Title
Improving Student Success in Calculus : : A Comparison of Four College Calculus Classes

Author
Bagley, Spencer Franklin

Publication Date
2014

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

SAN DIEGO STATE UNIVERSITY

Improving Student Success in Calculus: A Comparison of Four College Calculus Classes

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in

Mathematics and Science Education

by

Spencer Franklin Bagley

Committee in charge:

University of California, San Diego

John Eggers
Jeffrey Rabin
Elizabeth Simon

San Diego State University

Chris Rasmussen, Chair
Richard Levine

2014

The Dissertation of Spencer Franklin Bagley is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

San Diego State University

2014

DEDICATION

To my parents, Larene and Steve,

for their love, support, and encouragement

over far too many years of schooling.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

It takes a village to write a dissertation. I am indebted to many people for their help in the completion of this document, and wish to acknowledge them here.

Academically, my first thanks go to my committee. My advisor Chris Rasmussen always pushed me to write more, even when I thought I had said everything that I could, and this improved my text substantially. He sat through numerous meetings with me during which I threw noodles at the wall to see which ones stuck, and helped me clean up the ones that bounced off. Rich Levine's statistical expertise was invaluable, and he was always willing to answer questions, no matter how trivial. Jeff Rabin read this document and my proposal with a fine-toothed comb and caught many small errors to which I was blind because I wrote them. Beth Simon and John Eggers provided many helpful big-picture comments helping me understand what my work meant in the real world.

I also thank the other members of CRMSE for all they have taught me. I'm particularly grateful for Joanne Lobato, who has helped me become a much better writer, and for Lisa Lamb, for her gracious and sympathetic ear. I'm indebted to the four instructors in this study for allowing me unfettered access to their classes to collect data. The SDSU President's Leadership Fund provided financial support for this dissertation.

More personally, I thank my parents, Larene and Steve Bagley, for everything they have done for me; I could write another two hundred pages about them. My fellow graduate students, Megan Wawro, Dov Zazkis, Molly Kelton, Casey Hawthorne, John Gruver, David Quarfoot, C. David Walters, Hayley Miles-Leighton, Naneh Apkarian, and Raymond LaRochelle, and especially my cohort, Jess Ellis, Brooke Ernest Downey, Mike

Fredenberg, Natalie Selinski, and Bridget Druken, have attended a hundred of my talks, given me advice, explained theory, laughed and complained with me, and in a thousand other ways small and large made my graduate school experience better. My thanks also go to Dr. Martin Doucett for helping me find my way and myself through the hundred crises attendant to dissertation writing (and everything else). Finally, I thank my friends Lindsay Nielson, Matt and Sofia Krause, Caitlin Stallings, Cy Credille, and Kamrie Littlefield, for always being there to listen, and for putting up with me during all this.

VITA

Degrees Awarded

2006 Bachelor of Science, Mathematics, University of Utah

2010 Master of Science, Mathematics, University of Utah

2014 Doctor of Philosophy, Mathematics and Science Education,
University of California, San Diego and San Diego State University

Academic Appointments and Teaching Experience

2014 – 2015: Visiting Assistant Professor, Department of Mathematical Sciences,
University of Northern Colorado

2013 – 2014: Instructor, Department of Mathematics,
San Diego State University

2010 – 2013: Graduate Research Assistant,
University of California, San Diego, and San Diego State University

2008 – 2010: Tutor, Department of Mathematics,
University of Utah

Refereed Publications Under Review

Bagley, S., Rasmussen, C., Zandieh, M. (under review.) Inverse, composition, and identity: The case of function and linear transformation.

Selected Articles in Refereed Conference Proceedings

Bagley, S., & Rabin, J. (2013). Computational thinking in linear algebra. In (Eds.) S. Brown, G. Karakok, K. H. Roh, and M. Oehrtman, *Proceedings of the 16th Annual Conference on Research in Undergraduate Mathematics Education,* 2013, Denver, Colorado. Vol. 1, 53-64.

ABSTRACT OF THE DISSERTATION

Improving Student Success in Calculus: A Comparison of Four College Calculus Classes

by

Spencer Franklin Bagley

Doctor of Philosophy in Mathematics and Science Education

University of California, San Diego, 2014

San Diego State University, 2014

Professor Chris Rasmussen, Chair

The quality of education in science, technology, engineering, and mathematics (STEM) fields is an issue of particular educational and economic importance, and Calculus I is a linchpin course in STEM major tracks. A national study is currently being conducted examining the characteristics of successful programs in college calculus (CSPCC, 2012). In work related to the CSPCC program, this study examines the effects on student outcomes of four different teaching strategies used at a single institution. The four classes were a traditional lecture, a lecture with discussion, a lecture incorporating both discussion and technology, and an inverted model.

This dissertation was guided by three questions: (1) What impact do these four instructional approaches have on students' persistence, beliefs about mathematics, and conceptual and procedural achievement in calculus? (2) How do students at the local institution compare to students in the national database? And (3) How do the similarities and differences in opportunities for learning presented in the four classes contribute to the similarities and differences in student outcomes?

Quantitative analysis of surveys and exams revealed few statistically significant differences in outcomes, and students in the inverted classroom often had poorer outcomes than those in other classes. Students in the technology-enhanced class scored higher on conceptual items on the final exam than those in other classes. Comparing to the national database, local students had similar switching rates but less expert-like attitudes and beliefs about mathematics than the national average.

Qualitative analysis of focus group interviews, classroom observations, and student course evaluations showed that several implementation issues, some the result of pragmatic constraints, others the result of design choice, weakened affordances provided by innovative features and shrunk the differences between classes. There were substantial differences between the inverted classroom in this study and successful implementations in the literature. I identified a set of departures that forms a list of best practices for inverting classrooms. Students in all classes felt that prior calculus experience was a prerequisite for their current calculus class, and that class sessions felt rushed. These concerns implicate the constraints imposed by the curriculum shared by the four classes.

Chapter 1: Introduction

Calculus I is a course of great importance in college education. It is a required course in the college careers of many students, whether as a foundation for a math major or as a service course providing them with the mathematical tools necessary to succeed in another discipline. Given its broad importance and utility, Calculus I should be taught well; unfortunately, many students experience their calculus classes as uninspiring, dull, or unproductive, and as many as a quarter of the students in any given calculus class will not achieve a passing grade (Bressoud, Carlson, Pearson, & Rasmussen, 2012). Therefore, the mathematics education community is obligated to find and document productive approaches to calculus, then disseminate these approaches for use across the nation. Only in this way can all students obtain from their calculus classes the tools, skills, and attitudes they need to succeed in their education and careers.

This chapter introduces my study, which examines four different pedagogical approaches to teaching calculus and compares their effects on student outcomes, including conceptual and procedural achievement, persistence in STEM major tracks, and attitudes and beliefs about mathematics. I discuss several reasons why calculus education is so important, ranging from questions of national economic standing to individual student outcomes. I explain the relationship of my study to a national MAA study characterizing and developing deeper understanding of successful institutions. I explain why studying the teaching of college calculus resonates personally with me. Finally, I conclude by presenting the research questions that drive this study.

1.1. Why calculus teaching matters

The quality of education in science, technology, engineering, and mathematics (STEM) fields is an issue of particular educational and economic importance. The President's Council of Advisors on Science and Technology (PCAST, 2012) reported on current economic forecasts which indicate that if the United States is to maintain its position of global leadership in science and technology, U.S. universities must produce over the next decade approximately one million more graduates in the STEM majors than currently anticipated. However, PCAST (2012) reported that fewer than 40% of students who originally intend to major in a STEM field actually complete a STEM degree. Staunching the flow of potential majors out of STEM degree tracks is thus a viable target for improvement; merely increasing the retention of STEM majors from 40% to 50% would generate 750,000 more STEM degrees in the next decade.

One clear way to proceed is to examine the reasons students give for leaving STEM majors, and then attempt to address the issues thus uncovered. PCAST (2012) found that many of the reasons students give for abandoning a STEM degree point at uninspiring, unwelcoming, or poorly-taught introductory courses. Calculus I was among the courses most often cited; it is an important "gateway" course required by every STEM major, commonly the first mathematics course taken by incoming freshmen, and a tone-setter for future classes in mathematics and other STEM fields. Unfortunately, such classes are "frequently uninspiring, relying on memorization and rote learning while avoiding richer mathematical ideas," leaving many students with a picture of mathematics and other STEM fields as "dull and unimaginative" (p. 28).

PCAST's (2012) findings on why students leave STEM degrees echoed those reported by Seymour (2006). In exit interviews she conducted with both those who switched out of STEM major tracks and graduating seniors in STEM tracks, poor learning experiences were the most common complaint. Students' concerns included the lack of discussion of conceptual material, faculty's implicit or explicit dislike for teaching, courses that attempted to cover too much material too fast, and becoming bored with introductory courses even when incoming interest was strong.

It thus appears that productive solutions to the problem of STEM major retention may lie in the direction of improving the teaching of Calculus I, as well as identifying and documenting good teaching practices already in place. As one recommendation to address the national dearth of STEM majors, PCAST (2012) urged researchers to "launch a national experiment in postsecondary mathematics education to address the mathematics-preparation gap" (p. 27). In research closely linked to this recommendation, a team of researchers conducted a national study, sponsored by the NSF and under the aegis of the MAA, entitled Characteristics of Successful Programs in College Calculus (CSPCC, 2012). The MAA study surveyed 160 institutions across the United States, aiming to improve understanding of the demographic makeup of the body of students in calculus, to measure the impact of the various characteristics of calculus classes that are believed to influence student success, and to identify particularly successful programs. These programs were then made the focus of case studies to determine what institutional factors contribute to their success.

My study is an offshoot of this larger study being conducted by the MAA. However, rather than taking a broad look at programs across the United States, with all the variables such a large study entails, I focused on four classes with different pedagogical approaches at a single institution. I then compared the single-institution data to the broad, cross-institution data to examine how the classes in my study compare to successful classes elsewhere.

