibration accuracy. By basing one's perceptions in evidence such as the mathematical correctness of responses to content questions, students would become less susceptible to cognitive biases, thereby improving their calibration accuracy (through aligning their perceptions of mastery with data). These practices would also provide students with a better sense of what content they should likely continue practicing during independent study as well (Labuhn et al., 2010; Sheldrake et al., 2014; Zimmerman et al., 2011).

Limitations

As optional online tools were used to measure calibration, the sample for this study may have been influence by selection bias. While this selection bias is a possibility, it is important to contextualize the student populations use of these tools. Of the 356 consenting students (which includes over 85% of the student population), 77.8% and 68% of the population use the PSA and PCQ respectively. The resulting 61.2% of the consenting students is an intersection of these two groups of students.

Further choosing only students who had choose to use these tools is a control for student motivation. Self-efficacy and motivation have been said to support and constrain students' use of metacognitive monitoring and calibration, as a students' enactment of strategies are related to whether the student is motivated to succeed in the given task (Flavell, 1979; Pintrich, 2002). Tian and colleagues (2018) found that student self-efficacy mediates the relationship between metacognitive monitoring and academic achievement, so it important to control for students' self-efficacy or motivation. By choosing students who chose to use two optional online tools, the sample is in part limited to students who are motivated to complete optional assignments. Likewise, this study only included students who completed all exam in the course, further controlling for additional aspects of motivation through the duration of the entire course. Rather than a limitation, the choice of sample involved in this study is strength by including only students that have, to some extent, similar motivation in the course.

Chapter 4

Paper 3: You Only Get Out What You Put In: Calculus Students' Graphical Understandings of Outputs and Differences of Outputs of Functions

Understanding the derivative is non-trivial, as students' difficulties with the concept have been documented for decades (Lauten et al., 1994; Orton, 1983; Thompson, 1994; White & Mitchelmore, 1996). Some research has sought to categorize student errors with the derivative (e.g., Orton, 1983) while other research aims to identify the foundations or "cognitive roots" (Larsen et al., 2017) of the derivative that are hindering students from developing rich and productive meanings for the concept (Asiala et al., 1997; Byerley, 2019; Monk, 1994; Nemirovsky & Rubin, 1992; Thompson, 1994). Students' difficulties with the derivative can be traced back to impoverished understandings of various prior mathematical concepts, such as function (Carlson et al., 2002; Oehrtman et al., 2008), ratio and slope (Byerley, 2019; Byerley & Thompson, 2017; Nemirovsky & Rubin, 1992; Orton, 1983), rate of change (Thompson, 1994; Thompson & Carlson, 2017; Zandieh, 2000), covariation (Thompson & Carlson, 2017), and variable (White & Mitchelmore, 1996).

While students have proven proficient in computing derivatives of functions , students have difficulty understanding the rate of change of a function when given a graphical representation of a function (Larsen et al., 2017; Orton, 1983). When investigating calculus students' reasoning about derivatives, Nemirovsky and Rubin (1992) found that students tended to use either resemblance-based or variation-based approaches to matching graphs of functions with the corresponding derivative. Resemblance-based approaches aimed to match perceptual features of the two graphs (e.g., both graphs increasing over an interval), while variation-based approaches focused on local variation and the relationship between graphed quantities of a function and its derivative. Students' meanings for graph are likely related to students' use of these approaches, as students who understand a graph as a geometric object (e.g., a line) when matching graphs and derivatives may more readily attend to transformations of the shape of the graph (a resemblance-based approach) rather than coordinating changes in the quantities represented by the coordinates of points (a variation-based approach) (Moore & Thompson, 2015).

Carlson (1998) and Monk (1994) found that many students' meanings for the graph of a function were derived from visual attributes of the graph rather than measured values. Monk (1994) found that students confused a graph of velocity with a graph of position, and that students often interpreted two cars to collide when the graphs of the cars' velocities with respect to time intersect (Monk, 1992). These students' reasoning (i.e. iconic translation (Monk, 1992)) is based on thinking about the graph of a function like a picture of the scenario, where points moving along the graph represent the cars in the scenario rather than representing the states of two covarying quantities.

Based on prior research about students' graphical understanding of the derivative, students' understandings of a graphical representation of a function will impact the students' development of their understanding of derivative in a graphical context. As a first step towards understanding how these ideas are connected, this work examines how students understand the output of a function and the differences of outputs to be represented in a graph, as these two concepts (output and difference of outputs) necessary for understanding the average rate of change of a function.

I draw on David, Roh, and Sellers' (2019) location-thinking and value-thinking constructs (defined shortly) to characterize calculus students' understandings of function outputs in graphical contexts. David, Roh, and Sellers (2019) found that students' meanings for output, graph, and points on a graph can largely impact students' reasoning about mathematical statements. I hypothesize that students' meanings for output would impact their understanding of the differences of outputs, and their graphical understanding of the difference quotient. I imagine that students meaning for outputs would impact students' understanding of the relationship between a function and its corresponding derivative. My work uses the location-thinking and value-thinking constructs in a new context than David and colleagues' (2019) original study. This new context is influenced by Thompson's work on quantitative reasoning (Thompson, 2011) and his tasks on students' understanding of magnitude and measurement (Thompson et al., 2014). This study identifies how students with various meanings for output of a function may identify the difference of outputs on a graph.

In this report, I use the constructs location-thinking, value-thinking, and conceptions of magnitude to investigate calculus students' understandings of output of a function and differences of outputs of functions. Specifically, I address the following research questions:

1. What conceptions for output of a function do students use when engaging in a graphical task?
2. What conceptions for difference of the outputs of a function do students use when engaging in a graphical task?
3. How do students' conceptions for output relate to those for differences of output in a graphical task?

Theoretical Perspective

In this study, I adopt the radical constructivist perspective (von Glasersfeld, 1995). From this perspective, I assume that students' have individual meanings for mathematical content which are inherently inaccessible to others. Hence, the models of student's thinking I make are what I believe the student understands. I draw on my own interpretations of their written work, speech, and gestures to develop these models in effort to explain what the students are doing. As I am basing my models on what I am observing through their speech, gestures, and drawings, I am particularly modeling student's conveyed conceptions. Researchers aligned with this perspective do not claim that their interpretations of the student's thinking are correct; rather, we argue that the resulting model for the students' thinking is useful in explaining the student's actions and utterances.

Drawing on radical constructivism, I understand learning as a process of accommodating and assimilating to one's schemes (Steffe & Thompson, 2000; von Glasersfeld, 1995). For me, a scheme is "an organization of actions, operations, images, or schemes—which can have many entry points that trigger action—and anticipations of outcomes of the organization's activity" (Thompson et al., 2014, p.11). Assimilation involves adapting new information to fit into our existing schemas (Steffe & Thompson, 2000). Accommodation of a scheme involves modifications to the scheme that are permanent, in so far as the modification reemerges when the scheme is used (Steffe & Thompson, 2000). I focus on functional accommodations, where "an accommodation is functional if it occurs in the context of the scheme being used" (Steffe & Thompson, 2000, p. 290). When a student experiences a perturbation, the student is unable to assimilate the results of their activity to their schemes. If the perturbation is serious enough, the student may accommodate their schemes, thereby engaging in the activity differently (von Glasersfeld, 1995). Hence when (1) a student conveys a meaning for a concept in one way, (2) becomes confused or perturbed, and subsequently (3) conveys a different meaning, I use this data as evidence of the student assimilated to a scheme, where the student had (1) operated using their scheme, (2) experienced a perturbation, and (3) accommodated their scheme, thus having operated according to their modified scheme in an observably new way.

When describing student's conceptions, I draw on Thompson, Carlson, Byerley, and Hatfield's (2014) description of understandings and meanings. From their perspective, a student is said to have an *understanding* when they are in a cognitive state resulting from assimilation, while a *meaning* is the "space of implications that the current understanding mobilize" (p.13). Students' understandings and meanings can be considered either stable or in the moment. This distinction is determined by whether or not a student is assimilating to scheme. If a student's meaning is stable, then the understanding resulted from assimilation *to a scheme*, and 'the space of implications' is the actions, operations, images, schemes, and anticipations that arise from having assimilated to the scheme. When Thompson et al. (2014) describe the scheme in the case of a stable meaning, the meaning is the scheme. This distinction between stable and in the moment (or unstable) understandings/meanings will helpful when modeling students' thinking. In response to students being asked to represent the difference of outputs of a function, some students assimilated the question to their current meaning for output. Other students, however, accommodated their meanings for output to resolve the perturbations they were experienced when representing the difference of outputs.

Literature Review

Quantitative Reasoning

Quantitative reasoning in mathematics education is a form of reasoning about situations where one conceptualizes the quantities involved and relationships between them (Thompson, 2011; Thompson & Carlson, 2017). Rather than focusing on promoting symbolic manipulation and blind computation, quantitative reasoning focuses on how mathematical operations connect to realizable situations and what relationships exists between different measurable aspects of the situation. For example, when imagining a car in motion and some means to keep track of or measure time, one can think about how the car's position varies as well as how the two quantities (time and position or distance from a reference point) vary simultaneously. With this framing, one can ask questions about how the car's distance changes with time, how the distance traveled on average changes with respect to changes in time over an interval of time, and how the distance traveled instantaneous changes with respect to changes in time over an interval of time.

Thompson (2011) describes a quantity as one's conception of an aspect or attribute of a situation as being measurable. For example, one can conceive of a line segment or a curvilinear arc as having a length, which is some of amount or size. This line segment's length is measurable, and one may assign a numerical value to the length based on a unit of measure, such as inches, feet, or meters. While this numerical value changes depending on the unit of measure, the magnitude of quantity is invariant regardless of the measurement (Moore, Stevens, et al., 2019; Thompson et al., 2014). One does not need to reason with numerical values to engage in quantitative reasoning; one only needs to anticipate that the quantity has a measure at any given instance and associate some amount or size to the quantity at those instances.

Quantitative reasoning is closely related to covariational reasoning, which involves coordinating how two quantities simultaneously vary together (Thompson & Carlson, 2017). Covariational reasoning has been a focus for calculus students in particular because reasoning quantitatively about rates of change involves such coordination of quantities (Thompson, 1994). Research suggests that covariational reasoning is developmental and researchers often model students' covariational reasoning with multiple levels (Carlson et al., 2002; Thompson & Carlson, 2017). Students' in the moment level of covariation is commonly determined by students' covariation in graphical contexts, where students either represent the covariation of two quantities by creating a graph of the relationship, or by discussing the covariation of quantities represented in graphs (see Thompson & Carlson, 2017).

Graph

Graphs can represent the covariational relationship between two quantities (Carlson et al., 2002; Moore et al., 2013). In the calculus classroom, graphs of functions provide opportunity to explore concepts such as average and instantaneous rate of change between two quantities, however, students do not always reason about graphs covariationally (Byerley & Thompson, 2017; Nemirovsky & Rubin, 1992). Without instruction, graphs may be understood as a picture of a scenario (Carlson, 1998; Monk, 1994) or as an object on the page (Moore & Thompson, 2015). Hence, students' understanding of what a graph is and how it represents quantitative information is likely related to their success when reasoning about such concepts.

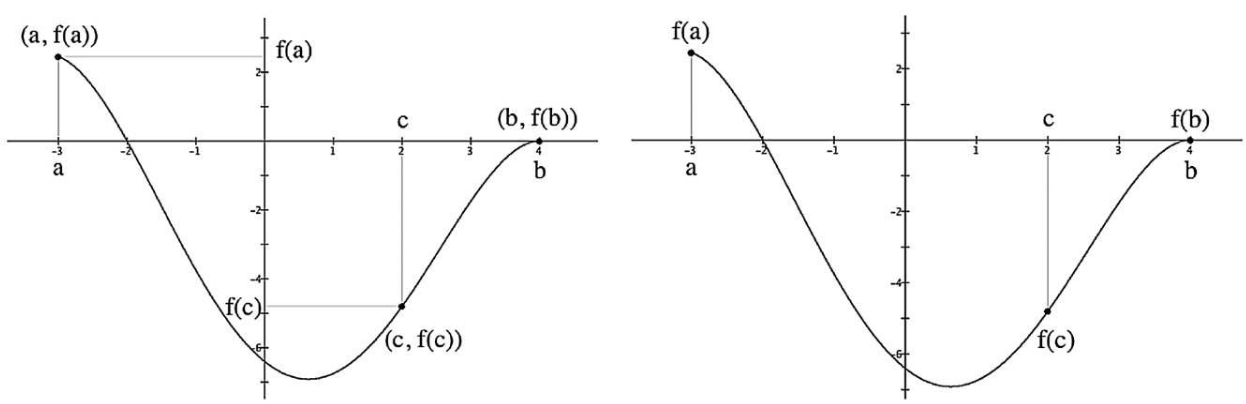
Moore and Thompson (2015) discussed students' understanding of graphs in terms of shape-thinking. They contrast students who view a graph as a wire on a coordinate system (static shape-thinking) with students who view a graph as a trace of two covarying quantities (emergent shape-thinking). Students using static shape-thinking tend to draw on perceptual features of the graph, while students using emergent shape-thinking recognize both the covariation of the two quantities as well as the shape of the graph (as a trace). Moore (2016) describes the mental activity that one engages in when using static or emergent shape-thinking, as either figurative or operative thought. Students' thoughts about graph that, in the moment, are "based in and constrained by sensorimotor experience (including perception)" (Moore, 2016, p. 324) are said to be *figurative*. Students thoughts where "figurative material was subordinate to mental operations" (p. 324) is said to be *operative*. These constructs are a productive way for distinguishing students who are viewing a graph as a source for encoding covariation of two quantities from those that are drawing perceptual features of graphs. While these constructs offer one way to categorize student's activity and thinking around graphs, I primarily draw on the constructs of David, Roh, and Sellers (2019) in this work to discuss students' conceptions of graphs, specifically how the output of a function is represented in a graph.

Location-Thinking and Value-Thinking

The goal of my research is to document how students' conceptions of graphs impact their development of other mathematical ideas. David, Roh, and Sellers's (2019) constructs value-thinking and location-thinking are particularly helpful for categorizing students' graphical thinking, particularly about the relationship between points on a curve and the output of a function. When investigating how undergraduate students reason about four mathematical statements related to the intermediate value theorem, David, Roh, and Sellers (2019) gave students six different graphs of functions and asked the students to evaluate whether the statements were true. While the students reasoned about the statements, the authors categorized the ways in which students were reasoning about the graphs. The constructs *location-thinking* and *value-thinking* were used to describe students' understanding of outputs of functions, graph of a function, and points on a graph of a function.

Students engaging in value-thinking are characterized as thinking of the output of a function for a given input value as a value. These students further understood a point on a graph as a pair of values [e.g. (input, output)], and a graph of a function as a set of input/output pairs. Students engaging in location-thinking are characterized as thinking of the output of a function for a given input value as a point lying in the Cartesian plane along the graph of the function. Further a student engaging in location-thinking treats a point on a graph as indistinguishable from the resulting output for a given input value and a graph of a function as a collection of spatial locations in the Cartesian plane.

David and colleagues (2019) coded students' thinking based on how they labeled mathematical objects embedded in the graph. For example, students who created a label along the vertical axis and labeled the marking *f(a)* would be coded as using value-thinking, as would a student labeling a point on the graph as a coordinate-pair *(a, f(a))*. Students who created labels on a point along the curve and labeled it *f(a)* were instead coded as using location-thinking, as they are conveying that the point on the curve is the output. David et al. (2019) provide two graphs of a function with labels indicative of value-thinking (Figure 4.1a) and location-thinking (Figure 4.1b) to illustrate the differences in students' labeling activity in Figure 4.1 below.



(a) (b)

Figure 4.1: A visual model of student work indicative of value-thinking (a), and location-thinking (b). (David et al., 2019)

David, Roh, and Sellers (2019) saw that students engaging in value-thinking tended to evaluate the truth of these statements in the normatively correct way, while students that were engaging in location-thinking more frequently gave incorrect responses when evaluating the truth of the statements. In their sample, one student in Advanced Calculus/Real Analysis (the course with the highest level of mathematical preparation in the sample) used location-thinking to reason about graphs. This provides an existence proof that students who have been deemed successful in courses such as Calculus I and Introduction to Proof still find location-thinking a productive means for reasoning about graphs.

My work connects students' meanings for graph and output (in terms of location-thinking and value-thinking) to later conceptions for students have for other mathematical conceptions in calculus. While David, Roh, and Sellers (2019) used graphs with numerical axes, I consider what meanings for graph, and outputs of functions students draw on when representing the outputs and differences of outputs of a function in non-numerical graph, where quantities are represented using magnitudes of drawn line segments.

Frames of Reference and Graphs

When a student reasons with a graph, there is always a coordinate system in which the graph is in relation to. However, students may not necessarily attend to the coordinate system the curve is in relation to (Moore et al., 2013; Moore, Silverman, et al., 2019). Investigating a student's frames of reference (Joshua et al., 2015; H. Y. Lee et al., 2019) provides a tool to inspect how to a student makes sense of a representational system. Joshua and colleagues (2015) describe that a student who is reasoning within a frame of reference must commit to (1) a unit to measure quantities, (2) a reference point to measure quantities in reference to, and (3) a directionality defining what is positive and negative. In terms of graphs, points in Cartesian coordinate systems are intended to represent data in a particular frame of reference, specifically the one Descartes intended. From this perspective, a person representing an input value along the x-axis with a normative view of a Cartesian coordinate system may be committing to the origin as a reference point, committing to the rightward direction as the direction representing the 'positive direction' (using the conventional orientation of the Cartesian coordinate system), and the numerical value/length in the rightward direction provides a multiplicative comparison of the measure of the quantity in relation to the unit. While the unit is vital for reasoning about measures of quantities and numerically representing quantities, a unit of measure is not necessary for reasoning about magnitudes of quantities, that is one's conception of a quantity' size without the need to measure the quantity (Thompson et al., 2014)

Frank (2016) provides a model for understanding a point in the Cartesian plane as representing two quantities simultaneously by projecting orthogonally onto the axes, as shown in Figure 4.2. Using this model, the two quantities are represented as the directed distance from the origin to the end point of the segment along the respective axes and can be discussed in terms of frames of reference.

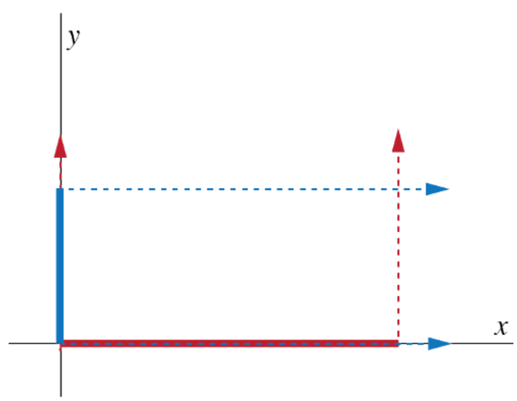


Figure 4.2: Frank's (2016) representation of a point as a projection of two quantities'

Consider the quantity represented by the y-coordinate of the point where the dotted lines intersect in Figure 4.2. In terms of frames of reference, a student conceiving of this magnitude (in blue) in the normative way would likely conceive of the directed distance from the reference point at the origin vertically (i.e., with the upward direction corresponding to the positive direction) to the labeled marking along the vertical axis. To associate a numerical value to this magnitude, one would need to commit to a unit to measure the magnitude and identify the measure of the magnitude in relation to the unit. In the context of graphs with unmarked axes however, students can represent the magnitude $f(a)$ with a line segment of length $f(a)$.

Value-Thinking, Frames of Reference, and Differences

In addition to marking along the vertical axis, a student may represent the magnitude of the output vertically from the horizontal axis upward to the point, in other words, from the marking a on the horizontal axis vertically to the point $(a,f(a))$, as shown in Figure 4.3. A student could further re-present the length of this line segment to other spatial locations on the graph, either by the request of a researcher or in efforts to compare the magnitude of two such quantities. Figure 4.3 below provides a variety of normative representations for the output of a when engaging in value-thinking.

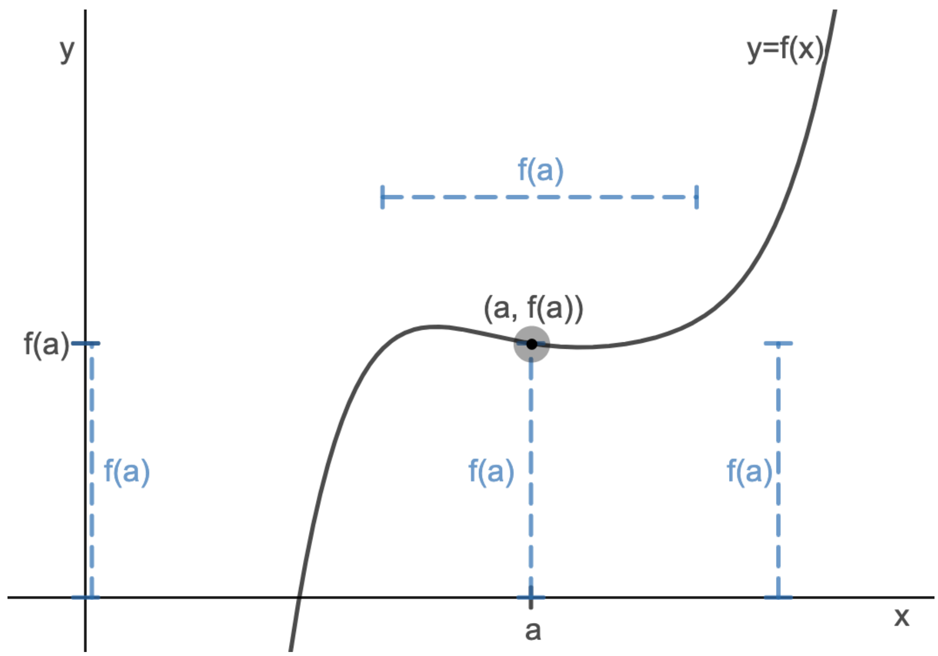


Figure 4.3: Different representations of the output quantity when using value-thinking.

When reasoning with non-numerical graphs, thinking of the output of a function in such as way may afford conceiving of the *difference* of two outputs of a function in a particularly productive way. Thompson (1993) describes two perspectives of a difference: the result of subtraction, and a quantitative difference. The first perspective involves the operation of subtraction between two numerical values, while the later involves a quantitative reasoning. Thompson (2011) elaborates on a quantitative difference as an additive comparison of two quantities which in turn is a quantity, which involves conceiving of "the amount by which one quantity exceeds another" (p. 203). One conceiving of a quantitative difference may position two magnitudes in such a way that they share a reference point and are oriented in the same direction. Then the magnitude of the difference can be represented as the length of the line segment between the ends of line segments representing the two magnitudes. Figure 4.4 highlights one way a student could potentially conceive of the difference of the two outputs *f(b)* and *f(a)* as a quantity. The magnitude of the line segments in blue represents the output *f(a)* in two positions, one vertically from the horizontal axis to the point *(a,f(a))* and one re-presented near a line segment representing the magnitude of *f(b)* in green, so that magnitudes *f(a)* and *f(b)* have spatially similar reference points. The magnitude of the red line segment represents *f(b)-f(a)*, the difference of the two outputs *f(b)* and *f(a)*. This second representation of *f(a)* is intended to exemplify how this representation can be used flexibly to perform both additive and multiplicative comparisons of two quantities.

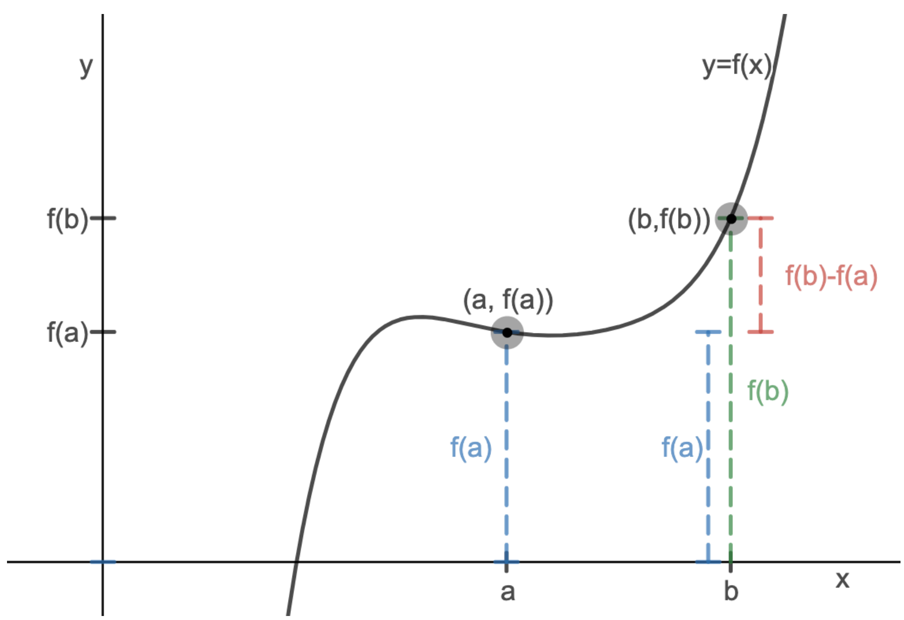


Figure 4.4: A graphical representation of two outputs of a function and the difference of the two outputs when using value-thinking in a quantitative context.

Value-Thinking and Location-Thinking in Non-Numerical Contexts

Value-thinking seems productive when representing outputs and differences of outputs in non-numerical graphing tasks. I hypothesize that a student using location-thinking in such a task would be able to represent two outputs of a function but would have difficulty representing the difference of the two outputs. Students using location-thinking consider the output of a function to be a point on the curve, but it is not clear to me how a student would make sense of difference of two geometric points. Additionally, I do not believe it is necessary for a student graphing or identifying a point on a graph to necessarily reflect on the measures of the coordinates used to graph the point. Frank (2016) describes how a student plotting a point on a graph in terms of the recipe ‘over *x* units and up *y* units’ may not conceive of a point as a representing two quantities. This is akin to a person following directions of how to get from one place to another, but not knowing where their current position is in relation to the starting position[1], which Skemp (1976) describes as an instrumental understanding. I expect students who do not conceive of a point as representing two quantities simultaneously to experience difficulty when representing the difference between two points and, hence, ultimately become perturbed.

Given my expectations for students using value-thinking and location-thinking in a non-numerical graphical context, I designed a task for students to represent outputs of functions and differences of outputs in a non-numerical graph. The graph includes axes with no a priori numerical values nor ‘tick marks’. I provide students lengths of line segments that represent the quantities as inputs for the function whose graph is given. This task is intended to problematize thinking rooted in location-thinking. While I expect students using value-thinking to progress productively through this task, I anticipate students engaging in location-thinking to have difficulty completing this task.

[1] I credit Dr. Cameron Byerley for this apt simile. The simile is very helpful for accounting the difference in performing an action and reflecting on the sensorimotor actions.

Methodology

Data Collection and Task Design

Data for this study was collected as part of a larger study aiming to understand how students' understanding of a graph of a function impacts the development of the derivative of a function for a fixed input. Data was collected from first-semester calculus students enrolled in a calculus course tailored to biological scientists in the summer of 2019 at a large western public university. Recruitment occurred at the beginning of the first two class periods of the course, where five students agreed to participate in at least one interview for the study. Students were asked to participate in two 90-minute semi-structured interviews during the first two weeks of the semester and were financially compensated for their time. All but one student (Lisa) participated in both interviews which were scheduled one week apart. Each interview occurred before the students' calculus course was scheduled to discuss rate of change in class. Three students were women, two students were men, and all five students were white. Each of the participants indicated having previously taken a first course in calculus within the past four years. No data was available regarding the student's mathematical performance in past coursework. Interviews with students were recorded and transcribed to capture student's utterances, drawings, and gestures.

One task was chosen to be the focus of this paper. The task (inspired by a task from Thompson and colleagues (2014)) used Cartesian axes oriented in the conventional manner and asked students to represent the outputs of two inputs and the difference between those outputs (see Figure 4.5).

Consider the graph of the function *f(x)* above. The lengths of colored line segments above represent the quantities a and h. Both the *x*- and *y*-axes have the same scale length.

Represent the following quantities to the right.

I. *f(a)*
II. *f(a + h)*
III. *f(a + h) - f(a)*

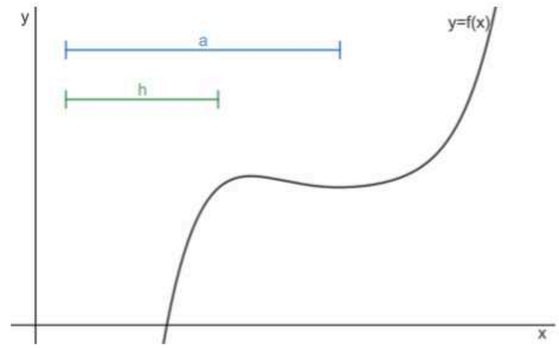


Figure 4.5: The interview task

The first two questions asking students to represent *f(a)* and *f(a+h)* were designed to characterize students' thinking of output as either value-thinking or location-thinking. The task was designed to necessitate a student to reason in a non-numerical, non-computational way, particularly with magnitudes, both as inputs and outputs of the function. Students need to reason with the magnitudes represented in the graph, thus allowing the interviewer/teacher/researcher to inquire what magnitudes on the paper students are attending to throughout the task. This choice was made so that students' understandings of output in the graphical context would be at the forefront of their mathematical reasoning, instead of computation. Analysis of students' responses to earlier versions of these items revealed that some students computed outputs and differences of outputs of functions numerically and only then matched the computed numerical value to a labeling represented on the graph. In this task, however, I anticipated that students using value-thinking would either need to draw on their measurement schemes to represent outputs or to create numerical values along the axes for some unit of measure. Students using location-thinking were expected to represent the output as a point along the curve.

To further investigate the relationship between students' understanding of output and differences of outputs, students were asked to represent *f(a+h)-f(a)*. As previously discussed, I anticipated that students using value-thinking would be able to productively represent the difference of the two outputs. By drawing on their measurement schemes, I expected students using value-thinking to understand or come to understand the ‘y-coordinate’ of a point as a vertical directed distance, either represented from the x-axis to the point on the curve or from the origin along the y-axis.

As discussed previously, I expected the representation of $f(a+h)-f(a)$ easier for students using value-thinking than for students using location-thinking. I expected students using location-thinking when representing $f(a+h)-f(a)$ to either experience difficulty conceiving of ‘the difference of two points’, or to potentially accommodate one’s meaning for difference to understand or come to understand the difference of two points as ‘the distance between two points’ as the length of the straight line segment between the two marked points.

Students were then asked to estimate the average rate of change of $f(x)$ from input a to $a + h$, though this item was not considered in this analysis except to highlight one student’s thinking about output and differences of outputs. Specifically, Alison’s conception of the difference of outputs became salient when discussing the numerator and denominator of the ratio of the average rate of change.

Data Analysis

Data was analyzed using theoretical thematic analysis (Braun & Clarke, 2006) where I coded instances of student’s reasoning about output as either location-thinking, or value-thinking. Instances that did not fit either of these constructs were coded as ‘other’ and I used inductive thematic analysis (Braun & Clarke, 2006) to code the remaining conceptions of output based on what mathematical object the student seemed to be referring to. Similarly I used theoretical thematic analysis to analyze student’s conceptions of the difference of output, $f(a+h)-f(a)$, using Thompson’s (1993) distinction of subtraction and qualitative difference. For those conveyed meanings that did not align with quantitative difference or subtraction, I used inductive thematic analysis to code what mathematical object the student seemed to be referring to.

During thematic analysis, I employed both ongoing and retrospective analysis techniques (Steffe & Thompson, 2000; Thompson, 2008) to generate and test hypotheses about students' conceptions of mathematical ideas. On-going analysis involved generating hypotheses about students' thinking during the interview based on my inferences of students' actions. At times, I also tested these hypotheses during the interview by asking the interviewee questions, such as asking why the student gestured in a specific way, what the student was referring to, or whether an idea was familiar to the student. Retrospective analysis involved analyzing the data after the interview to develop, test, and reject hypothesized models of students' thinking. Working models of students' thinking were iteratively refined until the models accounted for students' utterances, drawings, and gestures during the episodes. The resulting models are the result of the research team discussing and iteratively refining working models until the models accounted for students' utterances, drawings, and gestures during the episodes. The research team included the author and an expert in mathematics education research. As the data I drew on involved what was observable to me, these models are intended to describe the students' conveyed meanings.

Episodes

As I was expecting students to experience perturbations and accommodations to their schemes during the course of the task (e.g., the anticipated difficulty for student's using location-thinking to conceive of the difference of outputs), I coded students' activity with the task using multiple episodes. Different episodes were used to distinguish between different conceptions that student's conveyed during the task. For example, the interviewer at times attempted to shift the student's thinking during the task through a brief intervention. During such instances, students typically conveyed different conceptions than prior to the intervention, and their activity was considered a different episode. Without the support of an intervention, one student (Colin) further spontaneously conveyed a different conception during the task as the result of an accommodation to his scheme and, hence, was coded as a different episode.

Criteria for using Location-Thinking or Value-Thinking

Students in the sample seemed to represent input values in ways I was not expected. Drawing on the frames of reference construct, I considered David, Roh, and Sellers (2019) description of location-thinking and value-thinking using the frames of references (Joshua et al., 2015) construct. The students in David, Roh, and Sellers's (2019) paper seemed to recognize the input value in the normative way, regardless of whether using location-thinking or value-thinking. I found no evidence that those students were non-normative frames of reference for representing input values. Therefore, I used the frames of references construct (Joshua et al., 2015; H. Y. Lee et al., 2019) as a criterion for determining whether or not students were using location-thinking and value-thinking constructs, by identifying whether the student represented the inputs in a non-normative way. Students who used non-normative frames of reference for representing input were coded as 'other'. To this end, students that identified outputs as points on a curve would also have to use a normative frame of reference for a graph to be coded as using location-thinking. Likewise, students identifying a value or magnitude with a non-standard frame of reference would be coded as 'other'. The normative frame of reference representing an input of a function for a graph in Cartesian coordinate system involves using the origin as a reference point and representing the rightward direction along the horizontal axis as positive.

Results

In the theoretical thematic analysis (Braun & Clarke, 2006), three students (Abby, Colin, and Ethan) conveyed multiple conceptions for output, which were coded as different episodes. Abby and Colin's activity with the task used two distinct conceptions for output, and Ethan conveyed three distinct conceptions throughout the task. Alison and Lisa each conveyed one conception for output. Table 4.1 provides a summary of the codes for each episode.

Table 4.1: Codes for each episode

Episode	Frame of Reference for Input	Conceptions of Output	Conceptions for Difference of Outputs
Abby (First)	Non-Normative	Point	N/A
Abby (Second and Final)	Normative	Value-thinking	Quantitative Difference
Alison (First and Only)	Non-normative	Arc Length	Quantitative Difference
Colin (First)	Normative	Location-Thinking	Point
Colin (Second and Final)	Normative	Value-Thinking	Quantitative Difference
Ethan (First)	Non-normative	N/A	N/A
Ethan (Second)	Normative	Location-Thinking	Point
Ethan (Third and Final)	Normative	Value-Thinking	Quantitative Difference
Lisa (First and Only)	N/A	Arc Length	Quantitative Difference

During five of the nine episodes, students' conception of output aligned with either value-thinking (three episodes) or location-thinking (two) episodes, as seen in Table 4.1. For the remaining four episodes, students conveyed non-normative frames of reference when representing the input. During one of these episodes (Ethan's first episode), the student did not convey a meaning for output, as the student became perturbed quickly after describing the input value. During the remaining three episodes, students' conceptions were coded as 'other'. When reviewing the episodes coded as 'other,' I used inductive thematic analysis to create sub-codes to characterize what mathematical object the students identified as the output, which included 'arc length' and 'point'.