1.2. Personal interest in the study of calculus teaching

The quality of calculus instruction is one of the main forces driving my interest in mathematics education. I first became interested in mathematics education in community college, where I had an excellent calculus instructor. I became a mathematics major with a long-term plan of pursuing a Ph.D. in pure mathematics, because this seemed to be the path that led to teaching at the university level. As I went through my undergraduate studies, I commonly overheard other students complaining about mathematics and their instructors. This served to increase my desire to be a college math teacher; I wanted to share my passion for mathematics with students and help improve the state of lower-division undergraduate mathematics teaching.

It was this desire to be a excellent mathematics teacher that eventually drove me into the field of mathematics education. During my master's work in pure mathematics, I found myself wishing that I could study in a Ph.D. program that was more about pedagogy than math content. A colleague pointed me in the direction of mathematics education programs, and I was accepted to one several months later.

My interests in undergraduate mathematics education are broad, covering everything from proof to group theory to linear algebra. However, it has always been a core part of my motivation to improve teaching in lower-division classes, and particularly calculus, because they are the service courses that are taken, and largely loathed, by a wide swath of university students. Therefore, I felt that this was the level at which there was the most room, and the most need, for improvement. When the opportunity came for me to focus my dissertation research on improving calculus teaching, I was excited to make an impact on this important field.

1.3. The study and research questions

In the Fall 2012 semester, at a large public university in the southwestern United States, Calculus I was taught by four different instructors using four different instructional techniques. As will be detailed in chapter 3 and chapter 6, these instructional approaches differed on a number of axes, including interactivity, the use of technology, and the use of traditional lecture; these differences contributed to differences in opportunities for learning presented to the students in each class. I studied the effect of these differences in opportunities for learning on student outcomes, and provided explanatory, conceptual links between the similarities and differences in the instructional approaches and the similarities and differences in student outcomes. This study was driven by the following research questions:

1) What impact do the four different instructional approaches have on students':
    a) persistence in STEM major tracks?
    b) attitudes, dispositions, and beliefs about mathematics?
    c) conceptual and procedural achievement in calculus?
2) How do students at the local institution compare to students in the national database in their:

   a) persistence in STEM major tracks?

   b) attitudes, dispositions, and beliefs about mathematics?

3) How do the similarities and differences in opportunities for learning between the four classes contribute to the similarities and differences in outcomes?

In chapter 2, I review four bodies of literature that are pertinent to this study. In chapter 3, I discuss the theoretical perspective and methodology I used to answer these research questions. In chapter 4, I present my analysis of Research Question 1, which focused on quantitative results obtained by comparing the various outcome measures across the four classes. In chapter 5, I present my analysis of Research Question 2, focusing on quantitative comparisons between data from the local institution and data from the national sample. In chapter 6, I present qualitative data from student interviews and classroom observations, answering Research Question 3.

Chapter 2: Literature

I have identified four particularly relevant bodies of literature: (a) literature providing a rationale for studying college calculus; (b) literature on the role of affect and beliefs; (c) literature on the effect of the inverted classroom strategy; and (d) literature on the roles of discourse and technology in creating opportunities for learning. In this section I summarize salient articles I have examined in each of these categories, discuss the implications of the extant literature for my study, and outline the projected contribution of my study to the literature.

2.1. Rationale

As mentioned in the introduction, the report of the President's Council of Advisors on Science and Technology (PCAST, 2012) discussed the economic rationale for finding ways to improve the teaching of introductory STEM courses in general, and Calculus I in particular. Economic forecasts project an increasing proportion of STEM occupations, as well as other jobs relying on STEM disciplinary knowledge, including positions in nursing and skilled manufacturing. In order to supply enough qualified candidates for these positions, U.S. universities must produce, over the next decade, on the order of one million more STEM graduates than predicted by current graduation rates.

Searching for ways to bolster the numbers of STEM graduates, PCAST (2012) found that the pipeline of STEM majors is leaking students at an alarming rate; less than 40% of students who enter college intending a STEM major actually complete a degree in a STEM field. The students who leave cite varied reasons: some do not achieve the grades necessary to continue, despite high levels of interest and aptitude, indicating that they might benefit from better teaching or more institutional support; others, who do achieve the necessary grades, "describe the teaching methods and atmosphere in introductory STEM classes as ineffective and uninspiring" (p. 5). Many of these reasons for leaving implicate the need for better teaching of introductory courses, including and especially Calculus I, commonly considered a "gateway" course to many other STEM fields (p. 27).

In testimony before Congress, Seymour (2006; Seymour & Hewitt, 1997) also reported on reasons students give for choosing to switch out of STEM disciplines. Her findings parallel those reported above: "reports of poor learning experiences were by far the most common complaint *both* of those who had switched out of [STEM] majors *and* graduating seniors in those majors" (p. 3; emphasis in original). Students in her study implicated "over-stuffed" courses taught too quickly, unavailable and disinterested faculty who seemingly took little responsibility for student learning, and a dearth of "application, illustration, or discussion of conceptual material" as factors contributing to their dissatisfaction with the teaching of their courses (p. 4).

As to the causes of these poor learning experiences, Seymour (2006) pointed out that "the balance of status and rewards has, over time, tipped heavily towards research and away from teaching" (p. 2). This tilting has driven many faculty away from many of the interactive, and thus time-intensive, teaching functions, such as tutorials, seminars, and individual mentoring and advising, that had previously allowed students valuable face-to-face time with their professors. Therefore, "straight lecturing," a less-effective pedagogical strategy, has largely become faculty's dominant mode of contact with students (p. 2). Many of the interactive functions described above have recently become part of the job description of graduate teaching assistants. These TAs are largely untrained for these functions, and the overall quality of undergraduate education has suffered as a result. Seymour thus recommended that institutions place greater value on teaching, and implement systems of professional development for both TAs and faculty.

Contributing to the field's understanding of student persistence, Tinto (1975, 1997; Pascarella & Terenzini, 1980) developed a theoretical model of student dropout from higher education modeled on Durkheim's theory of suicide. He distinguished involuntary dropout resulting from academic failure from voluntary withdrawal resulting from other factors, and transfer or temporary dropout behaviors from permanent dropout. His model suggested that students' levels of goal and institutional commitment were affected by their experiences in both the academic and social systems of the institution, and that these levels of commitment were the most direct predictors of students' dropout decisions. For instance, voluntary withdrawal may be predicted by low initial levels of goal commitment coupled with insufficient social support, regardless of a student's level of academic performance. The academic and social components of this model are seen as nested spheres of activity, where academic activity occurs within the broader social sphere. Tinto (1997) concluded that "choices of curriculum structure … and pedagogy invariably shape both learning and persistence on campus," due to their effect on students' degree and manner of involvement in academic and social activities (p. 620). Persistence is thus fundamentally and inextricably linked to educational practice.

PCAST (2012) provided four overarching recommendations on addressing these questions of poor retention and poor teaching; I will focus here on the two that are most relevant to my study. The first of these is to "catalyze widespread adoption of empirically validated teaching practices" (p. 16). There is substantial empirical literature providing evidence in favor of progressive models of instruction; however, many introductory STEM courses are still taught using traditional methods "dominated by lectures and multiple choice tests" (p. 16). Pedagogical strategies that engage students in active learning result in better outcomes, both in achievement and in affect, and thus address many of the reasons students give for leaving STEM tracks. Therefore, PCAST argues, these strategies should be adopted by more teachers in more institutions across the country, in order to improve student outcomes.

In response to complaints from both industry and academe that entrants do not meet necessary mathematics standards, PCAST's (2012) second recommendation was to "launch a national experiment in postsecondary mathematics education to address the mathematics-preparation gap" (p. 27), including substantial support from the National Science Foundation for experiments examining bridge programs for high school students, mathematics courses designed and taught by faculty from mathematics-intensive fields such as engineering and computer science, and new pathways for producing mathematics teachers from mathematics-intensive fields. Studies like this one, comparing the effectiveness of diverse progressive pedagogical approaches, also contribute to this national experiment.

The story told by the reports summarized above is disconcerting. Despite the great importance of Calculus I for both the national economy and individual students' career goals, persistence is low, and this can be traced to the effects of poor teaching. The studies suggest further that supporting more innovative and engaging teaching methods will improve student outcomes and persistence. This study contributes to this literature by examining the effects of several different pedagogical models, including both traditional and innovative strategies, on student persistence in calculus in particular and STEM major tracks in general.

2.1.1. CSPCC

The present study is an outgrowth of another part of the "national experiment" recommended by PCAST: a national study, supported by the MAA, attempting to identify characteristics of successful programs in college calculus (abbreviated CSPCC; CSPCC, 2012). In this section, I describe this study, summarize some of the early results of this work (Bressoud, Carlson, Pearson, & Rasmussen, 2012; Bressoud, Carlson, Mesa, & Rasmussen, 2013; Rasmussen & Ellis, 2013; Sonnert, Sadler, Sadler, & Bressoud, 2014), and explain how my study dovetails with this ongoing research.

The CSPCC study consists of two phases. In the first phase, surveys (available on the website) were sent to over 14,000 students at 212 colleges and universities, selected by stratified random sampling, across the nation. These surveys collected information on student demographics, attitudes and beliefs, and high-school preparation. This data was examined to determine which institutions were most successful on several axes, including students' final grades, persistence, and attitudes and beliefs about mathematics. In the second phase, 17 institutions identified as successful were the subject of case study visits, comprising focus group interviews, class observations, and interviews of many administrative and teaching staff at each institution.

Researchers in the CSPCC team have begun to report results of the survey phase. Bressoud, Carlson, Mesa, and Rasmussen (2013) reported some basic descriptive statistics about the calculus students in the sample. Notably, 61% took a calculus class in high school; of these students, 62% took an Advanced Placement AB course, 13% took a BC course, and 34% earned a 3 or higher on one of the Advanced Placement exams. 11% of students were taking calculus for at least the second time in college. Students were generally confident in their preparation and abilities: 95% believed they had the knowledge and abilities necessary to succeed in the course, 58% expected to earn an A in their calculus class, and 94% expected to earn at least a B. However, the actual distribution of final grades was much lower, with 22% earning an A, 28% a B, and 27% earning a D, F, or withdrawing. This last statistic is remarkable; to emphasize, over a quarter of students in the calculus classes surveyed did not pass their class, despite their preparation and confidence entering the class.