The sections that follow outline the different meanings for output that students conveyed during the task. I report five illustrative episodes to show the various conveyed conceptions that were observed during the task. The first episode comes from Colin's first episode with the task to illustrate location-thinking. This episode is representative of the episodes where students using location-thinking. The second episode comes from Abby's second episode to illustrate value-thinking. Abby's activity in this episode differs from the other students using value-thinking because Abby tended to measure the magnitudes of outputs in terms of the unit a. The other students that used value-thinking, Ethan and Colin, represented the outputs as magnitudes along the vertical axis and below a point on the curve, respectively. The following three examples illustrate the different conceptions students conveyed for output that were neither value-thinking nor location-thinking. Episodes three and four discuss the conceptions of output conveyed by Alison and Lisa, both of which were coded as 'arc length'. The fifth and final episode presented comes from Abby's first episode, where Abby's conception of output was coded as 'point'.

Episode I: Colin (Location-Thinking)

Output

When beginning his work on the task, Colin's conveyed meaning of output aligned with location-thinking. To represent *f(a)*, Colin first labeled *a* along the horizontal axis by using a piece of paper to measure the length *a,* measuring right along the horizontal axis from the origin (his reference point), and creating a dot at the end of the length *a* (see Figure 4.6). To represent *f(a)*, Colin then identified the point on the curve that is "straight above that" by using the piece of paper to form a straight edge perpendicular to the horizontal axis from the marking *a* to find the point where the perpendicular line intersects the curve.

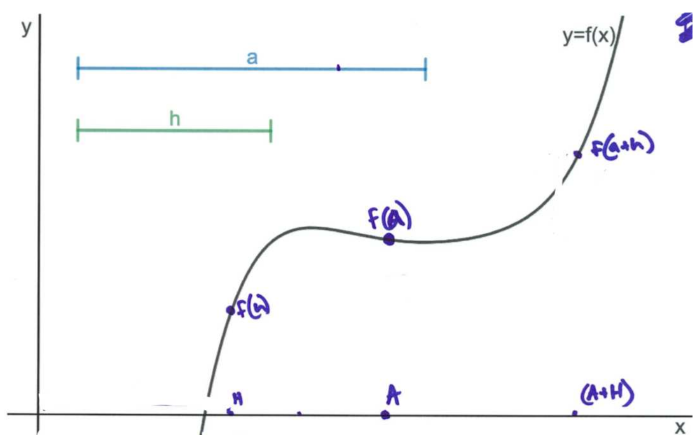


Figure 4.6: Colin's work after having plotted f(a), f(h), and f(a+h)

Colin represented *f(h)* and *f(a+h)* in a similar fashion, specifically measuring the input length along the horizontal axis using the origin as the reference point and identifying the point on the curve above the marking along the axis as the output for the given input. Like the other students, Colin concatenated the lengths *a* and *h* to form the length *a+h*. Note that Colin was not prompted to represent *f(h)*, a fact that he only noticed after he had already represented the output.

Given his representational activity for the outputs as points along the curve, Colin indicated a meaning for output consistent with location-thinking. Colin also conveyed this meaning when he described his work when representing *f(h)*, saying “so this down here is the value of *h* as an input, and the output using *h* is right there”, where then creates the dot on the graph corresponding to *f(h)*. Not only does this highlight that the point on the curve is the output of *h*’, but also highlights Colin’s attention to the spatial location of the point by saying ‘there’.

Colin’s Attention to the Coordinates of Points

Unlike Ethan, Colin additionally seemed to have a meaning for this point on the curve in reference to the axes. When describing the point on the curve, Colin wanted to represent the point in terms of the coordinates in Cartesian space. When representing an output, Colin commented on the lack of labeling along the axes. He indicated that the he could measure the y-coordinate of the point *f(a)* in terms of the input value, saying “this is one *a,* and the y-value is whatever fraction of *a* that is” gesturing first horizontally and then vertically from the label *a* on the horizontal axis to the point labeled *f(a)* (see Figure 4.7).

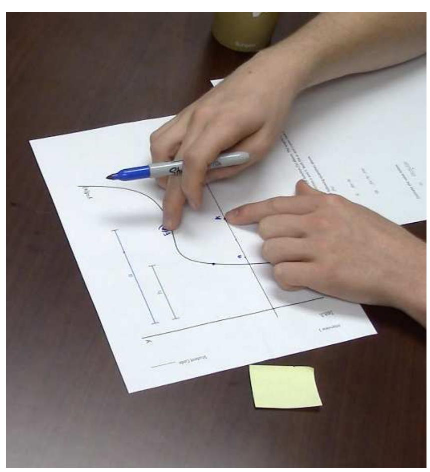


Figure 4.7: Colin’s gesturing while describing the ‘y-value’ of the point

Difference of Outputs

When describing the difference of the outputs *f(a+h)* and *f(a)*, Colin was explicit with his thinking. Colin understood the difference *f(a+h)-f(a)* as the point he previously identified as *f(h)* by drawing his attention to the markings for the input values of *a* and *a+h*. As Colin described his process for identifying *f(a+h)-f(a)*, he said "*f(a+h)* is right there [gestures to the point labeled *'f(a+h)'*], minus *f(a)* [gestures to the position on the horizontal axis labeled *'a'*] should just take us back to there [gestures to the point labeled *'f(h)'*]". Notice that when was Colin gesturing, he pointed to the position *a* along the horizontal axis, despite saying '*f(a)*'. In this moment, Colin seemed to be focused on variation in the *x*-coordinate of the points. When the interviewer asked Colin about his thinking, Colin described his thinking while motioning along the horizontal axis.

> *Colin: Okay for this one, I'm thinking 'f(a+h)' [gesturing to the point labeled 'f(a+h)'] should be the output of 'a+h' [gesturing to the marking 'a' on the horizontal axis] on the x-axis. But then for this one [the prompt 'f(a+h)-f(a)'], I'm thinking f(a+h) is there [gesturing to the point 'f(a+h)'], and then you're subtracting that [gesturing to the point 'f(a)']… So I'm trying to think, are you moving along the x-axis [does a sweeping gesture from the marking 'a+h' to the marking 'a' along the horizontal axis, see Figure 4.8]…So like if it was f(a-h), you'd go here [gesturing to the marking 'a' on the horizontal axis] and end up over here [gesturing on the horizontal axis between the 'a' and the origin]….Since it's f(a+h), we're over here [gesturing to the marking 'a+h' along the horizontal axis].*

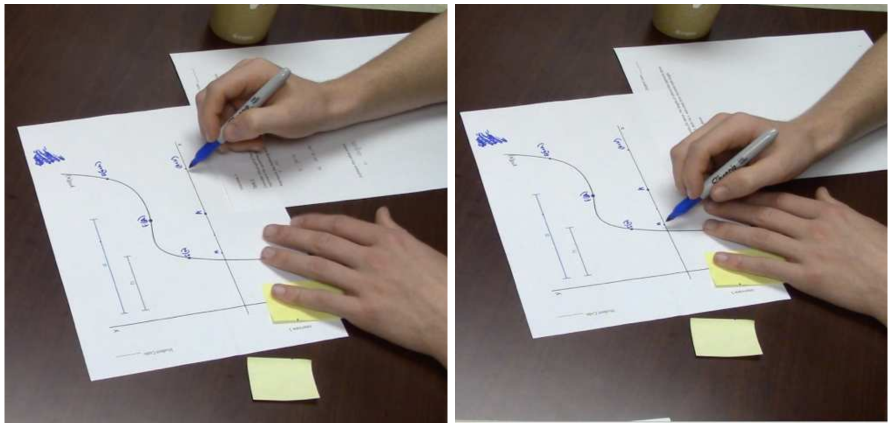


Figure 4.8: Colin's sweeping motion from *'a+h'* to *'h'*

As he gestures and described his thinking in the excerpt, Colin is focusing on the variation in the input values. Colin performs operations on the length *a* oriented on the horizontal axis, such as subtracting the length *h* (i.e., representing *a-h*) and adding the length *h* (i.e., representing *a+h*). When describing these operations, Colin is primarily focused on the horizontal axis, as evidenced by his gesturing only to the axis when referring to *f(a-h)*. Colin further does not mark the corresponding point vertically oriented on the curve, nor does he gesture or speak about this hypothetical point. Colin's identification of *f(h)* as the difference is consistent with what he would likely consider the output of the difference of the lengths *a+h* and *a*.

Summary of Colin's Thinking

Colin represented input values in the normative way, using the origin as reference point, measuring the length of his input rightward along the horizontal axis from the reference point, and creating a labeled dot at the end of the line segment. Colin then identified the output for the given inputs in a way consistent with location-thinking, by tracing upward from the marking along the horizontal axis to find the point on the curve that was vertically oriented from the marking. To Colin (and Ethan during his second episode), the output of the function was a point on the graph. Colin understood the difference of outputs in a way consistent with the output of the differences of inputs.

Colin's meaning for difference in this way was not stable. After reflecting on his work, revised his meaning of difference of outputs as a difference of coordinates pairs.

Colin: But if you're doing this [gestures to the point f(a+h) on the curve] minus this [gestures to the point f(a)], then you're subtracting two.. coordinates. Cause this [gestures to the point f(a+h) on the curve] is a coordinate. It's standing in for an 'x-comma-y-value'. So I'm wondering how would you take, um, 'x1, y1'[writes '(x_1, y_1)*'] and subtract 'x2, y2' [writes '*$-(x_2, y_2)$*'].*

Colin, here, is conveying another understanding for the difference of outputs: a difference of coordinates.

Episode II: Abby (Value-Thinking)

Output

Abby engaged in the task in a way consistent with value-thinking. While at times she described the point, Abby marked the output of inputs *a* and *a+h* along the vertical axis. A depiction of Abby's graphing actions is provided in Figure 4.9. Abby first represented the input *a* by using a self-made ruler to measure the length *a* rightward along the horizontal axis with reference point at the origin (see the horizontal blue line segment in Figure 4.9) and then created a tick mark where the end of the length *a* met the horizontal axis (see the black line and label along the horizontal axis in Figure 4.9). To represent the output *f(a)*, Abby then traced vertically from the marking to identify a point (from the tick mark *a* upward along the green vertical line in

Figure 4.9) and then horizontally to the vertical axis to create her another tick mark (see the green dotted horizontal line and black tick mark along the vertical axis labeled '*f(a)'* in Figure 4.9).

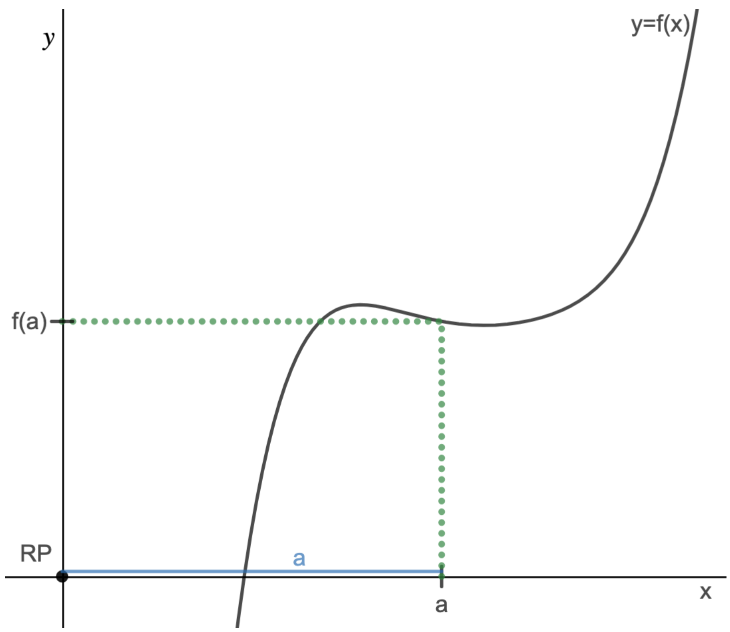


Figure 4.9: A depiction of the results of Abby's graphing activity for f(a).

In addition to identifying a position along the vertical axis as the output of the function, Abby seemed to understand the marking or 'tick' on the vertical axes as measurable. Abby measured the distance from the origin upward along the vertical axes using her self-made ruler. Abby particularly measured this distance in terms of the length of the input *a*, saying that "it'd be whatever this is [gesturing to the tick mark along the vertical axis]… which looks like it is half of one length of *a"*. Similarly Abby also identified the output of *a+h* similarly to how she identified the output of *a,* using the concatenated distance *a+h*. Likewise Abby recognized the marking she made as measurable, where she measured this output in terms of *a*, specifically as about three-fourths *a*.

Difference of Outputs

During this episode, Abby's conveyed meaning for the difference of output was coded as a quantitative difference. When Abby was asked to represent the difference of the output of $a+h$ and the output of a, she understood computations as representing the length of line segments. Abby used the numerical values she found for the two outputs (three-fourths a and one-half a, respectively) and subtracted the two outputs to get one-fourth a. While the computation is Abby did was subtraction, Abby was able to re-present the difference as a length elsewhere on the page, particularly representing the length as a portion of the line segment a, rather than relying on the distance between the output markings. Figure 4.10 is Abby's work after completing the task.

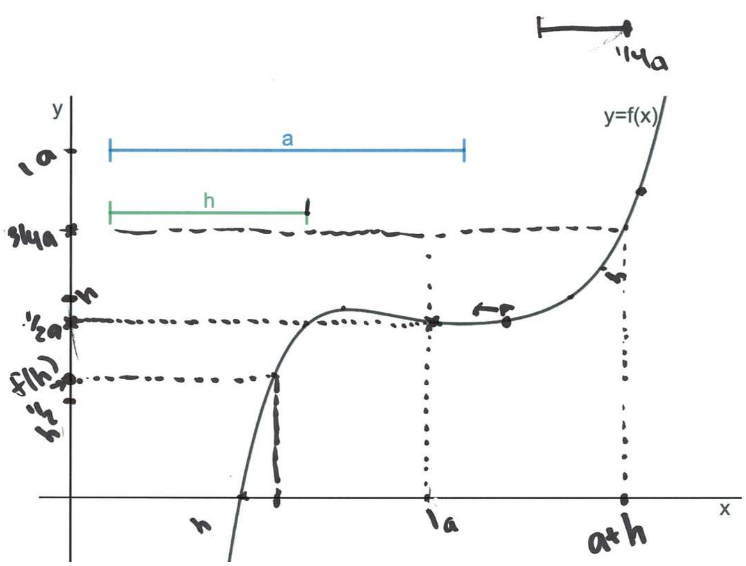


Figure 4.10: Abby's work during the second episode.

Note that the vertical dashed lines and horizontal dashed lines were used to identify a point on the graph and then the output of the inputs h, a, and $a+h$. Further the line segment in the top right corner of Figure 4.10 is Abby's re-presentation of the difference between the output of *a* and the output of *a+h*, which she labels one-fourth of *a*. As Abby wrote created this marking, she indicated the connected between the length she drew and her graph, saying “that would be one-fourth [of *a*] [draws the line segment labeled one-fourth *a* in Figure 4.10] which is the same as this [gestures to the vertical axes in Figure 4.11]”.

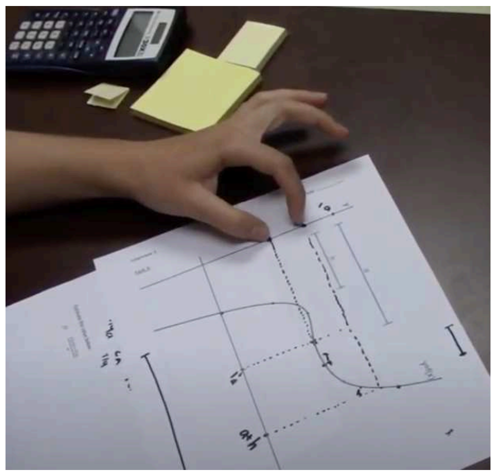


Figure 4.11: Abby's gesture to the vertical axis to represent *f(a+h)-f(a)*

Summary of Abby’s Thinking

Figure 4.12 is a visual model of Abby’s thinking during this episode. Abby’s frames of reference for representing inputs during this episode aligned with the normative frame of reference. Additionally, her attention to the horizontal axis when representing the outputs *f(a)* and *f(a+h)* indicated a meaning for output consistent with what David, Roh, and Sellers (2019) describe as value-thinking. Abby further indicated a recognition of outputs as magnitudes by using the length *a* as a unit to measure *f(a)* and *f(a+h)*. Abby recognized the difference of outputs as a quantitative difference by first subtracting the measures of *f(a+h)* and *f(a)* (three-fourths *a* and one-half *a*) and then connecting her computation to the magnitude represented in the graph (the vertical red line in Figure 4.12).