Bressoud et al. (2012) reported on further results that emerged when pre-term and post-term surveys were examined together. They found that on a 1 to 6 scale, with 6 being the most confident, the average score on a measure of student confidence dropped from 4.89 to 4.42, with an effect size of -0.46. Similarly large negative effect sizes were found on students' enjoyment of mathematics (-0.27) and intention to continue into Calculus II (-0.20).

Focusing more closely on "switchers," or those who initially intended to take Calculus II but then changed their mind, Rasmussen and Ellis (2013) found that switchers were disproportionally female (56.1%, while females made up only 41.5% of the STEM-intending population), disproportionally attended a large national university (45.6% of switchers, compared to 32.6% of STEM-intending students), did not differ significantly from persisters in their academic preparation, and lost more confidence and enjoyment in mathematics than persisters. However, they generally did well in Calculus I, with 81% achieving a grade of C or higher; this last finding implicates more factors than academic achievement in students' decision to stop taking calculus.

Tallman and Carlson (2012) also reported on the results of an analysis of final examinations submitted by the instructors in the sample. Applying an adaptation of Bloom's taxonomy of cognitive behaviors, the vast majority of items (78.7%) were coded as "recall and apply procedure," a fairly low-level cognitive behavior, while none were coded as "create," a behavior involving generative, synthetic thinking. However, instructors seemed to believe that their exams were less procedural than they actually were: when asked what percentage of points on a typical exam focused on procedural skills, the median response was 50%.

Sonnert et al. (2014) reported on the role of "good teaching" in students' beliefs. When the suite of questions about students' perception of instructors' behaviors was subjected to factor analysis, two clusters of questions emerged: one labeled "progressive teaching" and a second labeled "good teaching." Of these, "good teaching" was most strongly correlated with improved student attitudes about mathematics. Among other key behaviors, an instructor with a high "good teaching" score asked questions to check students' understanding, listened carefully to students' questions, acted as if students were capable of understanding calculus, and provided understandable explanations of key ideas. In addition, such instructors assigned exams that students felt were fair assessments of what they had learned, and graded exams and homework fairly. Notably, instructors at large state universities tended to rank lower on the "good teaching" scale than those at masters-granting, four-year, or two-year institutions.

My study is an offshoot of the CSPCC study. Instead of examining data from a broad national sample of colleges and universities, I examined data from four different classes at one institution. This has the benefit of controlling for many institution-level variables, such as institution size, funding, demographics, and student support resources. I gave the students and instructors in my classes the same surveys used by CSPCC, and was thus able to compare the data collected at my institution to the national sample, as well as to various subsets of the national sample, including the restricted sample of institutions identified for further study. This allowed me to compare my classes not just to each other, but also to a representative sample of classes nationwide. I anticipate that the results of both studies will inform each other in developing a rich picture of successful college calculus programs.

2.2. Affect and beliefs

To fully understand student success in calculus, it is necessary to examine students' attitudes, beliefs, and dispositions about mathematics. Students' mathematical beliefs, including confidence, self-efficacy, and self-concept, correspond strongly with achievement in mathematics classes (Pajares & Miller, 1995; Carlson, 1999, Schommer-Aikins, Duell, & Hutter, 2005). Pajares and Miller (1995) asked 391 students to provide various types of self-efficacy judgments, including their confidence in their ability to solve particular problems. They then asked the students to solve the problems on which their self-confidence had been assessed. They found that self-efficacy was strongly correlated with problem-solving performance (Pearson's $r = .69$).

Carlson (1999) studied the beliefs and behaviors of 34 graduate students, using an instrument called the Views About Mathematics Survey (VAMS), which is discussed in greater depth later in this review, to assess their beliefs. These students, who were evidently successful in mathematics, showed high levels of persistence and enjoyment in problems involving mathematical reasoning, and believed in the value of individual effort in reaching solutions.

Schommer-Aikins et al. (2005) gave questionnaires on epistemological and problem-solving beliefs and administered a test of mathematical problem-solving ability to 1269 middle-school students. Their step-wise regression analysis revealed that the more students believed in the usefulness of mathematics, and the less they believed in "quick/fixed" learning of mathematics, the better they were at problem solving and communicating their solutions. Additionally, a path analysis suggested that beliefs predicted overall grade-point average in mathematics. These findings are consistent with the researchers' prior work (Schommer, Calvert, Gariglietti, & Bajaj, 1997), which demonstrated that belief in quick learning is negatively correlated with academic performance, as measured by grade-point average.

These studies showing quantitative links between affect and performance do not address the reasons why these links exist; several other researchers (e.g., Schoenfeld, 1992; Carlson & Bloom, 2005; Carlson, Bloom, & Glick, 2008) have documented links between beliefs and behavior. For instance, when asked how many buses of a given seating capacity would be required to transport a certain number of people, many students wrote that the answer would be "31 remainder 12" (Schoenfeld, 1992, p. 71). Answers like these, Schoenfeld argued, result from the belief that the "bottom line" of mathematics problems is a straightforward application of an algorithm, and that the contexts given in many problems are little more than “cover stories” (p. 70). Schoenfeld attributed many of these beliefs to the influence of teachers, classrooms, and culture, noting that teacher beliefs “tend to come home to roost” in the beliefs of the next generation of students and teachers (p. 73).

Beliefs, attitudes, and emotions also have an important effect on problem-solving behavior. Carlson and Bloom (2005; Carlson et al., 2008) studied the problem-solving behaviors of 12 research mathematicians as they completed challenging problems. In the process of developing a general problem-solving framework, they found consistent application of affective resources across all the mathematicians. Sense-making behavior was driven by mathematicians’ strong curiosity and high interest. In the planning and conjecturing phase, the mathematicians were influenced by feelings of intimacy, familiarity, and ownership of the problem, as well as beliefs about the nature of mathematics. When constructing solutions, mathematicians showed aesthetic considerations, as well as feelings of mathematical integrity that drove them to validate their solutions. Additionally, feelings of frustration impeded the progress of even these experienced research mathematicians; however, they were able to employ various coping strategies, rely on their confidence in their mathematical abilities, and persist in finding a solution.

Carlson et al. (2008) agreed with current reform documents (e.g., National Council of Teachers of Mathematics, 1989, 2000; National Research Council, 1996) in recommending that teachers provide explicit scaffolding and attention to help students develop productive mathematical beliefs. In particular, they encouraged teachers to explicitly discuss with students how to manage emotions that occur during problem solving, arguing that "students should adopt the belief that frustration, disappointment, and elation are all natural responses to the problem solving process" (p. 286). To illustrate how such discussions may look in practice, they presented two examples from the classroom of Bloom. In these vignettes, she encouraged students to become intimate enough with problems that they could think about them in unrelated situations, helped them persist through frustration, prompted them to reflect on the efficiency and aesthetic quality of their solutions, asked them if they were pleased with their final solution, and encouraged them to reflect on their self-image as mathematicians and to experience pride and satisfaction in their hard work.

Similarly, Goldin (2004) reported on the use of problems in discrete mathematics to help students develop appropriate responses to feelings of frustration and impasse. He argued that the object should be to replace reactions of nervousness and anxiety, leading to the activation of psychological defenses, with reactions of curiosity and a sense that the problem is interesting, leading to exploratory behaviors. The problem he considered was to use a 3-liter pail and a 5-liter pail to obtain exactly 4 liters of water. This problem has several qualities which can be leveraged to support the development of productive beliefs about mathematics: it invites possibly non-successful trial solutions, suggests a hidden structure, and is conducive to exploratory questions like "What can you do with the buckets?" He emphasized that "there is in an important sense *no way to go wrong*" in attempting initial solutions to this problem (p. 57; emphasis in original).

A focus on improving students' beliefs implies a need for accurate assessments of students' beliefs. Many instruments have been developed to assess various dimensions of the affective domain in mathematics. For instance, Plake and Parker (1982) developed and validated the Revised Mathematics Anxiety Rating Scale (RMARS) to diagnose and help treat mathematics anxiety. The Fennema-Sherman Mathematics Attitude Scales instrument (Fennema & Sherman, 1976) was designed to gain information concerning males' and females' learning of mathematics, as well as factors influencing the decision to take mathematics courses beyond the required minimum; among others, it includes scales measuring mathematics anxiety, confidence, and usefulness of mathematics.

An instrument of particular importance to this study, the Views About Mathematics Survey (VAMS; Carlson, Buskirk, & Halloun, 1998), was developed to assess and characterize undergraduate students' beliefs about knowing and learning mathematics, including motivation, perseverance, personal control, and the nature of mathematics. Carlson et al. distinguished between "expert" views, corresponding to those most commonly held among professional mathematicians, and "naïve" views, corresponding to the view commonly attributed to those with little or no mathematics background. For example, an expert-like belief about mathematics is that it is "a coherent body of knowledge about relationships and patterns contrived by careful investigation," whereas a more naïve view is that mathematics is "a collection of isolated facts and algorithms" (p. 8).

Carlson et al. (1998) used this instrument to assess the beliefs of approximately 600 undergraduate mathematics students. Echoing the results of studies discussed earlier, they found that a majority of undergraduates held non-expert views about the nature of mathematics, and that course achievement correlated significantly with self-confidence and expert-like mathematical beliefs. In particular, high-performing students in their sample (those who received either an A or B in their course) were more likely to hold expert-like views about mathematics than were lower-performing students, and students with higher self-confidence were more likely to continue their study of mathematics. They found further that college students' views did not shift appreciably over the course of one semester, even when the instruction involved frequent group work and use of graphing technology.

The studies discussed above demonstrate conclusively that attitudes and beliefs are important components of student development, due to their correlation with achievement and their influence on behavior. Further, the literature has shown that students' beliefs are influenced by teaching and classroom structure. It is thus the responsibility of researchers and teachers to find and document teaching strategies that help students develop positive beliefs, measure these improvements in a reliable way, and disseminate the most productive strategies to be used in classrooms across the country. This study contributes to the literature by examining the effect of the four pedagogical strategies on student beliefs, attitudes, and dispositions toward mathematics.

2.3. Inverted classrooms

Inverted classrooms (Lage, Platt, & Treglia, 2000) are a revision of the traditional lecture-based classroom model. There are many different approaches to teaching an inverted classroom, but the common feature is that some proportion of lecture content is delivered outside of class time, often via internet videos. The class time thus freed up is typically spent in problem-solving activities with TA or professor assistance; thus, "events that have traditionally taken place *inside* the classroom now take place *outside* the classroom and vice versa" (p. 32; emphasis in original), hence the "inverted" moniker. Proponents of this model argue that it accommodates a wide variety of learning styles (Lage et al., 2000), allows students to study content at their own pace before coming to class, and frees up valuable faculty "face time" to be spent in direct engagement with students (Gannod, Burge, & Helmick, 2008).

Inverting a class is a theoretically-grounded way to increase student understanding. It moves less conceptually-demanding tasks, which require less expert help, outside of the classroom and replaces them in the classroom with more demanding tasks; thus, the utility of class time with the more-knowledgeable other physically present is maximized (Vygotsky, 1978; Talbert, 2014). Numerous studies have shown that student success increases when students are actively engaged (Freeman et al., 2014), and the inverted model frees up class time for active learning by moving lecture outside of class.

The inverted classroom has been studied in a number of disciplines in tertiary education, including physics (Deslauriers, Schelew, & Wieman, 2011), economics (Lage et al., 2000), computer science (Gannod, 2007; Gannod et al., 2008), and biology (Moravec, Williams, Aguilar-Roca, & O'Dowd, 2010). Many of these investigators have seen surprising improvement in learning outcomes over traditional lecture classrooms, as well as favorable reactions from students in their classes. Naturally, the reports in the literature will be skewed toward successful implementations; thus, a careful examination of success reports in the literature yields a list of commonalities that make an inverted classroom successful.

Deslauriers et al. (2011) compared student learning gains over one week of two large-enrollment introductory undergraduate physics classes, one taught by an experienced, highly-rated professor in traditional lectures, and the other taught using an inverted method by an instructor who was inexperienced but trained in physics education and pedagogy. Both classes covered a common unit on electromagnetic waves and completed a common end-of-unit test jointly developed by the instructors involved. The mean score on the end-of-unit test in the experimental section was 74%, while the mean score in the control section was 41%. The standard deviations in both sections were approximately 13%, so the score in the experimental section was more than two standard deviations higher than the score in the control section. Additionally, to assess students' reception of the inverted method, the experimenters asked students to complete an online survey after the unit. 90% of students in the experimental section indicated that they enjoyed the inverted technique, and 77% felt they would have learned more if the whole course had been taught in this style. It should be noted that the experimental section also utilized peer instruction and clickers, and it is thus difficult to isolate the effects of the inverted presentation from those of peer instruction.

Lage et al. (2000) studied students' perceptions of an introductory economics course taught using an inverted model. Lectures were available via videotape and PowerPoint with sound, and students were expected to come to class having read assigned sections of the textbook and having watched the relevant lecture. Class time was spent in small groups, conducting economic experiments or labs, or completing worksheets. On an end-of-term survey, students had favorable reactions to the course, with generally agreeing responses to survey questions such as "I prefer this classroom format to a traditional lecture." Instructors also noted that students appeared to be more motivated and enjoyed the group-work components of the course. Further, the researchers argue that their evidence suggests that the inverted model may help attract and retain female students, who are typically underrepresented in many STEM disciplines.

Gannod (2007; Gannod et al., 2008) used an inverted method with lectures delivered via video podcast to teach a software engineering course. They report no findings on student performance, but on an end-of-term survey, students in this course reported a great deal of confidence in their ability to program, and indicated that they felt the ability to pause, rewind, and replay portions of the podcasted videos was beneficial to their learning.

To free up class time for active learning exercises in an introductory biology class, Moravec et al. (2010) shifted some content into "learn before lecture" (LBL) activities. They moved four to five slides from the PowerPoint lectures used the year before into either narrated PowerPoint videos or PDF worksheets, made available two days before class. Students were assigned to submit electronic copies of either their completed worksheet or the notes they took on PDF versions of the PowerPoint slides, for which they received a token amount of class credit; over 90% of the students completed the assignments. On the final exam, students performed 21% better than students in previous years on the questions assessing content delivered through LBL activities, compared to <3% improvement on all other questions (typical of year-to-year variability in difficulty of exam questions). Additionally, students reported that the LBL activities were helpful in learning the course material and preparing for lectures, as well as reviewing material after turning in the relevant assignments.

Inverted models are also gaining acceptance, and becoming the focus of research attention, in the secondary classroom. Musallam (2010) gave advanced high school chemistry students a pre-training screencast before a lecture on chemical equilibrium. In comparison to a control group of students who only attended the lecture, students who received the pre-training scored significantly lower on self-reported mental effort, and performed almost one standard deviation better on a post-test.

After hearing positive reactions to podcasting from university students, Bergmann and Samms (2008) decided to podcast lectures in their high-school chemistry class and use class time for hands-on, inquiry-based activities. To assess the efficacy of the podcasting model, they used the same tests as they had used the year before; while the average scores in the podcasting year were nearly the same as those in previous years, the prerequisites had been lowered, suggesting that the podcasting model was more effective for student learning than the traditional model. Students appreciated the ability to work from home at their own pace and to pause and replay difficult parts of a lecture. The researchers also reported significant buy-in from initially-skeptical parents.

The inverted classroom in tertiary mathematics education has been the subject of recent attention (e.g., Bowers & Zazkis, 2012; Overmyer, 2013; Talbert, 2014; Wasserman, Norris, & Carr, 2013). Bowers and Zazkis (2012) reported that students in an inverted calculus classroom did not perform very well on a final exam consisting of difficult procedural items, and questioned students' ability to make critical connections between calculus topics. However, they did not compare results in the inverted classroom to results in another classroom.

Wasserman, Norris, and Carr (2013) compared student performance in an inverted Calculus III classroom (n = 41) and a traditional lecture classroom (n = 40). They found that while lower-achieving students' performance appeared to improve in the inverted classroom, and students in the inverted classroom appeared to perform slightly better on conceptual items, there were no statistically significant differences between the classes. This is perhaps attributable to the small number of students involved in the study.

Talbert (2014) used inverted classroom design principles to structure a series of in-class workshops in linear algebra. He created highly-structured pre-class assignments called "guided practice," which included learning objectives, a collection of resources, a set of exercises, and requirements for submitting responses. An example in-class workshop asked students to work in pairs or threes to explain whether given numbers and vectors are eigenvalues and eigenvectors of a particular matrix, and then to explore the results of repeatedly applying a stochastic matrix to different initial vectors. Students enjoyed these workshops, and every student rated themselves as either "satisfied" or "very satisfied" with their learning in the workshops.

The reports summarized above have several features in common. First, student reports of the affordances of the inverted model are fairly consistent: students enjoy the ability to pause and rewind lectures as necessary, and the ability to watch lectures on their own schedule. Also, student buy-in is consistent, with most students reporting that they feel the inverted model is good for their learning. The general trend of quantitative performance data in those studies that report it is that students perform better, or at least as well, in inverted classrooms than in traditional classrooms. Despite the differences in content area, delivery mechanism, and use of class time, these reports also share commonalities in their implementation: pre-lecture activities were tailored to the particular class, often personally created by the instructors or researchers; time formerly occupied by lecture was replaced with active-learning exercises with the substantial involvement of the instructor; and students were held accountable for completing pre-lecture activities. These commonalities can perhaps be considered a seed for a list of best practices for inverting a classroom.

These reports also suffer from several limitations. First, it is difficult to disentangle the effect of the "inverted" part of the inverted model from other altered pedagogical strategies; the learning gains attributed to the inverted model may just as convincingly be attributed to an increased number of contact hours or to the active-learning strategies implemented in place of lecture. Second, each of the comparison studies surveyed above compare the inverted model to traditional lecture, rather than to other innovative pedagogies, such as inquiry-based learning without a pre-class activity. The question of how the inverted model compares to these other innovations thus remains unanswered.

This study contributes to the literature on the inverted model by assessing the efficacy of one implementation of this pedagogical strategy in calculus. Additionally, it compares the inverted model with several other innovative pedagogical approaches, particularly active-lecture style classrooms.

2.4. Opportunities to learn

Watson (2003) has spoken of the mathematics classroom as "an arena in which there are various opportunities to learn mathematics" (p. 29). These opportunities to learn may be presented through various modalities; in this study, I am particularly interested in the roles of discourse and classroom culture (Boaler, 1998; Stipek et al., 1998; Kazemi & Stipek, 2001; Wood, Williams, & McNeal, 2006; Nickerson & Bowers, 2008; Mesa & Chang, 2010) and technology (Jackiw, 1995; Jiang & McClintock, 1997; Olive, 1998; Purdy, 2000; Scher, 2000; July, 2001; McClintock, Jiang, & July, 2002; Sinclair, Zazkis, & Liljedahl, 2004; Sinclair, 2008) in presenting students opportunities to learn.

2.4.1. Discourse and classroom culture

Nickerson and Bowers (2008) documented two interaction patterns that emerged in an upper-division course for secondary mathematics teachers. They called the first the Elicit-Respond-Elaborate (ERE) pattern, to point out its similarity to Mehan's (1979) well-known Initiate-Respond-Evaluate (IRE) pattern. In this pattern, the teacher Elicited student thinking on a particular mathematical object of interest, the students Responded, and the teacher Elaborated on students' perspectives, both to model appropriate mathematical vocabulary and ways of reasoning and to spark and encourage further discussion and justification. They termed their second pattern Proposition-Discussion (PD); in this pattern, either the teacher or the student would make a proposition, and then others would discuss it. On several occasions, Nickerson and Bowers observed a cycle between ERE and PD that they argued helped the class develop appropriate means of justification and supported more mathematically sophisticated ways of reasoning. They argued further that by the end of the class, students' views about the relative merits of calculational and conceptual approaches to mathematics had shifted significantly because of these patterns of interaction.