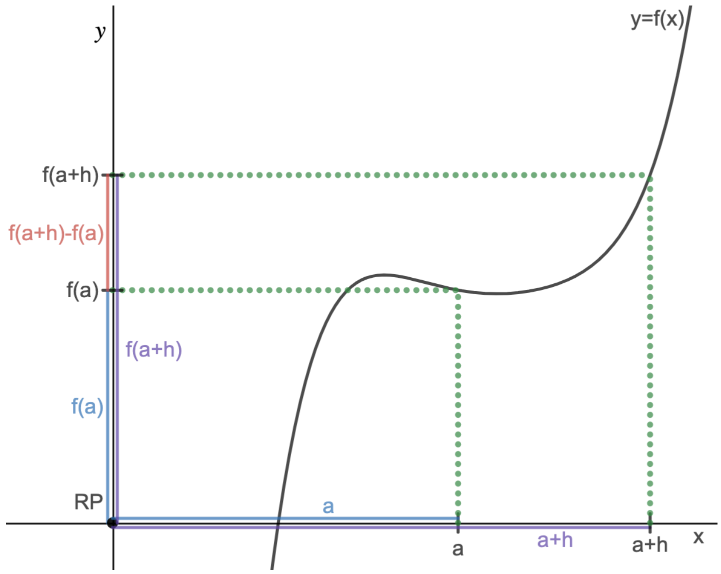


Figure 4.12: A model of Abby's thinking

Episode III: Alison (Other/Arc Length)

Input

When Alison engaged with this task, the meaning Alison conveyed for output aligned with neither location-thinking nor value-thinking. Alison described the output as an arc length of the curve. Prior to representing *f(a)*, Alison attempted three different ways of representing the input *a*, all of which were situated along the horizontal axis. Each attempt also involved measuring rightward along the horizontal axis, although Alison's choice of reference point varied. Alison's three choices of reference points are depicted in red (RP1), blue (RP2), and purple (RP3) in shown in Figure 4.13.

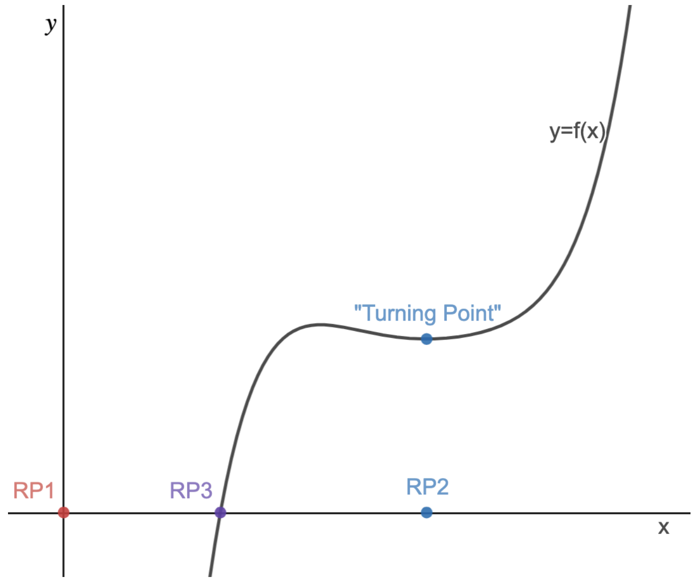


Figure 4.13: A depiction of Alison's three choices of reference points

On the first attempt to represent *f(a)*, Alison measured from the origin, labeling *a* (in black) in Figure 4.14. On her second attempt, Alison identified a point near what she called the "turning point" of the graph, (see Figure 4.13, and the position on the curve with a vertical bar that is near the point of inflection of the function in Figure 4.14) where she created a new marking on the x-axis. Alison then used the position on the horizontal axis below where the turning point is as the reference point, measuring rightward along the axis and creating a vertical dash off of the axis. As shown in Figure 4.14, Alison extended the x-axis to create this marking and extended the graph off of the page by using an additional piece of paper (see Figure 4.15). The meaning Alison conveyed for *a* is represented by the length and position of the red line segment (see Figure 4.14).

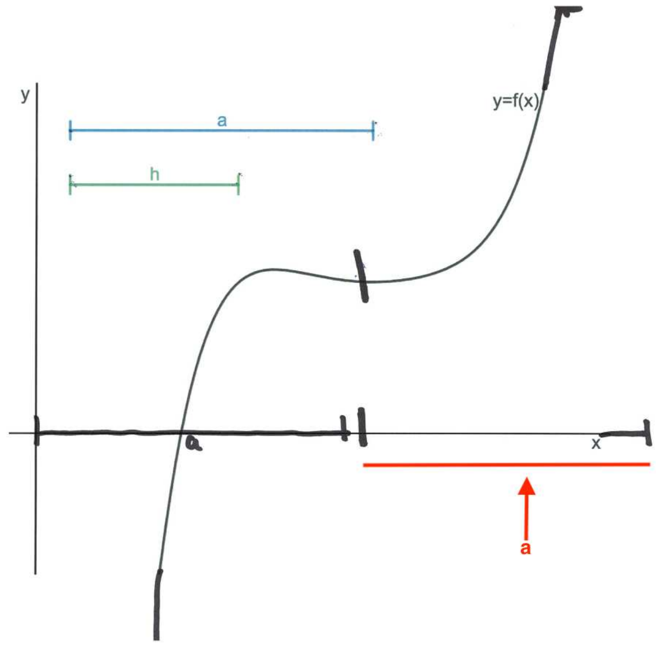


Figure 4.14: Alison's attempts at representing *a* along the horizontal axis

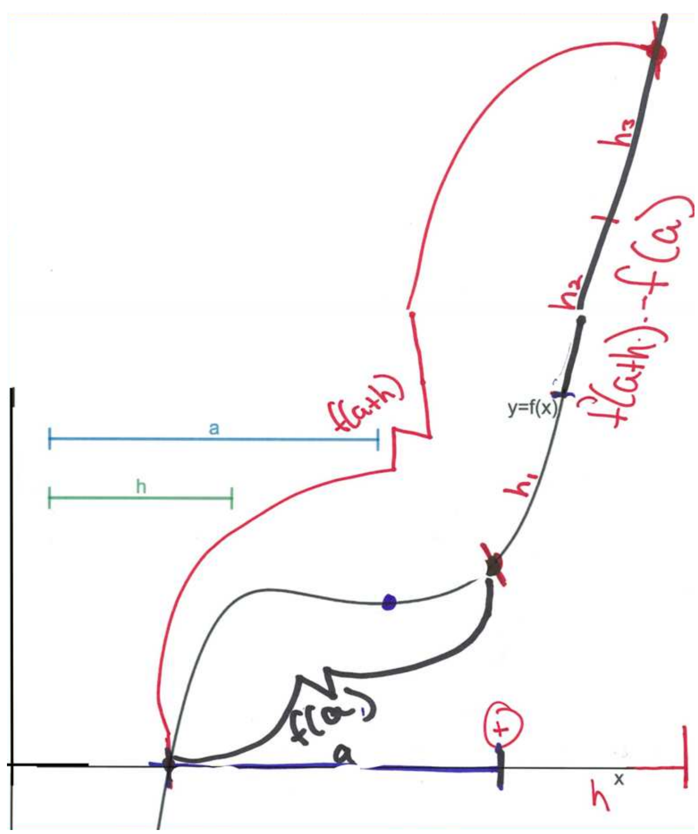


Figure 4.15: Alison's representations of outputs and differences of outputs

Output

Shortly after extending the horizontal axis and the curve, Alison described that what she had done seemed incorrect, and then used the point of intersection of the curve with the horizontal axis as the reference point (see the point 'RP3' in Figure 4.13). Alison indicated that *f(a)* was an arc length along the graph of *f(x)* over a finite interval. After extending the line segment *a* right-ward from the reference point, she labeled the length of the arc from the point of intersection to the point on the curve above the end of *a* (see Figure 4.15). When asked about what *f(a)* referred to, she said "It's that whole thing" while gesturing along the curve from one location on the graph to another, seemingly referring to the arc. At this moment, it was unclear whether Alison was referring to the arc as line or as an arc length, however it seemed that Alison was not understanding output as a point, and hence here activity was not coded as using location-thinking. Further Alison is not marking the output along the vertical axis nor as a directed distance from the horizontal axis to a point, and hence not categorized as using value-thinking. As the interview progressed, it became clear that Alison was referring to the arc length of a portion of the curve.

When representing *f(a+h)*, Alison likewise conveyed a meaning for output aligned with an arc length over a finite interval, as seen in Figure 4.15. As Alison later described, "that whole length would be *f(a+h)*" while gesturing along the curve from the reference point to the end of the curve. Alison represented *a+h* along the horizontal axis by concatenating the lengths *a* and *h* to form *a+h* and measuring rightward from the same reference point. After extending both the horizontal axis and the curve, she represented *f(a+h)* similarly to *f(a)*, using red braces from the point of the intersection to the point on the curve vertically above the end of the line segment *a+h*.

Difference of Outputs

When representing the difference *f(a+h)-f(a)*, Alison focused on the portion of the curve between the end points of the two arcs along the curve (i.e., those identified as *f(a)* and *f(a+h)*). After summarizing her prior work, Alison said "… if we're subtracting, then I would take from this point to this point" while gesturing to the end points of the arcs *f(a)* and *f(a+h)*, shown as red lines through the curve. It is unclear whether Alison is referring to the arc between the two points or the length of the arc, however Alison later indicated that she was referring to the arc length.

When later asked to estimate the value $\frac{f(a+h)-f(a)}{h}$, Alison indicated that the length of the arc was three lengths of *h* long. As seen in Figure 4.15, Alison measured the arc in terms of lengths of h, creating markings to designate the end of each successive length of *h* and labeling the number of lengths using h_1, h_2, and h_3. Given that I saw no evidence of Alison experiencing a perturbation while representing the outputs through estimating this value, I believe that Alison's meaning for output did not shift, and that she understood outputs as arc lengths for the entire duration of this task. It is worth noting that while an arc length can be viewed as a length or a value in relation to a unit of measure, arc length is not the normative understanding of the y-coordinate of a point. Value-thinking is not defined by output of a function being a value; rather value-thinking would, quantitatively, refer to the magnitude of directed from the horizontal axis to the corresponding point on the curve or an equivalent magnitude, potentially positioned elsewhere (e.g., along the vertical axis). The meaning Alison conveyed for output was hence coded as 'other'.

Summary of Alison's Thinking

Alison conveyed a meaning for output aligned with neither location-thinking nor value-thinking. A visual representation of Alison's meaning for output is shown in Figure 4.16, where the output *f(a)* is consistent with arc length. Alison's conveyed meaning for output involved first representing the corresponding input on the graph by (1) identifying a reference point along the horizontal axis, and (2) extending a line segment of appropriate length rightward from the reference point along the horizontal axis. Alison then identified points on the curve that were vertically oriented from the end points of *a*, and then conceived of output of the input as the arc length of the curve between the two identified points on the graph. Alison's conveyed meaning for difference of outputs aligned with Thompson's (1993) quantitative difference, as the extent to which one quantity (e.g., *f(a+h)*) exceeds another quantity (e.g., *f(a)*). As Alison's conception of *f(a+h)* and *f(a)* shared the same initial reference point, the magnitude of the difference of the two arc lengths normatively aligns with the distance between the two end points of the arcs along the curve.

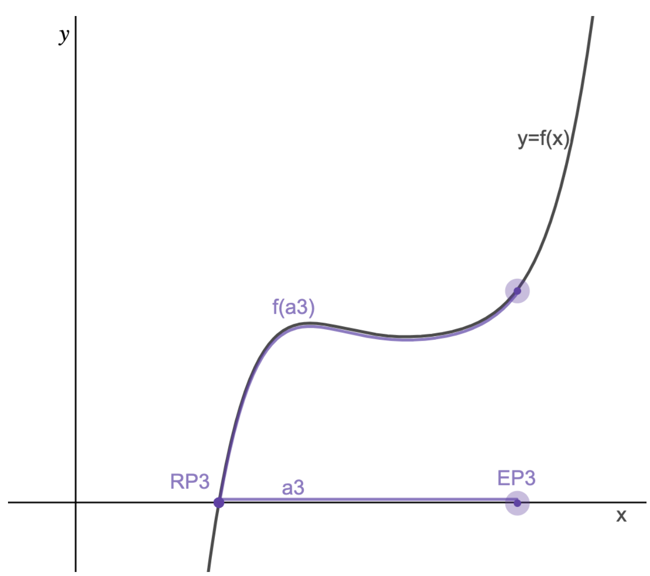


Figure 4.16: Model of Alison's work using her final choice of reference point

The model for Alison's thinking above helps explain why Alison's initial choices of reference points for representing *a* along the horizontal axis were unsatisfactory to Alison. When

Alison used the origin as the reference point ('RP1' in Figure 4.13), Alison was able to identify point on the curve from the end point of *a*, but not the point vertically oriented from origin. Alison attempted to initially extend the graph downward below the x-axis (see Figure 4.14). Since Alison was unable to find where the graph intersected the vertical axis, she could not identify the end of the arc whose length would be the output of *a*, as shown in Figure 4.17a. When using her second choice of reference point, Alison needed to extend both the horizontal axis and the curve to represent the input *a* along the axis and to identify the end of the arc that I would expect Alison to identify as the output. She would have had to extend the axis and curve even further to represent the arc whose length is *f(a+h)*. According to this model of Alison's thinking, she would also be unable to identify the end of the arc whose length is *f(a+h)* without extending the horizontal axis and the curve more than she already had, as shown in Figure 4.17b.

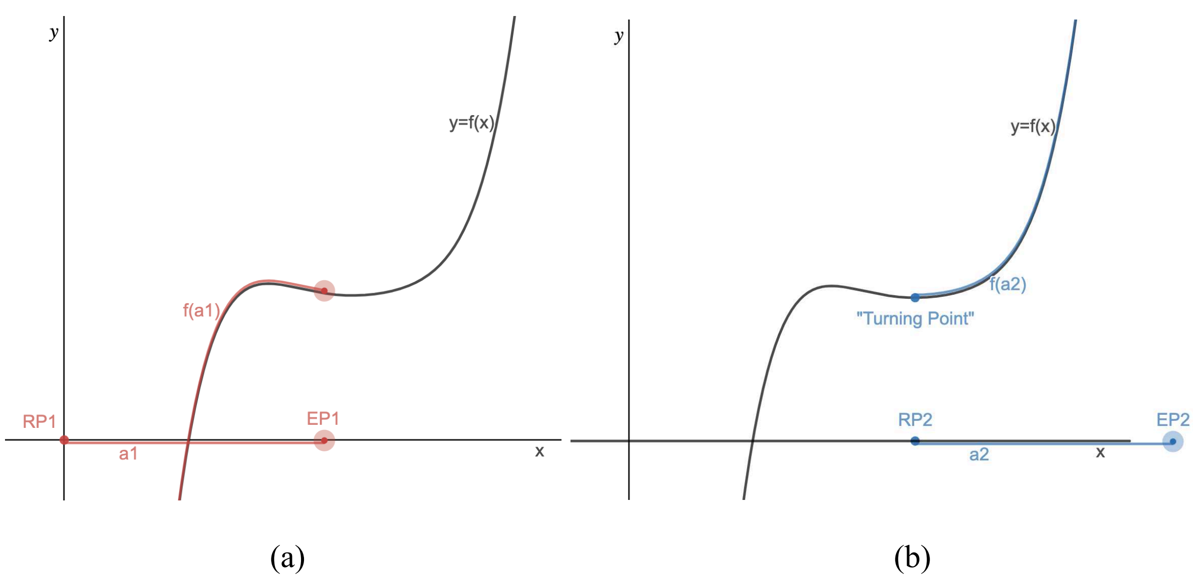


Figure 4.17: Model of Alison's meaning for output for her first (a) and second (b) choices of reference point.

Episode IV: Lisa (Other/Arc Length)

In this episode, Lisa does not convey a clear meaning for how to represent input values for the function *f(x)* in the graph. Hence, I do not include Lisa's meaning for input. However, I use frames of reference are used to describe Lisa's meaning for output.

Output

Lisa conveyed a meaning for output consistent with an arc length in a way that is distinct from how Alison represented output. The arc length Lisa identified was precisely the length of the input oriented along the curve. Lisa represented the arc length used the lowest point on the curve as the reference point for her measurements throughout the entire task. Figure 4.18 represents Lisa's activity with the task.

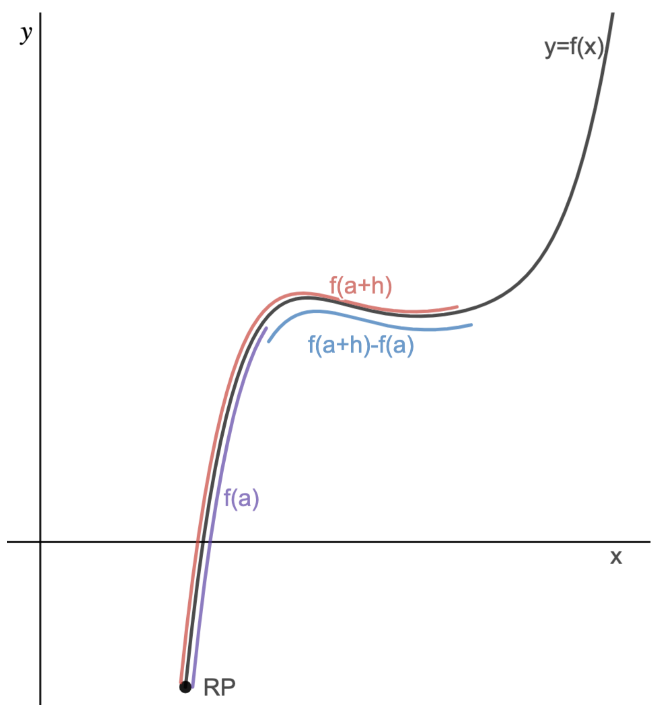


Figure 4.18: A model of Lisa's activity

Lisa represented *f(a)* by measuring along the arc from the reference point for a length *a* using a self-made ruler, where *f(a)* is depicted in purple in Figure 4.18. In the follows passage, Lisa indicated that *f(a)* is the arc length.

Lisa: Yeah, it would be like the whole distance from here [gesturing to the lowest blue 'x' in Figure 4.19] to here [gesturing to the other blue 'x' in Figure 4.19]..... I guess I should have written it over here [Lisa writes 'f(a)' near the x-axis] cause I wrote it as that point is 'f(a)', but it's like this whole thing

Interviewer: Okay, it's [referring to f(a)] the length of that line?

Lisa: Yeah.