Wood et al. (2006) examined 42 lessons from one conventional and four reform elementary mathematics classrooms, documenting examples of children's mathematical thinking and patterns of social interactions. They found that in the conventional classroom, students were most often engaged in recalling information, which they argued is not higher-order mathematical thinking and is thus less conducive to students' development. They also noted important differences between various reform-oriented classroom cultures: in one classroom whose dominant pattern was classified as "strategy reporting," students generally explained (expressing higher-order thinking) their strategies to the teacher in dyadic patterns of interaction, while in another classroom described as "inquiry/argument" dominant, there were more opportunities for open discussion between all students, more establishment of shared mathematical reasoning, and more kinds of mathematical thinking. They concluded that "social interaction patterns established in the classroom specifically affect how children construct mathematical knowledge in that classroom," and that "interaction patterns that required greater involvement from the participants were related to higher levels of expressed mathematical thinking" (p. 248).

In a similar comparison of interactive paradigms, Mesa and Chang (2010) studied the language used by two instructors teaching two undergraduate mathematics classes exhibiting high levels of student participation (as measured by number of student turns per minute). They used the framework developed by Martin and White (2005), in which engagement language is divided into *monogloss* voices, seeking to assert facts in a non-negotiable way, and *heterogloss* voices, seeking to engage the audience and either open (*expanding heterogloss*) or close (*contracting heterogloss*) options for dialogue. Using this framework to code instructor utterances, they found that while Class A seemed more interactive, with more student turns and more turns overall, Class B had a much higher proportion of heterogloss (63%, compared to 48%), and particularly *expanding* heterogloss (47% vs. 28%), utterances, and thus presented more opportunities for students to engage in classroom dialogue. Additionally, Instructor A produced a much higher proportion of monogloss utterances (51%, as compared to 34% for Instructor B), suggesting that he maintained a more authoritarian position in the classroom. Thus, while both classrooms were very interactive on their face, there were important and measurable differences in the type and quality of the interaction: "even in seemingly highly interactive settings, there may be little room for students to include their own perspectives or voices into the dialog. However, our analysis shows that it is possible to organize classroom discourse in a way that does [provide room for this]" (p. 97).

Kazemi and Stipek (2001) analyzed lessons on fraction addition from elementary-school classrooms, coding the discourse for several dimensions of motivational strategies. In particular, they examined the amount of "press for conceptual learning" (p. 59) created by each exchange, measured by how much an exchange emphasized effort and deemphasized performance, focused on understanding, and supported student autonomy. They found significant differences in the amount and type of justification required for an explanation to be acceptable: in high-press exchanges, students linked their problem-solving strategies to mathematical justification, whereas in low-press exchanges, students simply described the steps they took to solve the problem. They concluded that while all the teachers taught in progressive ways that supported and encouraged student participation, "the differences among the high- and low-press exchanges provide evidence for the need to go beyond these superficial teaching practices to examine the nature and degree of conceptual thinking" (p. 78). In the language of the emergent perspective (Cobb & Yackel, 1996; Yackel & Rasmussen, 2002), the classrooms had similar social norms but different sociomathematical norms, and differences in sociomathematical norms were linked to differences in student outcomes.

The findings of this report were consistent with previous quantitative work (Stipek et al., 1998) examining the effect of teachers' motivational practices, including enthusiasm and interest in mathematics, press for conceptual learning, and emphasis on student effort, on students' beliefs and learning, both procedural and conceptual. Stipek et al. found that a positive affective climate was a strong predictor of students' motivation, which was in turn significantly associated with learning gains on procedural items. Further, degree of press was significantly correlated with growth in conceptual understanding.

While not reporting explicitly on the role of discourse, Boaler (1998) described the impact of classroom culture on students' achievement and views on mathematics. In a three-year case study comparing students at Amber Hill, whose mathematics program was a traditional textbook-based approach, and students at Phoenix Park, where mathematics was more open-ended and project-based, she found dramatic differences in student attitudes and performance. She gave both sets of students an assessment involving an applied task (finding the volume of a house) and several procedural tasks. Phoenix Park students performed significantly better than Amber Hill students on the applied task, and performed comparably on the procedural tasks; students at Amber Hill did not believe they could use mathematics to solve real-world problems. Additionally, Amber Hill students held a view of mathematics as “rule following:” mathematics for them was about memorization and rote application of rules and formulas. At Phoenix Park, by contrast, students noted an emphasis on understanding and explaining methods and solutions.

The message of these studies is consistent and twofold: first, student outcomes are impacted by discourse and culture, and second, while interaction is an important feature of high-learning classrooms, interactive classrooms are not necessarily supportive of the development of higher mathematical thinking. Significant differences were found in each case between classrooms that, on their face, are similar in their emphasis on student involvement, engagement, and interaction; some models encouraged more substantive student engagement than others. Classrooms with interaction patterns that merely involve students in repeating known facts were shown in each instance to be less helpful than those with interaction patterns that encouraged student sense-making, justification, and conceptual understanding. In other words, there is interaction, and then there is interaction.

These studies also provide a methodological model for the rigorous study of classroom interactions, and particularly for differentiating between apparently similar forms of interactive discourse. My study contributes to this literature by examining the effects of various types of classroom discourse on student performance and beliefs.

2.4.2. Technology

The use of technology can also open opportunities for students to learn. In particular, Geometer’s Sketchpad (here abbreviated GSP; Jackiw, 1995), a dynamic geometry program in which students can construct and transform geometric figures through dragging and programming motion (Scher, 2000), provides students with opportunities to interact with typically-static figures in geometry, helping them deepen their geometric intuition (Olive, 1998; Purdy, 2000; McClintock et al., 2002).

Scher (2000) provided a historical account of the development of GSP. Eugene Klotz proposed an interactive computer program as an outgrowth of a project producing videotapes focusing on three-dimensional geometry, and recruited Nicholas Jackiw to program the software. Klotz's initial vision was that the program would be nothing more than "a way for students to draw accurate, static figures from Euclidean geometry," but when Jackiw showed Klotz how the Macintosh mouse could be used to directly manipulate vertices or sides of geometric figures, Klotz was "flabbergasted," and this became a key feature of GSP (p. 44). Jackiw argued for a lean set of pre-programmed constructions available from menus, together with a robust scripting feature that would allow users to define their own constructions; this approach, he argued, was more productive because it "forces [students] to confront and think through the geometry implicit in a construction" rather than simply making constructions available "as a magic recipe" (p. 46). These features and others, Scher argues, contribute to an aesthetic of exploration and play, and users "ultimately… come to see mathematics less as a collection of rules and procedures and more as an ongoing human endeavor" (p. 48).

July (2001; McClintock et al., 2002) studied tenth-grade students as they used GSP to construct two-dimensional projections of three-dimensional objects. She found that students' spatial ability, as measured by van Hiele level of geometric reasoning demonstrated on a pre-test and post-test, improved significantly, especially for lower-achieving students. She argued further that GSP helped students develop intuition and experiential understanding of three-dimensional objects, which could later be leveraged to scaffold more formal work in geometry. She concluded that carefully-designed GSP environments are viable ways to teach students three-dimensional geometry, despite their intrinsic restriction to a two-dimensional computer monitor.

Jiang and McClintock (1997) reported on the use of GSP with preservice teachers. They posed the teachers a problem about minimizing the length of a path between two points, A and B, on opposite sides of a river, which would need to be crossed with a bridge. The students initially conjectured that the bridge would best be placed at the midpoint of the straight segment AB, but found that this was not always the case; they thus “came to realize that they should always test their ideas by appropriate investigations” (p. 130). Eventually, after noticing salient characteristics of the diagrams produced by GSP, the students formulated and justified a correct conjecture, and expressed satisfaction in their efforts. The researchers pointed out the role of GSP in revealing geometric structure and thus scaffolding students’ intuition, and concluded that GSP is a useful tool in the “guess-investigate-conjecture-verify” process central to learning geometry (p. 135).

Purdy (2000) used GSP to help his high-school geometry students understand the “maximum-volume box problem:” what size corners should be cut out of a square sheet to maximize the volume of the box created by folding the flaps up? Students began working on this problem by assembling paper boxes, discovering the volume function, and finding the maximum on a graphing calculator. However, they were unsatisfied with this solution, so the instructor conjectured that a GSP sketch might help them develop more intuition for the solution. He said of the sketch he created that "it is dynamic and virtually tactile in the way that no paper model could be. The student can vary the side length and the cutout size, find the maximum volume, and confirm his or her results from the graphing calculator and the paper models" (p. 226), and his students were much more satisfied with their solutions after exploring the GSP sketch. He argued, however, that it would be best if the students were to create the sketch themselves, rather than have it given to them by the teacher; this would have the tangential benefit of giving students a great deal of practice with construction problems in an applied, non-contrived context.

Olive (1998) provided a broader look at the educational implications and uses of GSP. He related an anecdote of his 7-year-old son using GSP to decide that a degenerate triangle with all three of its vertices collinear is still a triangle, and to form and test conjectures about the motion of points. Surveying a number of reports of GSP usage in schools, he noted its use not just in geometry classrooms, but also to explore conic sections, trigonometry, graphs of functions, topics in calculus, perspective drawing, optics, and mechanical linkages; in several of these reports, the least advanced students benefitted the most from the use of GSP. He argued that a dynamic geometry approach has the potential to foster expert-like beliefs about mathematics by "[giving] students the opportunity to engage in mathematics as mathematicians, not merely as passive recipients of someone else's mathematical knowledge" (p. 399).