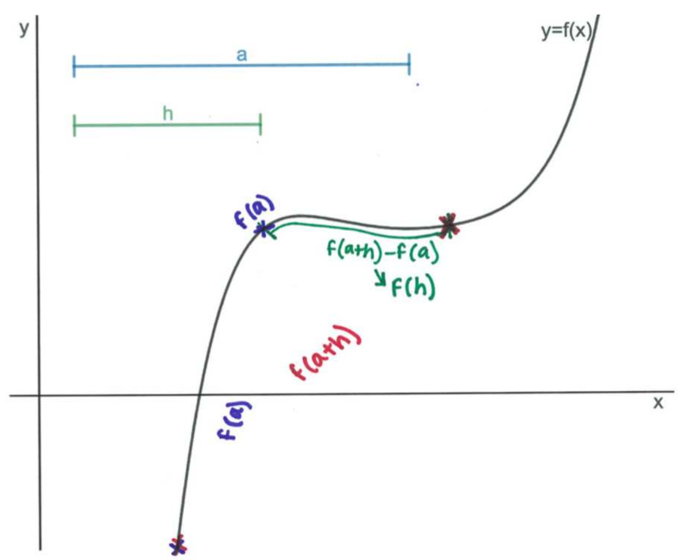


Figure 4.19: Lisa's work

Lisa seemed to be suggesting that the point along the curve marked with a blue 'x' is the end point of the line segment. It seemed that the line segment is the focus of Lisa's attention, and the end point is a feature of the line segment, instead of the focus of her attention. Particularly the blue ‘x’ markings represented the end points of the curve whose length was *f(a)*.

Lisa similarly conveyed a meaning of *f(a+h)* as the arc length along the curve of length *a+h* from the same reference point (depicted in red in Figure 4.18, and between the red ‘x’s in Figure 4.19). She created the length *a+h* by concatenating the lengths *a* and *h*. Lisa had difficulty representing *f(a+h)* given that the curve was not straight. Lisa then described a process of leveraging her previous work by (1) using the endpoint of the arc of length *f(a)* as a reference point, (2) measuring a distance of *h*, and then (3) creating an ‘x’ at the end point of the arc of length *f(a+h)*, where the *f(a+h)* was the distance along the curve from the original reference point.

> *Lisa: It would be easier to just measure from this point [gesturing to the endpoint of the arc with length f(a)] [for] this length [gesturing to a length ‘h’ in the prompt] to here [gesturing to what will be the endpoint of the arc of arc length f(a+h)]. Cause then from here to like over here [gesturing from the reference point and end point of the arc of length f(a+h)] it would be 'a+h'. Then it'd be from this x [the initial reference point in Figure 4.19] to a point over here [gesturing to where the endpoint of the arc of length f(a+h) will be in Figure 4.19].*

This segment highlights the sophistication of Lisa’s measurement process, as she recognized that she could also produce the end point of the arc with length *f(a+h)* by measuring an additional length of *h* from the end point of the arc with length *f(a)*.

Difference of Outputs

Lisa's description of the difference *f(a+h)-f(a)* aligns with what Thompson (1993) would call a quantitative difference of the two arc lengths, particularly the extent to which one arc length is larger than the other. She represented the difference as the arc length between the end points of the arcs of length *f(a+h)* and *f(a)*. Lisa then labeled the length of the curve in green in Figure 4.19, also in blue in Figure 4.18.

In the following except, Lisa described the difference *f(a+h)-f(a)*. Lisa gestured between the end points of the arcs of length *f(a)* and *f(a+h)*, referring to the length by which *f(a+h)* exceeded *f(a)*. Lisa was doing an additive comparison of the arc lengths *f(a+h)* and *f(a)*, specifically describing the length by which *f(a+h)* exceeded *f(a)*.

> *Lisa: Um, well, it would be.. So I have f(a+h) [gestures to the end points of the arc of length f(a+h)]... and then it would just be subtracting the distance of a, so would really just be the distance of h. So we just be…. from here to here [Lisa gestures from the end of the arc of length f(a) to the end of the arc of length f(a+h)]... Yeah. Cause this is 'a+h' [gesturing along the arc with length f(a+h)]. And then if you take away 'a', which is here [gesturing the length a given in the prompt], then it would just make this distance right here [gesturing to the end points of the arcs of length f(a) and f(a+h)].*

Much like how Lisa understood *f(a)* and *f(a+h)* as distances, Lisa conveyed a meaning of *f(a+h)-f(a)* as a distance, specifically an arc length. Lisa gestured to *f(a+h)-f(a)* similarly to when she used the end point of the arc of length *f(a)* to represent *f(a+h)*. However, Lisa used a frame of reference for $f(a+h)-f(a)$ where the reference point was the endpoint of the arc of length $f(a)$ rather than the point pf intersection.

Summary of Lisa's Thinking

Lisa's meaning for output did not align with location-thinking or value-thinking. Lisa conveyed a meaning for outputs of the function $f(x)$ as arc lengths. Lisa used 'x's to denote the endpoints of the arc of appropriate length and identified the arc length as the output of the function. Lisa's sophisticated reasoning for identifying the endpoint of the arc of length $f(a+h)$ and her coordination of frames of reference indicated that Lisa's meaning for output was stable. Lisa further conveyed a meaning for difference of outputs as a quantitative difference, where she compared the arc lengths additively. Lisa's meaning for difference of output also seemed to be stable. Later in the interview, Lisa derived a theorem: $f(a+h) - f(a) = f(h)$. When the interviewer asked about this, Lisa replied:

> *Lisa: ..cause this [gestures to the endpoints of arc with length f(a+h)-f(a)] is …the same distance as here [Lisa translates her index and middle fingers to the endpoints of arc with length f(a+h)-f(a)], so that's 'f(h)' cause then … the 'a's right here [Lisa gestures to her writing 'f(a+h)-f(a)'] would, like, cancel each other out, kind of.*

In this except, Lisa was assimilating the result that $f(a+h)-f(a)$ was equivalent to $f(h)$ to 'canceling the a's' in the symbolic expression. Given that many students refer to simplifying algebraic expressions with the phrase 'canceling', Lisa seemed to be identifying this result to a symbolic operation that she has does in the past.

Episode V: Abby (Other/Point)

Input and Output

In Abby's initial activity, indicated a non-normative frame of reference for representing input. Abby used the point of intersection between the curve and the horizontal axis as the point of reference, which coincides with Alison's third and final choice of reference point. Abby further indicated a directionality for her frame of reference along the curve, generally in towards the portion of the graph to the top-right. Abby's work and a model of Abby's thinking are provided in Figure 4.20a and Figure 4.20b, respectively.

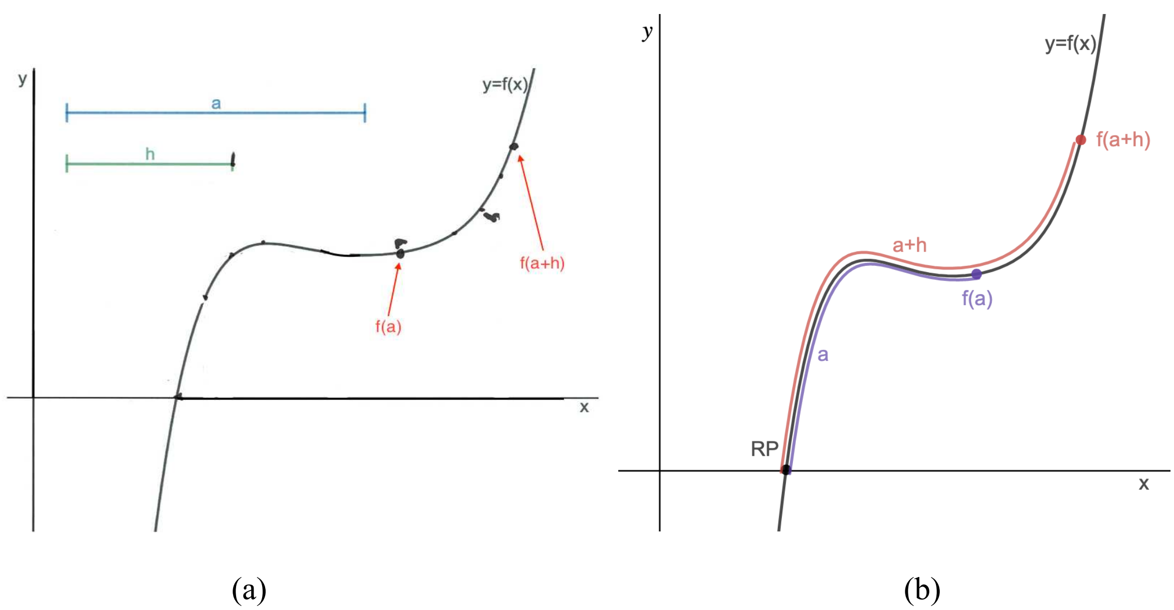


Figure 4.20: Abby's Work (a), and a Model of Abby's Meaning for Output (b)

To represent *f(a)*, Abby measured a distance of length *a* along the graph, from the reference point where the curve intersects the *x*-axis to the point on the curve that forms an arc length of length *a*. Abby discussed the process of identifying the point through a measurement process along the curve akin to measuring the "length of a trail" on a map, stating:

Abby: So then this [length] *a*, if you're doing like the maneuvering like I would do if I was trying to, like, figure out the length of a trail, this would be the length of *a* from here to here [gesturing along the curve]... like I took like the whole length of *a* here [gesturing to her reference point] and then all I do is, like, on the actual line itself, like, just continue to, like, align this, like, measurement that I'm using [gesturing to her self-made ruler] with the actual line, and just continuously turn it until I get to [length] *a*.

Here, Abby is describing the process of measuring a curved line using a straight ruler. Abby measured an arc length of the curve from the reference point for a length of *a*, and identified the end point of the arc, indicating that this represented $f(a)$. To identify the point $f(a+h)$, Abby first formed $a+h$ by concatenating the lengths *a* and *h*, and then measured an arc length from the same reference point along the curve for a length of $a+h$. The end point of this segment was identified as $f(a+h)$. Prior to representing the differences of $f(a+h)$ and $f(a)$, the interviewer chose to intervene with the student to propose another way of thinking of the lengths of the line segments *a* and *h*. A model for Abby's conveyed meaning for output are depicted in Figure 4.20b.

Discussion

Output of a Function

This analysis aligns with other research that suggests that calculus students bring a variety of understandings of output of a function to the classroom (David et al., 2019; Moore & Thompson, 2015). David, Roh, and Sellers (2019) described students' conceptions of output of a function as being represented by a point on a curve with the corresponding x-value (location-thinking) or as the y-coordinate of this point (value-thinking), and the data in this report suggests that these meanings for output are conveyed in the context of non-numerical graphs as well. The results of analysis additionally provide an existence proof for other conceptions of the output of a function as well, that I will call *arc length-thinking*. Two of the students (Alison and Lisa) conveyed a meaning for output as an arc length of the curve, particularly when they first engaged with this activity. These conceptions for Alison and Lisa were stable meanings as both students were able to represent the difference of these two outputs without being perturbed and were consistent with their conceptions of output. For this reason, it does not seem likely that this is the first time that Alison and Lisa have reasoned in this way, and Alison and Lisa were assimilating to a scheme.

Another student (Abby) also conveyed a difference meaning for output, particularly as a point on the curve. While in some ways similar to location-thinking, Abby's thinking was distinct as she represented input values along the curve rather than parallel to the horizontal axis. Specifically, Lisa used a reference point at the intersection of the curve with the horizontal axis. Like the Alison and Lisa, Abby's conveyed meaning during the first episode was stable. Though her meaning for difference of output was not conveyed, Abby indicated assimilating the task to measuring a trail, suggesting that her meaning for output was stable during her activity. While this analysis framed student's location-thinking to specifically use the normative frame of reference, there is room for considering this thinking as location-thinking using a non-normative frame of reference. Future research may want to consider location-thinking and value-thinking in terms of non-normative frames of reference.

Difference of Outputs

Throughout the seven episodes where students conveyed a conception for the difference *f(a+h)-f(a)*, five of the episodes involved students conveying conceptions aligned with a quantitative difference, while none of the episodes involved students conceiving of difference as subtraction. As this task was designed to entail quantitative reasoning, it is not surprising that most students viewed the difference of outputs as a quantitative difference, where those quantities were what the student identified as an output.

In these episodes, two students conceived of the difference of outputs as a point on the curve, both of whom were using location-thinking. Both students came to understand difference in this way by focusing on variation along the horizontal axes and conceiving of the difference of outputs in a way consistent with the outputs of the difference in inputs. While the students did not conceive of *f(a+h)-f(a)* as a quantity, they represented the input of the output *f((a+h)-a)* in a way consistent with a quantitative difference of *a+h* and *a*.

In addition to quantitative difference, Thompson (1993) also described subtraction as a meaning for difference. This meaning for difference, however, was not present in this data set. While this could be a feature of the sample size (five students), I also believe that the task influenced this. The task was intended to necessitate quantitative reasoning by removing numerical markings from the axes and representing quantities with magnitudes. Without available measures, students would need to take it upon themselves to choose a unit of measure to measure as all lengths in reference to (which in terms of frames of reference involves committing to a unit (Joshua et al., 2015)). If a student were to do this, I believe it would be unlikely that they would forget what the numerical markings meant, as they would have created them.

Relationship Between Output and Difference of Output.

Of the seven episodes where students conveyed conceptions for the difference of outputs, students also conveyed conceptions for output. Students that described *f(a+h)-f(a)* as a quantitative difference either used value-thinking or indicated a meaning for output as an arc length. In each of these cases, students represented outputs as the line segment and represented the difference as an additive comparison.

Additionally, in each episode where students described the difference of outputs as a quantitative difference, the students conveyed a stable meaning for both output and difference of output, whether representing output in a standard way (e.g., value-thinking) or non-standard way (e.g., arc-thinking). I hypothesize that these were stable meanings for the output of a function because of the consistent nature of the students reasoning, as these students represented both the output and the difference of outputs as magnitudes. The consistency of students' conceptions and the lack of evidence of the students experiencing perturbations during this activity suggest that the students were accommodating to a scheme.

Both episodes where students used location-thinking also involved the student indicating that *f(a+h)-f(a)* was a point on the curve. These were also the only students who represent the difference in this way. Both students seemed to focus on the variation along the horizontal axis and represented the difference of output in a way consistent with the output of the difference of inputs. Students' understanding of the difference of output as a point, however, was in the moment, as both students seemed perturbed and uneasy with their claim. In fact, after reflecting, Colin considered an alternative understanding for difference of output, particularly as a difference of coordinate-pairs. This illustrates that Colin's conception of the difference of outputs was unstable (i.e., in the moment) and that a difference of coordinate-pairs is another potential meaning for difference of outputs, specifically when using location-thinking. Colin's deduction that the difference of outputs would need to be the difference of coordinates pairs (based on his meaning for outputs as points) was "unfamiliar" to him. Colin then accommodated his meaning for output to align with value-thinking.

Conclusion and Implications

My work builds on David, Roh, and Sellers's (2019) constructs to describe student's meanings for outputs of graphs, particularly in quantitative contexts. David et al. describe students' meanings for output as points (location-thinking) and y-coordinates of points (value-thinking). Both Alison and Lisa's thinking suggest another meaning for output: arc length of a curve, or arc-thinking. These students both used non-normative frames of reference for representing the input value when conveying this meaning, however, there is evidence that Alison's meaning has the potential for being consistent with a normative frame of reference for input. Alison's initial choice of reference point was normative, however she changed to another reference point. My model for Alison's thinking would suggest that using the origin as reference point was problematic because she could not identify one end of the arc length she would be considering. This particularly was a feature of the task, as the curve did not visibly intersect the vertical axis. I conjecture that Alison would have not changed her reference point if the curve intersected the vertical axis and had been visible continuous and, hence, would have used a normative frame of reference while representing the output of a function using arc-thinking.

Educators can use this data by first acknowledging that students in the classroom may have non-normative meanings for how outputs of functions are represented in graphs. This research aligns with the work of others to suggest that a meaning for graph should not be taken as shared (David et al., 2019; Moore, Stevens, et al., 2019; Moore & Thompson, 2015).

Understanding the ways that students understand graphs of functions and how graphs represent quantities (e.g., value-thinking, arc-thinking, location-thinking) can help teachers anticipate these meanings in the classroom, supporting the interpretation of students' activity and reasoning about graphs during in-the-moment instruction. Teachers can further use this information to strategically choose examples of functions such that the graph of the function under a particular coordinate system may entail different variation in the output quantities depending on how the student is conceiving of output, rather than an example that may be ambiguous under different meanings for output. For example, the graph of the linear function $f(x) = 2x - 3$ in Cartesian coordinates can similarly be seen as increasing at a constant rate of change by a student engaging in value-thinking (see Figure 4.21a) or by a student who is conceiving of the arc length of a curve using a normative or non-normative reference point (see Figure 4.21b for a non-normative reference point).

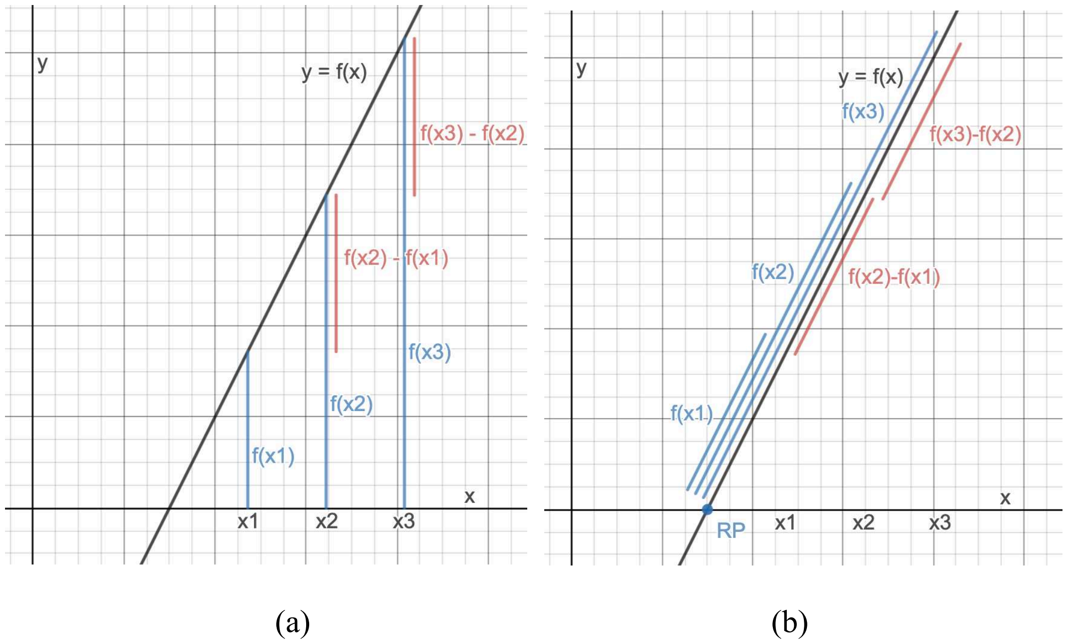


Figure 4.21: Linear change in output when using value-thinking (a) and when output using arc-thinking (b).

To support students in accommodating their meanings to be more productive, it is essential that students find their reasoning problematic, or at least different from what is considered normatively correct. Using such an example may not support students recognizing differences in their reasoning compared to their peers, and likely would not support students in transitioning from their non-normative meaning for output. Using examples that entail different variation may afford classroom conversations about graphs, how mathematicians represent data, and how the instructor and the mathematical community intends to depict variation between two quantities using a Cartesian coordinate system, thereby supporting students in transitioning to normative conceptions of output and the difference of output.

Differences of Outputs and Learning in Calculus

This work contributes to the body of research around students' thinking of concepts in calculus. To help illustrate how these outputs of a function and the differences of output relate to rate of change of a function, I draw on Zandieh's (2000) work, where she deconstructs the concept of derivative as a function into three layers: function, limit, and ratio. These layers are connected in the framework through process-object pairs, where a process conception of one layer can draw on the object (or pseudo-object) conception of another layer, forming a chain of process-object pairs (ratio to limit to function). In this way, Zandieh says that the derivative *function* is the *limit* of a *ratio*. In a symbolic context, this ratio is often referred to as the difference quotient or $\frac{f(b)-f(a)}{b-a}$. Work by Byerley (2019) and Byerley and Thompson (2017) describe how students' conceptions for slope, fraction, and measure can impact their conceptions for rate of change. They argue that students' understanding of a rate of change is impacted by their conceptions of a ratio, which is the ratio of changes of an output quantity with respect to changes of the corresponding input quantity, or the change in y with respect to the change in x. Extending Zandieh's (2000) chain, one can argue that the ratio is a ratio of differences, specifically the changes in one quantity (output quantity) in relation to the changes in another quantity (input quantity). The numerator of this fraction is precisely the difference in outputs of a function, and students' conceptions of these differences (and hence outputs) are essential in students' conceptions of the derivative.

This analysis also provides evidence of some of the downstream conceptions that one may expect for the difference of outputs for those using value-thinking, location-thinking, and arc length-thinking. In a quantitative context similar to the one described in this paper, educators may expect students focusing on this difference of output as the magnitude of a vertical line segment as being indicative of value-thinking, and students focusing on this difference as the magnitude of an arc as being indicative of arc length-thinking. Students focusing on difference of output as a point or as the subtraction of coordinate-pairs may be indicative of location-thinking.

Given the limited sample size, however, I suggest further research to document other meanings that students may convey for difference of output, as well as other concepts along Zandieh's (2000) chain such as average rate of change and instantaneous rates of change.

Frames of Reference and Supporting Accommodation to Normative Meanings

This analysis suggests that the origin as the normative reference point for graphs in Cartesian coordinates is non-obvious for students. When first engaging with this task, four of the five students reasoned about the quantities represented in the graph using non-normative frames of reference, particularly non-normative reference points for representing the input of the function. Additionally, two of the five students conveyed a directionality for their frame of reference along the curve rather than along (or parallel to) the horizontal axes. While this perspective is productive when considering arc length or parametric equations, teachers may want to support students in coming to normative frames of reference when interpreting graphs of functions in Cartesian coordinates. I suggest teachers provide students opportunities to reason with quantitative situations and create graphs that represent the relationship between the covarying quantities in the situations. Creating graphs would likely support students in coming to normative frames of reference, especially if students first created graphs with numerical axes. Through such activity, students would need to reflect on how quantities are represented in a graph, further supporting the identification of normative meaning of output of a function.

To further support students coming to normative conceptions of graphs in the classroom, instructors can also be careful to label the axes of graphs as well as the curve of a graph in Cartesian coordinates as '$y = f(x)$' rather than '$f(x)$'. Such precision can support students in connecting their mathematical meanings across algebraic and graphical contexts. For example, Colin leveraged this expression as he transitioned from location-thinking to value-thinking.

After being perturbed when considering $f(a+h)-f(a)$ as a difference of coordinates, Colin began to consider the output as the y-coordinate of the point he had been thinking was the output. Colin said the following when reflecting on this moment.

> *Colin: … If you'd asked me at the beginning, having not just thought about it for a long time, [I] would have said that it [f(a)] was the point. But now that I think about, like, it's like $y = f(x)$, and all we did was substitute the a there, so I would say that this [f(a)] is the y-value.*

Colin provides an example of how with a student with the appropriate supports can naturally transition from unproductive meanings to productive meanings.

Chapter 5

Conclusion

Discussion

This dissertation identifies factors related to student success in Calculus I with the intention of informing interventions for students that may be at-risk of failing Calculus I. The first paper suggested that students who regulate their precalculus content knowledge tend to perform better in Calculus I than students that do not. However, due to coarse nature of the coding scheme for the SRL score, it is difficult to give targeted advice to students identified as at-risk given both the complex nature of SRL and the limited information captured by the tools. The second paper in this dissertation helps clarify some of the relationships that were involved in the first paper. Drawing on students' calibration of their precalculus content knowledge, my work suggests that students who tend to underestimate their precalculus mastery tend to have higher scores on Exam 1 than students who overestimate their abilities. Students who tend to be more accurate in their precalculus mastery also tend to perform better on Exam 1 than students whose perceptions of precalculus mastery are inaccurate. Students who are more accurate their assessments of precalculus mastery further tend to benefit more from help-seeking than students who are inaccurate in their assessments of precalculus mastery. This deeper analysis of calibration and help-seeking provides additional recommendations for students who have be identified as at-risk, such as the hypothetical student Derek.

Recall that Derek is a student with high precalculus confidence, low demonstrated mastery, and who has neither reflected on his precalculus content quiz results nor sought help to improve his precalculus mastery. Given Derek's precalculus confidence and mastery, in the first paper I suggested that Derek could regulate his precalculus knowledge by engaging in the reflection phase of SRL and using the feedback from the precalculus content quiz to inform his future studies. Derek could do so by reflecting on his quiz score by using the precalculus reflection tool and seeking help with precalculus material either online or through in-person precalculus-focused workshops. By doing these actions, Derek would have a higher SRL score. The second paper elucidates on how instruction and intervention could better support Derek. I found that students who are inaccurate in their calibration may not benefit from help-seeking as much as students who are more accurately calibrated. This could be the case because students who are attuned to their content knowledge would more likely be efficient and accurate in identifying what content areas would be most productive to focus on during a tutoring session, thereby increasing the effectiveness of help-seeking. As Derek exhibits low mastery and high confidence in his precalculus ability, Derek would likely be categorized as having high calibration bias and low calibration accuracy. Results from the second paper suggest that not only should Derek seek help to improve his precalculus content knowledge, but that helping Derek become more accurate in his abilities also plays an important role. While seeking help from various resources would likely support his growing precalculus content knowledge, his perceptions of mastery may hinder the effectiveness of that help. Specifically, Derek's inaccurate calibration may inhibit him from targeting the content areas that would be most beneficial for his conceptual growth. By improving Derek's calibration accuracy, Derek will be better positioned to identify the content with which he needs support.

While the second paper highlights the importance of calibration, past researchers suggest that a strong normative foundation in precalculus content knowledge is also important for success (Agustin & Agustin, 2009). Quantitative data from the first two papers suggest that many students entering calculus have not yet mastered concepts related to graphs. As described in the analysis of the PCQ in the introduction to this dissertation, less than 12% of students enrolled in a calculus course who took the precalculus content quiz answered both items related to graph correctly, and over 60% of students made an incorrect response to one of these multiple answer items (i.e., answered at least one question incorrectly). Clinical interviews in the third paper of this dissertation corroborates this finding, as all five Calculus I students demonstrated non-normative conceptions for output when first engaging with the task. Not only were the conceptions of output of a function incorrect, but one category of these conceptions (i.e., arc length-thinking) was stable, where the students had consistent and coherent ways of making sense of difference of outputs throughout the task. The stability of students' non-standard conceptions of graphed quantities supports quantitative findings that students in Calculus I, in general, had low levels of calibration accuracy and high levels of calibration bias on items related to graphs. Together, these findings from this quantitative and qualitative data suggest that graphs of functions may be one content area to target when trying to improve students' calibration accuracy because (1) there seems to be room for growth for many Calculus I students and (2) incorrect meanings may engender stable conceptions that are counterproductive to students in the course. Students like Derek may have a consistent system of conceptions for precalculus content, such as graphs of functions and how quantities are represented in graphs, but have not yet recognized that these conceptions will be problematic. Once students recognize that their non-normative meanings are unproductive, they can begin to accommodate meanings to ones that are more productive and standard. Until those students find their meaning unproductive, however, students with stable non-normative meanings for graph would likely exhibit high confidence in their abilities to answer items related to graphs correctly.

I hypothesize that students' non-normative, yet consistent, conceptions of graph may explain why students in Calculus I are inaccurate in their perceptions of mastery of graph, where students' high calibration bias on graphing items (in the introduction to this dissertation) may be indicative of non-normative stable meanings (in the third paper of this dissertation). While this dissertation study provides evidence toward this claim, further work ought to investigate the relationships between students' calibration bias and their stable non-normative meanings of graph. To further investigate whether high calibration bias is indicative of non-standard meanings of graphs, I suggest that future researchers conduct clinical interviews with students where students indicate their confidence in content related to graphs and answer questions related to graph. I further suggest that future researchers measure students' calibration bias and accuracy using students' confidence of correctly completing the task before seeing the item (i.e., predictions) and after responding to the item (i.e., *postdictions*; Labuhn et al., 2010; Schraw et al., 1993). Measuring the alignment of students' perceptions of mastery after the task with their observed mastery would give insight into students' perceived success *after* they have already completed the items rather than before they have seen the items (e.g., in this dissertation). It is unclear whether students' confidence in answering questions related to graphs in this study was related to their perceived success with the task or whether the task was unexpected for the student. Using students' confidence in having completed the item correctly after answering the question would more likely capture data about students' perceived success with a given task.

I suggest that future work investigating high calibration bias and stable non-normative meanings for graphs additionally account for students' motivation for learning mathematical content. Unlike students in the first and second papers, the participants in the third paper were enrolled in a calculus course that is a terminal course for many programs of study that the course serves. As such, students' motivation to learn mathematical content may differ for those in the third paper and those in the first and second papers. Similarly many students in the calculus course for biological scientists (in the third paper) defer to take the course in their third or fourth year of their undergraduate career, while most students in the calculus course for engineers (in the first and second papers) take the course in their first or second year. These differences in the courses and their student populations may indicate differences in students' intrinsic interest in learning the calculus content and differences in students' mathematical preparation for the respective courses.

Implications

Graphs are commonly used in the Calculus I classroom, and it is important that students understand graphs in a productive and normative way. Results from this dissertation suggest that interventions for at-risk Calculus I students should aim to improve students' calibration accuracy of precalculus content, both generally and specifically with graphs of functions. While graphs of functions should be a focus for such interventions, it would also be beneficial for instructors to consider integrating intentional time in class to focus on how graphs of functions represent information about how two quantities covary early in the semester, such as on the first day, and throughout the semester. Qualitative data suggest that Calculus I students may hold a variety of conceptions for graph, and quantitative data could indicate that large numbers of students hold such non-normative meanings, as evidenced by the high numbers of overconfident students enrolled in Calculus I. By integrating activities and discussion around graphs of functions into class, a large proportion of Calculus I students can potentially be perturbed from non-normative to normative conceptions of graph. Further, early and regular activity around graphs in the classroom could support instructors’ understanding of how students in the class make sense of graphs, which in turn could impact how they facilitate the development of students’ normative conceptions of graph. Use and sense-making of multiple representations have been stressed in elementary through secondary literature (National Council Of Teachers Of Mathematics, 2000), and my work suggests that continued use of graphical representations and cohesion with other representations is also crucial for student learning.

On the first day, instructors could have students work in small groups to identify different quantities in graphs and discuss their results with other groups. Instructors can strategically choose functions whose graphs would elicit different responses depending on various expected conceptions that students hold, such as value-thinking, location-thinking, or arc length-thinking. Further, the instructor could provide graphs of functions devoid of numerical axes to have students’ think about how graphs represent information. Discussing relationships between visual representations and other representations of function, such as algebraic representations, may support students in connecting their graphical meanings for function with other representations to support more coherent meanings across representations (Knuth, 2000).

Additionally, this work highlights the importance of one aspect of SRL: Calibration. Students looking to regulate their content knowledge must be aware of their content knowledge, and it is productive for students’ judgements about their content knowledge to be accurate. The analysis in the second paper provides description for how calibration and help-seeking together correlate to student success. Instructors can support their students in becoming more accurate in their perceptions by having students reflect about their mathematical content knowledge regularly inside and outside the classroom (Bol et al., 2012; Zimmerman et al., 2011).

Zimmerman et al. (2011) found that students who tended to reflect more on assignments tended to be less overconfident in their abilities and performed better on exams than students who did not reflect as much. By having students practice judging their own mastery and reflect more, instructors can foster an environment where students can improve their calibration accuracy (Bol et al., 2012), thereby improving the effectiveness of their academic help-seeking. Given the detailed suggestions we can make about these calibration and help-seeking, I suggest that researchers looking to identify at-risk students using SRL look at how specific aspects of SRL impact student success rather than SRL as one construct. This way, educators and administrators can use this information about these competencies to identify supports for students, similarly to what I described for Derek.

Future Work

In future work, I plan to investigate how different meanings students' have for graphs engender conceptions for other mathematical concepts, such as average rate of change and the derivative at a point. For example, Stump (2001) described a preservice teacher's conception of slope of a function consistent with the length of the hypotenuse of a triangle rather that the ratio of vertical change to horizontal change. I suspect that this conception of slope as hypotenuse may be related to non-normative conceptions of output on a graph because slope as hypotenuse would be consistent with arc learning-thinking. In other future work, I will investigate what supports students in transitioning from non-normative conceptions of output (e.g., location-thinking and arc length-thinking) to normative ones (e.g., value-thinking).

Regarding calibration, I will continue to examine the role calibration plays in student success and help-seeking, as well as how to support students in becoming more accurately calibrated. Research suggests that students who reflect more frequently tend to have higher levels of calibration compared to those who do not reflect as frequently (Zimmerman et al., 2011). Consequently, I recommend that instructors provide more opportunities for students to reflect on their content knowledge both inside and outside of class. For example, Pilgrim et al. (2020) describe questions intended to prompt students into thinking about whether they can successfully complete an item similar to a previously worked problem without additional support. Likewise, researchers have found that prompting students to identify what concepts they grasped and what their strengths and weaknesses were supported students in becoming more accurately calibrated in both individual and group settings (Bol et al., 2012).

Lastly, I plan to investigate effects of self-efficacy with help-seeking in order to determine how this correlates with student performance. Self-efficacy has also been correlated with both student success and self-regulation (Tian et al., 2018; Worthley, 2013). Incorporating self-efficacy into the models used in the second paper, after accounting for incoming ability, may clarify the relationship amongst these three competencies and illuminate nuanced suggestions for intervening with students. For example, helping Derek become more accurately calibrated would likely involve decreasing Derek's self-efficacy. It would be important to understand how Derek's deflated self-efficacy would impact his performance in the course. Likewise, interviews with students would help uncover other potential effects that calibration, self-efficacy, and help-seeking have on students' affect with Calculus I and persistence into Calculus II.

References

Agustin, M. Z. N., & Agustin, M. A. (2009). Algebra and precalculus skills and performance in first-semester calculus. *International Journal of Case Method Research and Application*, *21*(3), 232–236.

Asiala, M., Cottrill, J., Dubinsky, E., & Schwingendorf, K. E. (1997). The development of students' graphical understanding of the derivative. *The Journal of Mathematical Behavior*, *16*(4), 399–431. https://doi.org/10.1016/s0732-3123(97)90015-8

Baker, R. S., Corbett, A. T., & Koedinger, K. R. (2004). Detecting student misuse of intelligent tutoring systems. In *International conference on intelligent tutoring systems* (pp. 531–540). Springer. https://doi.org/10.1007/978-3-540-30139-4_50

Bandura, A. (1981). Self-referent thought: A developmental analysis of self-efficacy. In J. Falvell & L. Ross (Eds.), *Social cognitive development: Frontiers and possible futures* (pp. 200–239). Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman.