Sinclair (2008; Sinclair et al., 2004) examined students' use of two microworlds, called Number Worlds and the Colour Calculator, to explore concepts and solve problems in elementary number theory. Sinclair and her colleagues found that their students used the technology to gain insight and intuition, to produce graphical displays suggesting underlying mathematical patterns, to test conjectures, and to confirm results that they had derived in non-empirical ways. For instance, after interacting with the Number Worlds microworld, several of their students developed a stronger image of the "every *n*th property" of multiples; one student said that "multiples follow a pattern in that every 3rd number is highlighted if we want multiples of 3, it gives us an actual image, not just words to describe it" (p. 240). This is illustrative of the effect computers may have in helping students develop rich, robust concept images to accompany their concept definitions (Tall & Vinner, 1989).

Sinclair (2008) also presented a vignette in which she described her use of GSP to help her explore the behavior of "reflex triangles" formed by reflecting each of the vertices of a triangle across the opposite side. By dragging one vertex of the initial triangle, she was able to develop conjectures about the "niceness" of various types of triangles, and gain visual insight into productive avenues of reasoning. Further, she noted the connections between GSP's emphasis on dynamic representations of geometric figures and the type of "transformational reasoning" (Harel & Sowder, 1998) that appears to be vital to constructing formal proofs. This suggests that GSP may be a useful tool for helping students develop transformational reasoning and thus better proof-writing abilities.

Zazkis (2013) studied the use of a GSP applet called the Tangent Intuition Applet (TIA), which was created for use in a technology-enhanced calculus class. The TIA makes use of several "slope-widgets," consisting of two connected points, one of which bisects a small line segment intended to act as a tangent. The other point controls the slope of the tangent. When the tangent points are placed on the graph of a function, and the slopes of the tangents are adjusted to correspond to the local slope of the function, the slope-control points lie on the graph of the derivative of the function. Zazkis argued that this applet could help build intuition for the meaning of slope at a point, as well as the connection between derivative at a point and derivative of a function. Zazkis found that several students who had used the applet in class continued to reason in ways inspired by the applet, even in paper-and-pencil settings where the applet was not available.

Again, the message from these studies is consistent. GSP and other technologies are useful tools for helping students develop intuition and conceptual understanding of geometric ideas; since geometric ideas find themselves applied in a wide variety of mathematical contexts, GSP's benefits spill over from geometry into algebra, trigonometry, calculus, and a host of other fields. A common feature of all these reports is that the learning gains only accrue to the person holding the mouse. It seems that it is not enough to show students GSP sketches; the power of GSP appears to be in students themselves interacting with the drawings. My study contributes to this literature by examining the effects of GSP usage by college calculus students on their conceptual understanding.

Chapter 3: Methodology

3.1. Theoretical perspective

In this work, I align myself with a learning theory known as the emergent perspective (Cobb & Yackel, 1996; Rasmussen, Yackel, & King, 2003). This theoretical stance coordinates constructivist accounts of individual psychological development with sociocultural accounts of growing participation in communities of practice. The emergent perspective holds that participation in classroom activity "constitute[s] the conditions for the possibility of learning" (Cobb & Yackel, 1996, p. 185), and that an individual's psychological development is enabled and constrained by their participation in classroom activities. Therefore, classes that present more opportunities for students' engagement and participation in classroom activities are seen as presenting more opportunities for student learning.

Gresalfi (2009; Gresalfi, Barnes, and Cross, 2012) conceptualize of opportunities for students' engagement as *affordances*. This term is borrowed from Gibson's (1979) work on perceptual affordances, which are the set of actions made possible by a certain object. For example, the affordances of a chair include sitting, but the affordances of a whiteboard do not. Similarly, tasks and teaching strategies provide affordances for student engagement. Gresalfi further conceptualizes of affordances as varying in *forcefulness* or *strength*, defined as the imperative a student is likely to feel to take up the affordance. In other words, if taking up an affordance is required, or if not taking it up would be a violation of norms or rules, then it is seen as strong; if there is no consequence for not taking up an affordance, then it is seen as weak. However, although increasing the forcefulness of an affordance increases its chances of being taken up by more students, it does not guarantee that the affordance will be taken up by all students; students must still choose to take up an affordance, and their choice is influenced by their attitudes and beliefs.

The emergent perspective holds that there is a reflexive relationship between individual student knowledge and classroom mathematical practices. Students' individual knowledge develops within the classroom microculture and is thus influenced by the shape of classroom activity, including classroom interaction patterns and collective ways of thinking. On the other hand, these collective ways of thinking emerge from the classroom participants, and are gradually modified as individuals' mathematical knowledge grows (Bowers & Nickerson, 2001).

The emergent perspective also provides an account of the mutual and reflexive development of individual beliefs and classroom norms. For instance, individual beliefs about mathematics are developed in the context of a classroom community in which certain social and sociomathematical norms regulate the pattern of interactions; individual beliefs are thus influenced by those norms. However, the norms are negotiated by the students and their instructor, all of whom are members of the classroom community. Since students' individual beliefs influence their interactions in and contributions to the classroom community, beliefs play a role in the negotiation of classroom norms (Cobb & Yackel, 1996; Yackel & Rasmussen, 2002).

This theoretical framework provides an orientation toward the types of data to collect and the types of analyses to conduct: because of its focus on the relation between classroom and individual development, the emergent perspective implies that I should collect both collective classroom and individual psychological data, and analyze the links between them, to build a complete account of learning. Further, because of its emphasis on beliefs and norms, the emergent perspective suggests that I give proper attention to the affective dimension of learning and development, not just measures of psychological achievement.

The emergent perspective is embedded in each of my three research questions. While quantitative questions like the first two could easily be asked within any theoretical framework recognizing the importance of the individual psychological dimension of learning, the emergent perspective's emphasis on beliefs and attitudes informs my interest in these as phenomena to investigate. It also informs the hypotheses embedded in the first two research questions: since the pedagogical strategies, and thus the learning opportunities they present students, differ on their face, I anticipate differences in student outcomes. Indeed, if I embraced a theoretical perspective that would lead me to think there would not be differences, I might not even be led to ask questions like these.

This perspective is what gives theoretical power to my third research question. In the first two, I look for differences in student outcomes between the classes. However, the mere identification of differences is not theoretically satisfying; the causes of the differences would remain unexplored and unexplained. The third research question is an attempt to find theoretically-plausible links between any differences uncovered in the first two research questions and the differences in the classes, and the emergent perspective is the theory that provides a rationale for linking differences in classes to differences in outcomes.

While many analyses conducted from the emergent perspective focus on documenting the emergence of classroom norms and mathematical practices, I am more interested in students' individual understanding and the effects of the classroom on its development. Accordingly, I have not collected the detailed, longitudinal classroom data that would be necessary to carry out such an analysis. My focus is on presenting a general picture of the intellectual life of the classroom and drawing conceptual links between this picture and student outcomes.

3.2. Setting

The setting for this study is a large and highly diverse public university in the southwestern United States. Undergraduate enrollment for Fall 2012, during which this study was conducted, was approximately 26,000, with large groups of Hispanic (28.8%) and Asian (14.0%) students joining the white plurality (37.6%). 55.3% of undergraduate students were women. No terminal doctoral degree in the mathematical sciences is offered at this institution, although it partners with several nearby universities to offer joint doctorates in mathematics education, computational sciences, and computational statistics. Several terminal master's degrees are offered, including in pure and applied mathematics, statistics, and biostatistics. The institution did not participate in the national MAA survey study of calculus programs.

This study focuses on four different Calculus I classes taught at this university during the Fall 2012 semester. I gave the teachers the pseudonyms Rachel, Corbin, Viktor, and Julie, and I refer to their classes as the Lecture class, the Lecture with Discussion (LD) class, the Lecture with Discussion and Technology (LDT) class, and the

Inverted class, respectively. Brief descriptions of each of their classes are below; more thorough descriptions of a typical day in each classroom are presented in chapter 6.

Rachel's class, the Lecture class, was a traditional, large-lecture approach to calculus. The whole class, approximately 180 students, met for 50-minute lectures three times a week, and broke into five sections of 30-40 students for 50-minute recitation sections led by a graduate TA once a week. During these recitation sections, the TA would assist students in solving homework problems, deliver supplemental lecture material, and answer questions on class material. Homework was completed online using Wiley+. There were five exams spaced approximately equally throughout the semester. Rachel held the position of lecturer and had average student reviews.

Corbin's class, the LD class, was a more interactive, student-centered lecture; he would commonly break from lecture to ask students to solve problems he wrote on the whiteboard. The class of approximately 60 students met for 105-minute lectures twice a week, and there were no TA recitations. There were two midterms spaced approximately equally in the semester. Corbin held the position of lecturer and had excellent student evaluations.

Viktor's class, the Inverted class, was taught using an inverted model: lecture content was delivered outside of class time via internet videos selected from resources such as Khan Academy, and class time was used by students to solve problems in small groups. The students, approximately 100, met twice a week for 105 minutes. At the end of each class, one problem was selected to be handed in and graded as homework. Viktor's class had no TA recitations, but the TAs were involved in helping answer student questions during class time; Viktor did not attend class. There were two exams spaced approximately equally in the semester. Viktor held the position of associate professor and had average student evaluations.

Finally, Julie's class, the LDT class, was an interactive, student-centered, technology-intensive lecture with TA recitations. Approximately 120 students met three times a week for 50-minute lectures, then broke into four sections for recitations led by a graduate TA with approximately 40 students. Homework was completed online using Wiley+. During her lectures, Julie made use of Geometer's Sketchpad applets whose aim was to help develop students' intuition for calculus concepts. There were five exams spaced equally throughout the semester. Julie held the position of associate professor and had good student ratings.

3.3. Data collection

The data corpus is comprised of four categories of data: enrollment data, surveys, measures of achievement, classroom observations, and focus group interviews. Consonant with the emergent perspective, I have collected data pertaining to both individual and classroom attributes. The relationship between research questions and data sources is summarized in Table 3.3.1, and elaborated in the sections that follow.

Table 3.3.1. Relationship between data sources and research questions

<table>
<tr><th colspan="2">Research Question</th><th>Data Sources</th></tr>
<tr><td rowspan="3">1) What impact do the four different instructional approaches have on students':</td><td>a) persistence in STEM major tracks?</td><td>• Local survey data
• Focus group interviews</td></tr>
<tr><td>b) attitudes, dispositions, and beliefs about mathematics?</td><td>• Local survey data
• Focus group interviews</td></tr>
<tr><td>c) conceptual and procedural achievement in calculus?</td><td>• CCR
• CCI
• Final exam</td></tr>
<tr><td rowspan="2">2) How do students at the local institution compare to students in the national database in:</td><td>a) persistence in STEM major tracks?</td><td>• Local survey data
• National survey data</td></tr>
<tr><td>b) attitudes, dispositions, and beliefs about mathematics?</td><td>• Local survey data
• National survey data</td></tr>
<tr><td colspan="2">3) How do the differences in opportunities for learning between the four classes contribute to the differences in outcomes?</td><td>• Classroom observations
• Focus group interviews
• Results of RQ 1 and 2</td></tr>
</table>

3.3.1. Surveys

Near the beginning of the term, students completed surveys designed by the MAA for use in the ongoing Characteristics of Successful Programs in College Calculus study. These surveys included questions about students' demographic information, mathematical preparation (e.g., their ACT or SAT scores and their experiences in high school mathematics classes), beliefs and attitudes about mathematics (e.g., their confidence in their mathematical abilities and their level of productive disposition), and college and career plans (e.g., in which field they intend to pursue a career, and whether they intend to take Calculus II). Near the end of the term, students took a similar survey, with many of the same questions measuring beliefs and attitudes, as well as a set of new questions examining their perceptions of their experience in Calculus I. These surveys are attached as Appendix A.

Surveys were administered via the internet. Students were sent an email through the course management system describing the survey and providing a link to the survey website. To ensure a high response rate, the survey was worth a token amount of class credit. The survey was designed in such a way that students can leave any question blank without affecting their ability to complete the rest of the survey. A unique identifier (student ID, assigned by the university) collected at the end of each survey allowed me to assign students credit for completing the survey and to link pre-term and post-term responses with other data collected from each student, but was deleted before any analysis of the data.

The survey data were used primarily in answering the first and second research questions. The surveys are the primary data source for assessing students' attitudes, beliefs, and dispositions about mathematics, and they also provide valuable demographic and baseline preparation data. Additionally, the questions about major and career tracks allowed me to examine students' persistence in STEM major tracks.

Since these surveys were the same instruments used in the CSPCC study, I was able to compare students' responses not only across the four classes at the institution under study, but also to responses given at 160 institutions across the country. I focused particular attention on comparing the four classes at the local institution to the institutions selected by the CSPCC researchers for further study, master's-granting institutions, and Ph.D.-granting institutions. I compared to both master's-granting institutions and Ph.D.-granting institutions since the local institution shares characteristics with both groups.

3.3.2. Measures of achievement

To provide a baseline measurement of students' preparation for calculus, students completed the Calculus Concept Readiness (CCR) instrument (Carlson, Madison, & West, 2010). This was used to control for any differences from class to class in the populations of students at the beginning of the term. At the beginning and the end of the term, students completed the Calculus Concept Inventory (CCI) instrument (Epstein, 2006), which has been developed to assess students' understanding of key concepts of differential calculus. Additionally, all students in the four classes took a common final exam that was jointly designed by all four instructors. These assessments allowed me to examine students' ability in calculus, both procedural and conceptual; thus, they are the primary source of data for answering Research Question 1c.

In each of the classes, I or the instructor read a standardized description of the CCR and CCI instruments, emphasizing that while performance on the assessments will not influence students' grade in any way, they were still to be taken seriously. Similarly to the surveys, a token amount of class credit was given to students who complete the CCI and CCR, regardless of their scores, to ensure a high response rate. The CCR is an online assessment; a handout was distributed both in class and via email to students explaining how to take the assessment.

The CCI is a paper-and-pencil multiple-choice assessment that was administered during class time, either in discussion section or during lecture. I proctored the administration of the CCI in each section. Thirty minutes were allowed for students to complete the CCI. I met individually with several students who were absent on the day the CCI was scheduled to be administered, and proctored a test session for them.

The common final exam was jointly developed by the four instructors to ensure that it represents topics that are covered in all four classes. Students in all four sections took the exam at the same time in the same room. It was graded by the instructors, their TAs, and me. To ensure consistency in the grading process, each problem was graded by the same person across all the exams.

Personally-identifiable information, in the form of university-assigned student ID numbers, was used to cross-link student responses to each assessment. However, all personally-identifying information was removed from all student responses before analysis.

3.3.3. Classroom observations

I identified two topics, related rates and the fundamental theorem of calculus, that were taught by each of the instructors, and observed the class sessions of each class that addressed these topic. Each topic was discussed for approximately 100-120 minutes of class time.

During these observations, I took extensive field notes documenting daily activities and common patterns of interaction in the classrooms. These sessions were audiorecorded to aid in qualitative analysis. I paid particular attention to the patterns of discourse used in each of the classrooms, and the opportunities these patterns open for substantive student participation in the work of the classroom, as well as the uses of technology and worked examples.

This data allowed me to produce detailed descriptions of the daily life of each classroom. Additionally, this data contributed to my analysis in Research Question 3, identifying theoretically-plausible explanatory relationships between the opportunities for learning provided by the structure of each class and student persistence, attitudes, and achievement.

3.3.4. Focus group interviews

In each class, a group of four to ten student volunteers participated in focus group interviews near the end of the term. Volunteers were solicited from class sessions a few days before the interviews were scheduled.

Each focus group was composed of students from only one class. These interviews focused on students' subjective assessment of their enjoyment of their calculus class. Students were asked to rate specific class features (e.g., students in Julie's class were asked to discuss the use of technology in their class), as well as provide general feedback about the overall structure of each class (e.g., students were asked to describe a typical day in class). I conducted the focus group interviews as semi-structured interviews, loosely following a protocol modified from that used by the CSPCC researchers. The protocol is attached as Appendix B. These interviews were audiorecorded and transcribed to aid in qualitative analysis.

This data served several purposes in my analysis. First, it helped me build a detailed picture of the daily life of the classroom, and understand how students felt about how their class is conducted. Students were asked what they liked and disliked about their classes, which gave me insight into their subjective enjoyment of the various pedagogical strategies. Additionally, the data I collected in focus group interviews helped triangulate the survey data relating to students' attitudes, dispositions, and beliefs about mathematics. Further, all of this information contributed to my analysis of Research Question 3, in drawing explanatory links informed by students' perspectives between variations in pedagogy and variations in outcomes.

3.4. Data analysis

The data described above were analyzed using a mixed-methods approach. In broad strokes, statistical methods were used to answer questions 1 and 2, and the answers to these questions fed into the qualitative analysis used to answer question 3. The emergent perspective provides the theoretical grounding for postulating links between student outcomes, as explored in the first two research questions, and student perceptions of their classroom experiences, examined in the third.

3.4.1. Research Question 1

Research Question 1 comprises three sub-questions, comparing persistence, attitudes and beliefs, and conceptual and procedural achievement.

3.4.1.1. Comparison of persistence in STEM major tracks

Before explaining the statistical tests that I employed to determine differences between classes, I first explain how I measured persistence. As a proxy for STEM intention, which I define as intent to graduate in a STEM major track, I used student enrollment in Calculus II (or intention to take Calculus II). Intention to take Calculus II is a proxy for STEM intention because most STEM major tracks require Calculus II, while most non-STEM majors do not. Following Rasmussen and Ellis (2013), students were sorted into four categories: *culminators, persisters, switchers,* and *converters.* Culminators are those students who did not at any point in the term intend to take Calculus II; often, they are those whose major only requires Calculus I, and are thus non-STEM majors. Persisters are those students who both began and ended the term intending to take Calculus II. Switchers are those students who began the term intending to take Calculus II, but changed their minds during the semester and decided not to take Calculus II. Finally, converters are those who began the term without intentions of taking more calculus, but then decided during the term to take Calculus II. For the purposes of this study, I am primarily interested in the numbers of switchers and persisters.

I used two methodologies to determine switchers and persisters: first, the methodology employed by Rasmussen and Ellis (2013), which I refer to as the CSPCC methodology, and second, a methodology using student enrollment records.

The CSPCC methodology relied on student responses to the start-of-term survey (STS) and end-of-term survey (ETS). I used several survey questions from both the STS and the ETS to determine students' intentions. The two most important questions were

STS 26 and ETS 3, which both ask, "Do you intend to take Calculus II?" Students could respond to this question with "yes," "no," or "I don't know yet" on the STS or "I'm not sure" on the ETS. If a student marked "yes" on both STS 26 and ETS 3, they were classified as a persister; if they marked "yes" on STS 26 and "no" on ETS 3, they were classified as a switcher. I did not include in my analysis students who marked "no" on STS 26, because these students were not initially STEM-intending.

If students marked "I don't know yet" or "I'm not sure" for one of these two questions, they were more difficult to classify, and I examined their responses to several other survey questions. ETS 5 asked students, "When you started this class, did you intend to take Calculus II?" If a student answered "yes" to this question and marked "I don't know yet" or provided no response to STS 26, I classified them as initially-STEM-intending (as if they had answered "yes" to STS 26).

ETS 4 was particularly useful for classifying students who answered "I'm not sure" on ETS 3. The prompt for this question was "If you do not intend to take Calculus II, check all reasons that apply," with a list of checkboxes:

> I never intended to take Calculus II.
> (*) I changed my major and now do not need to take Calculus II.
> (*) My experience in Calculus I made me decide not to take Calculus II.
> I have too many other courses I need to complete.
> To do well in Calculus II, I would need to spend more time and effort than I can afford.
> (*) My grade in Calculus I was not good enough for me to continue to Calculus II.
> (*) I do not believe I understand the ideas of Calculus I well enough to take Calculus II.

The starred responses are particularly informative, because they say directly that something about a students' Calculus I experience made the student change their mind about taking Calculus II. If an initially-STEM-intending student ("yes" on STS 26 or ETS 5) marked one of these starred responses, I classified them as a switcher. If a case was still too ambiguous to code after examining these triangulation questions, I excluded it from analysis; there were very few such cases. This methodology is summarized in the flowchart in Figure 3.4.1.1.1.