Bandura, A., & Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-motivation. *Journal of Personality and Social Psychology*. https://doi.org/10.1037/0022-3514.41.3.586

Blair, R. M., Kirkman, E. E., Maxwell, J. W., & Society, A. M. (2013). *Statistical abstract of undergraduate programs in the mathematical sciences in the United States: Fall 2010 CBMS survey*. http://www.ams.org/profession/data/cbms-survey/cbms2010-Report.pdf

Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology*, *54*(2), 199–231.

https://doi.org/10.1111/j.1464-0597.2005.00205.x

Bol, L., & Hacker, D. J. (2001). A comparison of the effects of practice tests and traditional review on performance and calibration. *Journal of Experimental Education*. https://doi.org/10.1080/00220970109600653

Bol, L., & Hacker, D. J. (2012). Calibration research: where do we go from here? *Frontiers in Psychology*, *3*(2), 229. https://doi.org/10.3389/fpsyg.2012.00229

Bol, L., Hacker, D. J., Walck, C. C., & Nunnery, J. A. (2012). The effects of individual or group guidelines on the calibration accuracy and achievement of high school biology students. *Contemporary Educational Psychology*. https://doi.org/10.1016/j.cedpsych.2012.02.004

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

Breidenbach, D., Dubinsky, E., Hawks, J., & Nichols, D. (1992). Development of the process conception of function. *Educational Studies in Mathematics: An International Journal*, *23*(3), 247–285. https://doi.org/10.1007/BF02309532

Bressoud, D. M. (2015). The Calculus Student. In D. M. Bressoud, V. Mesa, & C. L. Rasmussen (Eds.), *Insights and Recommendations from the MAA National Study of College Calculus* (pp. 1–16). Mathematical Association of America. https://doi.org/10.1227/01.NEU.0000028161.91504.4F

Bressoud, D. M., Carlson, M. P., Mesa, V., & Rasmussen, C. (2013). The calculus student: insights from the Mathematical Association of America national study. *International Journal of Mathematical Education in Science and Technology*, *44*(5), 685–698. https://doi.org/10.1080/0020739X.2013.798874

Bressoud, D. M., Mesa, V., & Rasmussen, C. L. (2015). Insights and Recommendations from the

MAA National Study of College Calculus. In D. M. Bressoud, V. Mesa, & C. L. Rasmussen (Eds.), *Insights and Recommendations from the MAA National Study of College Calculus*. https://doi.org/10.1227/01.NEU.0000028161.91504.4F

Bressoud, D. M., & Rasmussen, C. L. (2015). Seven Characteristics of Successful Calculus Programs. *Notices of the American Mathematical Society*, *62*(02), 144–146. https://doi.org/10.1090/noti1209

Broadbent, J., & Poon, W. L. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. In *Internet and Higher Education*. https://doi.org/10.1016/j.iheduc.2015.04.007

Byerley, C. (2019). Calculus students' fraction and measure schemes and implications for teaching rate of change functions conceptually. *The Journal of Mathematical Behavior*, 0–1. https://doi.org/10.1016/j.jmathb.2019.03.001

Byerley, C., Campbell, T., & Rickard, B. (2018). Evaluation of Impact of Calculus Center on Student Achievement. *Proceedings of the 21st Annual Conference on Research in Undergraduate Mathematics Education*.

Byerley, C., Moore-Russo, D., James, C., Johns, C., Rickard, B., & Mills, M. (2019). Defining the Varied Structures of Tutoring Centers: Laying a Foundation for Future Research Cameron. *Proceedings of the 22nd Annual Conference on Research in Undergraduate Mathematics Education*. https://doi.org/10.1017/CBO9781107415324.004

Byerley, C., & Thompson, P. W. (2017). Teachers' Meanings for Measure, Slope, and Rate of Change. *Journal of Mathematical Behavior*, *48*, 168–193.

Carlson, M. P. (1998). A cross-sectional investigation of the development of the function concept. *CBMS Issues in Mathematics Education*, *7*, 114–162.

https://doi.org/10.1090/cbmath/007/04

Carlson, M. P., Jacobs, S., Coe, E., Larsen, S., & Hsu, E. (2002). Applying Covariational Reasoning While Modeling Dynamic Events : A Framework and a Study. *Journal for Research in Mathematics Education*, *33*(5), 352–378.

Carlson, M. P., Madison, B., & West, R. (2010). The Calculus Concept Readiness (CCR) Instrument: Assessing student readiness for calculus. *ArXiv Preprint ArXiv:1010.2719*. http://arxiv.org/abs/1010.2719

Carlson, M. P., Madison, B., & West, R. D. (2015). A Study of Students' Readiness to Learn Calculus. *International Journal of Research in Undergraduate Mathematics Education*, *1*(2), 209–233. https://doi.org/10.1007/s40753-015-0013-y

Carlson, M. P., Oehrtman, M., & Engelke, N. (2010). The precalculus concept assessment: A tool for assessing students' reasoning abilities and understandings. *Cognition and Instruction*, *28*(2), 113–145. https://doi.org/10.1080/07370001003676587

Champion, J. (2009). Self-Efficacy in Social Cognitive Theory. In M. Zandieh (Ed.), *Proceedings of the 12th Annual Conference on Research in Undergraduate Mathematics Education*.

Champion, J., & Mesa, V. (2018). Pathways to Calculus in U.S. High Schools. *PRIMUS*. https://doi.org/10.1080/10511970.2017.1315473

Cleary, T. J., Callan, G. L., & Zimmerman, B. J. (2012). Assessing Self-Regulation as a Cyclical, Context-Specific Phenomenon: Overview and Analysis of SRL Microanalytic Protocols. *Education Research International*, *2012*(June), 1–19. https://doi.org/10.1155/2012/428639

Cobb, P., & Yackel, E. (1996). Constructivist, Emergent, and Sociocultural Perspectives in the Context of Developmental Research. *Educational Psychologist*, *31*(3–4), 175–190.

https://doi.org/10.1080/00461520.1996.9653265

Creswell, J. W., & Plano-Clark, V. L. (2011). Choosing a mixed methods design. *Designing and Conducting Mixed Method Research*.

Dabbagh, N., & Kitsantas, A. (2005). Using web-based pedagogical tools as scaffolds for self-regulated learning. *Instructional Science*, *33*(5–6), 513–540. https://doi.org/10.1007/s11251-005-1278-3

David, E. J., Roh, K. H., & Sellers, M. E. (2018). Teaching the Representations of Concepts in Calculus: the Case of the Intermediate Value Theorem. *PRIMUS: Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *October*, 1–23. https://doi.org/10.1080/10511970.2018.1540023

David, E. J., Roh, K. H., & Sellers, M. E. (2019). Value-thinking and location-thinking: Two ways students visualize points and think about graphs. *Journal of Mathematical Behavior*, *February*. https://doi.org/10.1016/j.jmathb.2018.09.004

Desoete, A., & De Craene, B. (2019). Metacognition and mathematics education: an overview. *ZDM - Mathematics Education*, *51*, 565–575. https://doi.org/10.1007/s11858-019-01060-w

Dorans, N. J. (1999). Correspondences Between ACT™ and SAT® I Scores. *ETS Research Report Series*. https://doi.org/10.1002/j.2333-8504.1999.tb01800.x

Dunning, D. (2011). The Dunning-Kruger Effect: On Being Ignorant of One's Own Ignorance. In J. M. Olson & M. P. Zanna (Eds.), *Advances in Experimental Social Psychology* (44th ed., pp. 247–296). Academic Press. https://doi.org/10.1016/B978-0-12-385522-0.00005-6

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*. https://doi.org/10.1371/journal.pone.0157447

Ferla, J., Valcke, M., & Schuyten, G. (2010). Judgments of self-perceived academic competence and their differential impact on students' achievement motivation, learning approach, and academic performance. *European Journal of Psychology of Education*, *25*(4), 519–536. https://doi.org/10.1007/sl

Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*. https://doi.org/10.1037/0003-066X.34.10.906

Frank, K. M. (2016). Plotting Points: Implications of "Over and Up" on Students' Covariational Reason. In M. B. Wood, E. E. Turner, M. Civil, & J. A. Eli (Eds.), *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education.* (pp. 573–580). The University of Arizona.

Garofalo, J., & Lester, F. (1985). Metacognition , Cognitive Monitoring , and Mathematical Performance. *Journal for Research in Mathematics Education*, *16*(3), 163–176. https://www.jstor.org/stable/748391

Hagman, J. E. (2019). The Eighth Characteristic for Successful Calculus Programs: Diversity, Equity, & Inclusion Practices. *PRIMUS*. https://doi.org/10.1080/10511970.2019.1629555

Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching*, *48*(8), 952–984. https://doi.org/10.1002/tea.20439

Honicke, T., & Broadbent, J. (2016). The influence of academic self-efficacy on academic performance: A systematic review. In *Educational Research Review*. https://doi.org/10.1016/j.edurev.2015.11.002

Ironsmith, M., Marva, J., Harju, B., & Eppler, M. (2003). Motivation and Performance in

College Students Enrolled in Self-Paced versus Lecture-Format Remedial Mathematics Courses. *Journal of Instructional Psychology*.

Jo, I. H., Yu, T., Lee, H., & Kim, Y. (2015). Relations between student online learning behavior and academic achievement in higher education: A learning analytics approach. In *Lecture Notes in Educational Technology* (pp. 275–287). https://doi.org/10.1007/978-3-662-44188-6_38

Joshua, S., Musgrave, S., Hatfield, N., & Thompson, P. W. (2015). Frames of reference. *Proceedings of the 18th Annual Conference on Research in Undergraduate Mathematics Education*, 31–44. https://doi.org/10.7208/chicago/9780226317267.003.0002

Kline, T., & Dibbs, R. (2018). *Reflection , Calibration and Achievement In Introductory Calculus*. *1*(2), 21–36.

Knuth, E. J. (2000). Student Understanding of the Cartesian Connection: An Exploratory Study. *Journal for Research in Mathematics Education*, *31*(4), 500–507. https://doi.org/10.2307/749655

Kruger, J., & Dunning, D. (1999). Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. In *Journal of Personality and Social Psychology* (Vol. 77, Issue 6).

Kutner, M., Nachtsheim, C., & Neter, J. (2004). *Applied Linear Regression Models* (4th Editio). McGraw-Hill Irwin.

Labuhn, A. S., Zimmerman, B. J., & Hasselhorn, M. (2010). Enhancing students' self-regulation and mathematics performance: The influence of feedback and self-evaluative standards. *Metacognition and Learning*, *5*(2), 173–194. https://doi.org/10.1007/s11409-010-9056-2

Larsen, S., Marrongelle, K., Bressoud, D. M., & Graham, K. (2017). Understanding the Concepts of Calculus: Frameworks and Roadmaps Emerging From Educational Research. In J. Cai (Ed.), *Compendium for research in mathematics education* (1st ed., pp. 526–550). National Council of Teachers of Mathematics.

Lauten, A. D., Graham, K., & Ferrini-Mundy, J. (1994). Student understanding of basic calculus concepts: Interaction with the graphics calculator. *Journal of Mathematical Behavior*, *13*(2), 225–237. https://doi.org/10.1016/0732-3123(94)90026-4

Lee, H. Y., Moore, K. C., & Tasova, H. I. (2019). Reasoning within quantitative frames of reference: The case of Lydia. *Journal of Mathematical Behavior*. https://doi.org/10.1016/j.jmathb.2018.06.001

Lee, J.-E., & Recker, M. (2017). Measuring students' use of self-regulated learning strategies from learning management system data: An evidence-centered design approach. In *a4Li.Sri.Com*. https://doi.org/10.13140/RG.2.2.24971.75047

Lingel, K., Lenhart, J., & Schneider, W. (2019). Metacognition in mathematics: do different metacognitive monitoring measures make a difference? *ZDM - International Journal on Mathematics Education*, *51*, 587–600. https://doi.org/10.1007/s11858-019-01062-8

Ma, J., Han, X., Yang, J., & Cheng, J. (2015). Examining the necessary condition for engagement in an online learning environment based on learning analytics approach: The role of the instructor. *Internet and Higher Education*, *24*, 26–34. https://doi.org/10.1016/j.iheduc.2014.09.005

McKenzie, K., Gow, K., & Schweitzer, R. (2004). Exploring first-year academic achievement through structural equation modelling. *Higher Education Research and Development*. https://doi.org/10.1080/0729436032000168513

Monk, G. S. (1992). Students' understanding of a function given by a physical model. In G.

Harel & E. Dubinsky (Eds.), *The concept of function: Aspects of Epistemology and pedagogy* (pp. 175–193). Mathematical Association of America.

Monk, G. S. (1994). Students' Understanding of Functions in Calculus Courses. *Humanistic Mathematics Network Journal*, *1*(9), 21–27. https://doi.org/10.5642/hmnj.199401.09.07

Moore, K. C. (2016). Graphing as figurative and operative thought. *Proceedings of the 40th Conference of the International Group for the Psychology of Mathematics Education*, *1–4*(August), 323–330.

Moore, K. C., & Carlson, M. P. (2012). Students' images of problem contexts when solving applied problems. *Journal of Mathematical Behavior*, *31*(1), 48–59. https://doi.org/10.1016/j.jmathb.2011.09.001

Moore, K. C., Paoletti, T., & Musgrave, S. (2013). Covariational reasoning and invariance among coordinate systems. *Journal of Mathematical Behavior*, *32*(3), 461–473. https://doi.org/10.1016/j.jmathb.2013.05.002

Moore, K. C., Silverman, J., Paoletti, T., Liss, D., & Musgrave, S. (2019). Conventions, habits, and U.S. teachers' meanings for graphs. *Journal of Mathematical Behavior*, *53*(December 2017), 179–195. https://doi.org/10.1016/j.jmathb.2018.08.002

Moore, K. C., Stevens, I. E., Paoletti, T., Hobson, N. L. F., & Liang, B. (2019). Pre-service teachers' figurative and operative graphing actions. *Journal of Mathematical Behavior*, *February 2018*. https://doi.org/10.1016/j.jmathb.2019.01.008

Moore, K. C., & Thompson, P. W. (2015). Shape thinking and students' graphing activity. *Proceedings of the 18th Annual Conference on Research in Undergraduate Mathematics Education*, 782–789.

Morris, L. V., Finnegan, C., & Wu, S. S. (2005). Tracking student behavior, persistence, and achievement in online courses. *Internet and Higher Education*, *8*(3), 221–231. https://doi.org/10.1016/j.iheduc.2005.06.009

Murray, S. B. (1931). On Certain Deficiencies of College Freshmen in Algebra. *Mathematics News Letter*. https://doi.org/10.2307/3027877

National Council Of Teachers Of Mathematics. (2000). Principles and Standards for School Mathematics. *School Science and Mathematics*. https://doi.org/10.1111/j.1949-8594.2001.tb17957.x

Nemirovsky, R., & Rubin, A. (1992). *Students' Tendency To Assume Resemblances between a Function and Its Derivative*.

Oehrtman, M., Carlson, M. P., & Thompson, P. W. (2008). Foundational Reasoning Abilities that Promote Coherence in Students' Function Understanding. In M. P. Carlson & C. Rasmussen (Eds.), *Making the Connection: Research and Teaching in Undergraduate Mathematics Education* (pp. 27–42). Mathematical Assoication of America. https://doi.org/10.5948/UPO9780883859759.004

Orton, A. (1983). Students' Understanding of Differentiation. *Educational Studies in Mathematics*, *14*, 235–250.

Pajares, F., & Graham, L. (1999). Self-Efficacy, Motivation Constructs, and Mathematics Performance of Entering Middle School Students. *Contemporary Educational Psychology*, *24*(2), 124–139. https://doi.org/10.1006/ceps.1998.0991

Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematical problem-solving. *Contemporary Educational Psychology*. https://doi.org/10.1006/ceps.1995.1029

Pajares, F., & Miller, M. D. (1994). *Role of Self-Efficacy and Self-Concept Beliefs in*

Mathematical Problem Solving: A Path Analysis. *86*(2), 193–203.

Pardo, A., Han, F., & Ellis, R. A. (2016). Exploring the relation between self-regulation, online activities, and academic performance. *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge - LAK '16*, 422–429. https://doi.org/10.1145/2883851.2883883

Pilgrim, M. E. (2010). *A Concepts for Calculus intervention: Measuring students attitudes toward mathematics and achievement in calculus*. (Publication No. 3419112) [Doctoral dissertation, Colorado State University]. ProQuest Dissertation Publishing.

Pilgrim, M. E., Burks, L., Hill-Lindsay, S., & Ryals, M. (2020). Links between Engagement in Self-Regulation and Performance. *Proceedings of the 23rd Annual Conference on Research in Undergraduate Mathematics Educationo*. https://doi.org/10.1017/CBO9781107415324.004

Pilgrim, M. E., & Gehrtz, J. (2016). An Alternative Calculus I Course. *Mathematics and Computer Education*, *50*(2), 120–129. https://search.proquest.com/docview/1794881645?accountid=14537

Pintrich, P. R. (2000). The Role of Goal Orientation in Self-Regulated Learning. In *Handbook of Self-Regulation* (pp. 451–502). Academic Press. https://doi.org/10.1016/b978-012109890-2/50043-3

Pintrich, P. R. (2002). The role of metacognitive knowledge in learning, teaching, and assessing. *Theory into Practice*, *41*(4), 219–225. https://doi.org/10.1207/s15430421tip4104_3

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(4), 385–407. https://doi.org/10.1007/s10648-004-0006-x

Pintrich, P. R., & De Groot, E. V. (1990). Motivational and Self-Regulated Learning Components of Classroom Academic Performance. *Journal of Educational Psychology*. https://doi.org/10.1037/0022-0663.82.1.33

Pintrich, P. R., Smith, D. A. F., Garcia, T., & Mckeachie, W. J. (1993). Reliability and Predictive Validity of the Motivated Strategies for Learning Questionnaire (Mslq). *Educational and Psychological Measurement*. https://doi.org/10.1177/0013164493053003024

Pintrich, P. R., Smith, D., Garcia, T., & McKeachie, W. (1991). A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ). *Ann Arbor. Michigan*. https://doi.org/ED338122

President's Council of Advisors on Science and Technology (PCAST). (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics*. http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-v11.pdf

Ramdass, D., & Zimmerman, B. J. (2008). Effects of Self-Correction Strategy Training on Middle School Students' Self-Efficacy, Self-Evaluation, and Mathematics Division Learning. *Journal of Advanced Academics*. https://doi.org/10.4219/jaa-2008-869

Reinholz, D. L. (2009). *An analysis of factors affecting student success in Math 160 Calculus for Physical Scientists I.* [Unpublished Master's thesis]. Colorado State University. Retreived from https://mountainscholar.org/handle/10217/20937.

Rickard, B., & Mills, M. (2018). The effect of attending tutoring on course grades in Calculus I. *International Journal of Mathematical Education in Science and Technology*. https://doi.org/10.1080/0020739X.2017.1367043

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science*, *17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Sadler, P., & Sonnert, G. (2018). The path to college calculus: The impact of high school mathematics coursework. *Journal for Research in Mathematics Education*, *49*(3), 292–329.

Schoenfeld, A. H. (1992). Learning to Think Mathematically: Probelm Solving, Metacognition, and Sense Making in Mathematics. In *Handbook for Research on Mathematics Teaching and Learning*.

Schraw, G. (1998). Promoting general metacognitive awareness. *Instructional Science*. https://doi.org/10.1007/978-94-017-2243-8_1

Schraw, G., Potenza, M. T., & Nebelsick-Gullet, L. (1993). Constraints on the calibration of performance. *Contemporary Educational Psychology*. https://doi.org/10.1006/ceps.1993.1034

Schunk, D. H., & Pajares, F. (2004). Self-efficacy in education revisited: Empirical and applied evidence. In D. M. MacInerney & S. Van Etten (Eds.), *Big Theories Revisited* (pp. 115–138). Information Age.

Seymour, E., & Hewitt, N. (1997). *Talking about leaving*. Westview Press.

Seymour, E., Hunter, A.-B., & Weston, T. J. (2019). Why We Are Still Talking About Leaving. In *Talking about Leaving Revisited*. https://doi.org/10.1007/978-3-030-25304-2_1

Sheather, S. (2009). *A modern approach to regression with R*. Springer.

Sheldrake, R., Mujtaba, T., & Reiss, M. J. (2014). Calibration of self-evaluations of mathematical ability for students in England aged 13 and 15, and their intentions to study non-compulsory mathematics after age 16. *International Journal of Educational Research*, *64*, 49–61. https://doi.org/10.1016/j.ijer.2013.10.008

Shilo, A., & Kramarski, B. (2019). Mathematical-metacognitive discourse: how can it be developed among teachers and their students? Empirical evidence from a videotaped lesson and two case studies. *ZDM - Mathematics Education*. https://doi.org/10.1007/s11858-018-01016-6

Skemp, R. (1976). Relational Understanding and Instrumental Understanding. *Mathematics Teaching*, *77*(1), 20–26.

Smith, W., Webb, D., Bowers, J., Voigt, M., & Team, T. S. (2017). SEMINAL: Preliminary Findings on Institutional Changes in Departments of Mathematics. In W. Smith, B. R. Lawler, J. Bowers, & L. Augustyn (Eds.), *Proceedings of the sixth annual Mathematics Teacher Education Partnership conference* (pp. 121–128). Association of Public and Land-grant Univeristies.

Sonnert, G., & Sadler, P. (2014). The impact of taking a college pre-calculus course on students' college calculus performance. *International Journal of Mathematical Education in Science and Technology*, *45*(8), 1188–1207. https://doi.org/10.1080/0020739X.2014.920532

Steen, L. A. (1987). Calculus for a new century: A pump, not a filter. Papers presented at a colloquium. *Mathematical Association of America*.

Steffe, L. P., & Thompson, P. W. (2000). Steffe, L. P., & Thompson, P. W. (2000). Teaching experiment methodology: Underlying principles and essential elements. *Research Design in Mathematics and Science Education*, 267–307.

Stump, S. L. (2001). Developing preservice teachers' pedagogical content knowledge of slope. *Journal of Mathematical Behavior*. https://doi.org/10.1016/S0732-3123(01)00071-2

Thanheiser, E., Philipp, R. A., & Fasteen, J. (2014). Motivating Prospective Elementary

Teachers To Learn Mathematics Via Authentic Tasks. *Proceedings of the 38th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, *5*, 233–240.

Thanheiser, E., Philipp, R. A., Fasteen, J., Strand, K., & Mills, B. (2013). Preservice-Teacher Interviews: A Tool for Motivating Mathematics Learning. *Mathematics Teacher Educator*. https://doi.org/10.5951/mathteaceduc.1.2.0137

Thompson, P. W. (1993). Quantitative reasoning, complexity, and additive structures. *Educational Studies in Mathematics*, *25*(3), 165–208. https://doi.org/10.1007/BF01273861

Thompson, P. W. (1994). Images of rate and operational understanding of the fundamental theorem of calculus. *Educational Studies in Mathematics*, *2–3*(26), 229–274.

Thompson, P. W. (2008). Conceptual Analysis of Mathematical Ideas: Some Spadework at the Foundation of Mathematics Education. *Annual Meeting of the International Group for the Psychology of Mathematics Education*.

Thompson, P. W. (2011). Quantitative reasoning and mathematical modeling. In L. L. Hatfield, S. Chamberlain, & S. Belbase (Eds.), *New perspectives and directions for collaborative research in mathematics education.* (WISDOMe Mo, Vol. 1, pp. 33–57). University of Wyoming.

Thompson, P. W., & Ashbrook, M. (2019). *Calculus: Newton, Leibniz, and Robinson Meet Technology*. http://patthompson.net/ThompsonCalc/

Thompson, P. W., & Carlson, M. P. (2017). Variation, covariation, and functions: Foundational ways of thinking mathematically. In J. Cai (Ed.), *Compendium for research in mathematics education* (1st ed., pp. 421–456). National Council of Teachers of Mathematics.

Thompson, P. W., Carlson, M. P., Byerley, C., & Hatfield, N. (2014). Schemes for thinking with magnitudes: A hypothesis about foundational reasoning abilities in algebra. In *Epistemic algebra students: Emerging models of students' algebraic knowing* (pp. 1–24).

Thompson, P. W., & Milner, F. (2018). *Teachers' Meanings for Function and Function Notation in South Korea and the United States*. 55–66. https://doi.org/10.1007/978-3-319-99386-7_4

Tian, Y., Fang, Y., Li, J., Wang, L., & Greenshaw, A. J. (2018). The effect of metacognitive knowledge on mathematics performance in self-regulated learning framework-multiple mediation of self-efficacy and motivation. *Frontiers in Psychology*, *9*(December), 1–11. https://doi.org/10.3389/fpsyg.2018.02518

Van Vliet, F., & Mirin, A. (2020). f(x) Means y: Students' Meanings for Function Notation. *Proceedings of the 23rd Annual Conference on Research in Undergraduate Mathematics Educationo*.

von Glasersfeld, E. (1995). *Radical constructivism: A way of knowing and learning.* Falmer Press. https://doi.org/10.4324/9780203454220

Wakefield, N., Champion, J., Bolkema, J., & Dailey, D. (2018). Diagnostic Effects of an Early Mastery Activity in College Algebra and Precalculus. *International Journal of Research in Undergraduate Mathematics Education*, 1–17. https://doi.org/10.1007/s40753-018-0075-8

White, P., & Mitchelmore, M. (1996). Conceptual Knowledge in Introductory Calculus. *Journal for Research in Mathematics Education*, *27*(1), 79. https://doi.org/10.2307/749199

Winne, P. H., & Baker, R. S. (2013). The potentials of educational data mining for researching metacognition, motivation and self-regulated learning. *Journal of Educational Data Mining*, *5*(1), 1–8. https://doi.org/10.1037/1082-989X.2.2.131

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlosky, & A. Graesser (Eds.), *Metacognition in educational theory and practice* (pp.

277–304). Lawrence Erlbaum.

Winne, P. H., & Jamieson-Noel, D. (2002). Exploring students' calibration of self reports about study tactics and achievement. *Contemporary Educational Psychology*. https://doi.org/10.1016/S0361-476X(02)00006-1

Winne, P. H., & Perry, N. E. (2000). Measuring Self-Regulated Learning. In *Handbook of Self-Regulation*. https://doi.org/10.1016/b978-012109890-2/50045-7

Worthley, M. R. (2013). *A mixed methods explanatory study of the failure/drop rate for freshman STEM calculus students*. (Publication No. 3565581) [Doctoral dissertation, Colorado State University]. ProQuest Dissertation Publishing.

Worthley, M. R., Gloeckner, G. W., & Kennedy, P. A. (2016). A Mixed-Methods Explanatory Study of the Failure Rate for Freshman STEM Calculus Students. *PRIMUS: Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *25*(2), 125–142. https://doi.org/10.1080/10511970.2015.1067265

Yackel, E., & Cobb, P. (1996). Sociomathematical Norms, Argumentation, and Autonomy in Mathematics. In *Source: Journal for Research in Mathematics Education* (Vol. 27, Issue 4). https://about.jstor.org/terms

Yackel, E., Rasmussen, C., & King, K. (2000). Social and sociomathematical norms in an advanced undergraduate mathematics course. *Journal of Mathematical Behavior*, *19*(3), 275–287. https://doi.org/10.1016/s0732-3123(00)00051-1

Zandieh, M. (2000). A Theoretical Framework for Analyzing Student Understanding of the Concept of Derivative. *Research in Collegiate Mathematics Education IV*, *8*, 103–127.

Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, *22*(6), 413–419.

https://doi.org/10.1016/j.learninstruc.2012.03.004

Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13–39). Academic Press. https://doi.org/10.1016/B978-012109890-2/50031-7

Zimmerman, B. J. (2002). *Becoming a Self-Regulated Learner : An Overview*. *41*(2), 64–70. https://doi.org/10.1207/s15430421tip4102

Zimmerman, B. J., & Kitsantas, A. (2014). Comparing students' self-discipline and self-regulation measures and their prediction of academic achievement. *Contemporary Educational Psychology*, *39*(2), 145–155. https://doi.org/10.1016/j.cedpsych.2014.03.004

Zimmerman, B. J., & Martinez-Pons, M. (1986). Development of a structured interview for assessing student use of self-regulated learning strategies. *American Educational Research Journal*, *23*(4), 614–628. https://doi.org/10.3102/00028312023004614

Zimmerman, B. J., Moylan, A., Hudesman, J., White, N., & Flugman, B. (2011). Enhancing self-reflection and mathematics achievement of at-risk urban technical college students. *Psychological Test and Assessment Modeling*, *53*(1), 108–127.

Zimmerman, B. J., & Schunk, D. H. (2001). *Self-regulated learning and academic achievement: Theoretical perspectives* (2nd ed.). Routledge.

Appendix A

SRL Score Coding Scheme

Coding for Each Data Source				Corresponding SRL Score
PSA	PCQ	PRT	PHR	
High Confidence	High Performance	Did Use	Did Use	5
High Confidence	High Performance	Did Use	Did Not Use	4
High Confidence	High Performance	Did Not Use	Did Use	3
High Confidence	High Performance	Did Not Use	Did Not Use	4
High Confidence	Low Performance	Did Use	Did Use	4
High Confidence	Low Performance	Did Use	Did Not Use	2
High Confidence	Low Performance	Did Not Use	Did Use	3
High Confidence	Low Performance	Did Not Use	Did Not Use	2
High Confidence	Did Not Use	Did Use	Did Use	3
High Confidence	Did Not Use	Did Use	Did Not Use	1
High Confidence	Did Not Use	Did Not Use	Did Use	3

High Confidence	Did Not Use	Did Not Use	Did Not Use	1
Low Confidence	High Performance	Did Use	Did Use	4
Low Confidence	High Performance	Did Use	Did Not Use	3
Low Confidence	High Performance	Did Not Use	Did Use	3
Low Confidence	High Performance	Did Not Use	Did Not Use	2
Low Confidence	Low Performance	Did Use	Did Use	N/A
Low Confidence	Low Performance	Did Use	Did Not Use	2
Low Confidence	Low Performance	Did Not Use	Did Use	3
Low Confidence	Low Performance	Did Not Use	Did Not Use	1
Low Confidence	Did Not Use	Did Use	Did Use	3
Low Confidence	Did Not Use	Did Use	Did Not Use	2
Low Confidence	Did Not Use	Did Not Use	Did Use	N/A
Low Confidence	Did Not Use	Did Not Use	Did Not Use	N/A
Did Not Use	High Performance	Did Use	Did Use	N/A
Did Not Use	High Performance	Did Use	Did Not Use	N/A
Did Not Use	High Performance	Did Not Use	Did Use	N/A

Did Not Use	High Performance	Did Not Use	Did Not Use	N/A
Did Not Use	Low Performance	Did Use	Did Use	3
Did Not Use	Low Performance	Did Use	Did Not Use	1
Did Not Use	Low Performance	Did Not Use	Did Use	N/A
Did Not Use	Low Performance	Did Not Use	Did Not Use	N/A
Did Not Use	Did Not Use	Did Use	Did Use	2
Did Not Use	Did Not Use	Did Use	Did Not Use	1
Did Not Use	Did Not Use	Did Not Use	Did Use	1
Did Not Use	Did Not Use	Did Not Use	Did Not Use	0

Appendix B

Coded Relationships between Precalculus Self-Assessment Items and Precalculus Content Quiz Items

		PSA Items															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	2	1
	2	1	1	1	0	0	0	0	0	0	0	0	0	0	0	1	2
	3	0	0	1	2	2	0	0	0	0	0	0	0	1	0	0	0
	4	0	0	0	0	1	2	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0
PCQ Items	6	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0
	7	2	2	2	2	2	0	0	0	0	0	0	0	2	0	0	0
	8	0	0	0	0	0	0	0	2	0	0	2	0	0	0	0	0
	9	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	0
	10	0	0	1	0	0	0	0	2	2	0	2	0	0	0	0	0
	11	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	12	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0

Note. Coded relationships between PSA and PCQ items are coded as either ‘0’, ‘1’, or ‘2’, where ‘0’ represents no relationship between the items, ‘1’ represents a relationship between the items, and ‘2’ represents a strong relationship between the items.

Appendix C

Data Regarding Resource Usage and Population Sample

Participants in this study included students enrolled in a predominantly Caucasian western university's Calculus I course in the fall semester of 2016. Of the 426 students enrolled in the course, 401 students consented to all aspects of the study. As many students formally drop the course or informally disengage from, this study intended to only investigate students who were engaged throughout the entire course. The data set for this study hence used only students who completed all exams in the course, leaving 356 students. Further this study intended to study the calibration of these students, so only students who completed the PSA and the PCQ were included in the sample. Of the 356 students, 77.8% and 68% of the students completed the PSA and PCQ, respectively, of which 218 students completed both (see Table C.1).

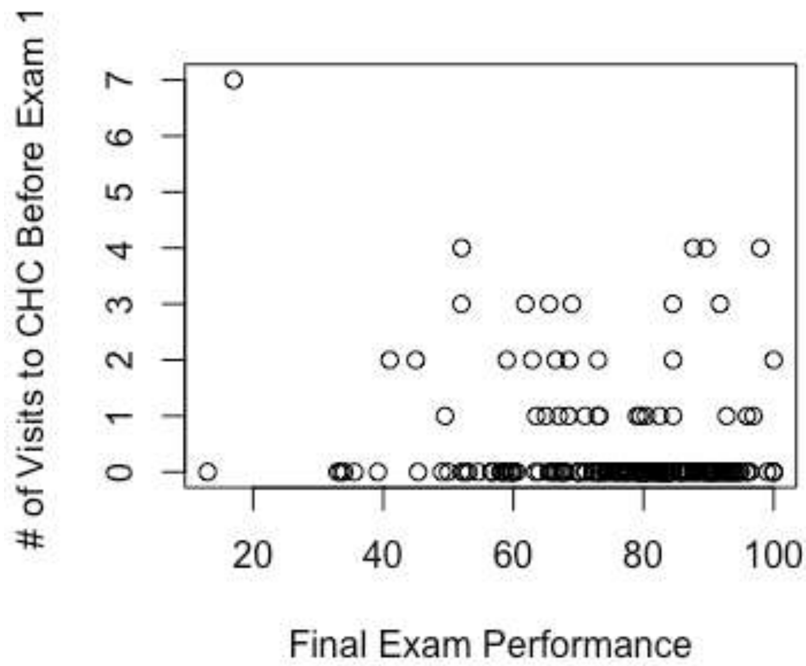
Table C.1: Number of consenting students by PSA and PCQ usage (percentage of entire population)

		PSA		
		No use	Did use	Total
PCQ	No use	55	59	114 (32%)
	Did use	24	218	242 (68%)
	Total	79 (22.2%)	277 (77.8%)	356

In addition to the calibration bias, calibration accuracy, and SAT/ACT scores, students' attendance to a Calculus Help Center (CHC) was recorded. Office hours of instructors and teaching assistants were held in the CHC. Students could freely visit the CHC during the workday for scheduled appointments with their instructor or for drop-in appointments with an instructor of a different section of the course. It is worth note that Calculus I was the lowest level course that the CHC served, so any instructor could one working in the CHC could help Calculus I students with whatever question they may have.

Including only students whom have consented to the study, completed all exams, choose to complete the PSA and the PCQ, and whose SAT Math or ACT Math scores were available, there are a total of 196 students in the sample. Two students were further removed from the data set from being outliers with regards to CHC visits before Exam 1and Final Exam performance, giving 194 students to be used for this research study.

Including only students whom have consented to the study, completed all exams, choose to complete the PSA and the PCQ, and whose SAT Math or ACT Math scores were available, there are a total of 196 students in the sample. Two students were further removed from the data set from being outliers with regards to CHC visits before the first exam or final exam performance (See Figure C.1), giving 194 students to be used for this research study.



Note. The two left-most points were excluded from the analysis as they were outliers.

Figure C.1: Scatterplot of Student Final Exams and Number of Visits to CHC Before Exam 1