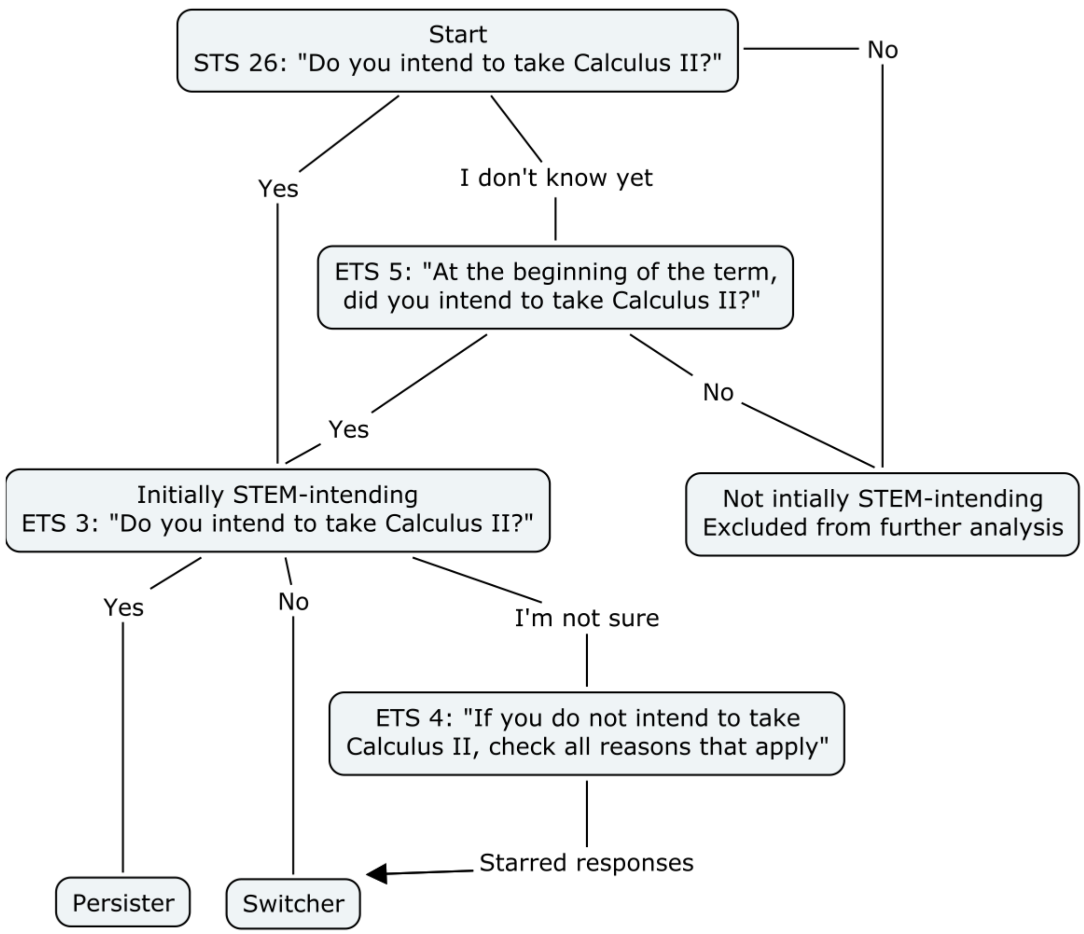

Figure 3.4.1.1.1. CSPCC methodology flowchart

The other methodology I employed to identify switchers and persisters relied on student enrollment data. For each of the next four semesters (Spring, Summer, and Fall

2013, and Spring 2014), I obtained enrollment records for each section of Calculus I and Calculus II. This allowed me to determine which students actually enrolled in Calculus I and Calculus II. Furthermore, the enrollment data included each student's current major code, and I classified each of the majors as STEM or non-STEM. Majors with a substantial mathematics component, in particular those requiring Calculus II, were classified as STEM. I thus had a direct measurement of the STEM intention of every student that was enrolled in Calculus I in the Fall 2012 semester.

An initially STEM-intending student was classified as a switcher if they were never again enrolled in any calculus course, or if they were declared in a non-STEM major if they were ever again enrolled in Calculus I. An initially STEM-intending student was classified as a persister if they enrolled in Calculus II by Spring 2014 and remained declared in a STEM major throughout this period.

Once switchers and persisters had been identified, I used chi-square analysis to determine whether there were significant differences between the proportions of switchers and persisters in each of the four classes. If the classes had no impact on student persistence, I would expect switchers and persisters to be approximately equally distributed across the four classes. Chi-square analyses allowed me to test this hypothesis, and determine which classes, if any, were more effective in encouraging persistence in STEM major tracks.

3.4.1.2. Comparison of attitudes, dispositions, and beliefs about mathematics

There were 12 beliefs items on the STS and 15 beliefs items on the ETS. I used ANOVA to compare the beliefs questions from the ETS, item-by-item, across the four classes. This allowed me to determine whether one class of students had more expert-like beliefs at the end of the term than the other classes.

While several beliefs questions are worded identically on the STS and the ETS, most are worded differently. To enable me to conduct more comprehensive comparisons of these questions, I conducted a conceptual analysis of the beliefs items on both the STS and the ETS, with the intent of grouping them together. The items grouped together in two sets: the first set assessed affective beliefs about mathematics (that is, how one feels about mathematics, one's confidence in their mathematical abilities, etc.), and the second set assessed cognitive beliefs about the nature of mathematics (that is, what one thinks mathematics is, or how it is supposed to be done or learned). I then created aggregate scores which measured students' affective and cognitive beliefs about mathematics. First, I converted the Likert-scale responses to each of the beliefs items into z-scores, and negated the z-scores of reverse-coded items. I then calculated the aggregate affective beliefs score by taking the average of the z-scores on the items in the affective beliefs cluster, and calculated the aggregate cognitive beliefs score similarly. Once these aggregate scores had been created, I compared the mean scores of the four classes using ANOVA. I also used ANCOVA to compare end-of-term beliefs while controlling for incoming beliefs.

3.4.1.3. Comparison of conceptual and procedural achievement in calculus

When comparing the growth of student understanding of calculus, the tacit assumption is that the populations of students do not differ significantly from class to class. This assumption can be validated by the careful examination of the baseline data I collected. I used ANOVA to determine if there were significant differences between classes in students' CCR scores. Differences here would indicate that students in one class were better prepared for calculus than those in other classes.

Once the assumption of equality in student population had been tested, the next step was to test for differences in achievement. One measure of student achievement is post-term scores on the CCI; this can be viewed as a measure of end-of-term conceptual understanding. Similarly, I employed normalized gain (Simon, Kohanfars, Lee, Tamayo, & Cutts, 2010) as a measure of gains in conceptual understanding. I used ANOVA to determine if there were differences between the four classes in either of these measures.

Another measure of student achievement is student scores on the final exam. This measures not only students' conceptual understanding, but also their ability to work specific problems. I conducted ANOVA and post-hoc analyses to test whether any classes outperformed others. I also categorized the final exam questions into procedural and conceptual categories, and used ANOVA to compare the classes' scores on these two subsets of the final exam.

The baseline data, and CCR scores in particular, can be used not only to determine whether there are differences in the population of each class, but also to control for individual student differences when comparing scores across classes. Each of the ANOVAs described above was repeated as an ANCOVA, with baseline data serving as the covariate. This tested the assumption that differences uncovered by ANOVA were attributable to the effects of the different pedagogical approaches, rather than any differences in students' mathematical preparation.

3.4.2. Research Question 2

Research Question 2 comprises two sub-questions, comparing persistence and beliefs at the local institution to the national database.

3.4.2.1. Comparison of persistence in STEM major tracks

For this analysis, since I compared to data collected by CSPCC researchers, I used the CSPCC methodology for determining switchers and persisters. I compared the proportions of switchers and persisters in the local data to those in each of four slices of the national database: the overall data, the data from selected institutions, the data from master's-granting institutions, and the data from Ph.D.-granting institutions. I used chi-square analyses to assess the significance of differences in proportions.

3.4.2.2. Comparison of beliefs and attitudes about mathematics

The analysis for this question was similar to that outlined in section 3.4.1.2 for Research Question 1b. Using the same conceptual groupings of STS and ETS items reported in that section, I created aggregate scores which measured students' affective and cognitive beliefs about mathematics. I then used a set of t-tests to compare the average scores on each of these variables at the local institution to those in each of the four subsets of the national database described earlier. I also employed ANCOVA to control for start-of-term beliefs while comparing end-of-term beliefs scores in each of the four national groups.

To develop a more detailed understanding of how beliefs shifted over the term, and to determine whether there were differences in the patterns of change, I examined students' responses to the items that were identical on the STS and ETS, and compared the change in these items for each of the five groups of interest. I used two-way ANOVAs and t-tests to assess the significance of differences.

3.4.3. Research Question 3

The theme of this research question is to draw explanatory links between differences in the classes and differences (or non-differences) in outcomes. Thus, the first task is to identify the differences in classrooms, and in particular, in the opportunities to learn presented to students in each of the four classes. I examined the classroom observation data and the focus group interview data to develop a description of the opportunities to learn afforded by each pedagogical strategy.

Through classroom observations, I documented from the outsider perspective the types of activities that are common in each classroom. Through focus group interviews, I interrogated the insider perspective on these activities, to determine how much insiders feel they are emphasized.

After conducting classroom observations, I employed grounded theory (Strauss, 1987) to analyze the focus group interview data. I took several analytic passes through the data. On the first pass, I employed open coding to tag student comments. I then employed axial coding to combine the codes developed in the first pass into themes. Several themes were unique to individual groups, while other themes recurred in several focus group interviews.

Once the analyses of Research Question 1 and 2 were completed, I had data illuminating the differences and the non-differences in the classes' outcomes. Significant differences in student outcomes are interesting results that should be explained by reference to the varied opportunities for learning presented by each of the classes. This is the portion of my analysis most directly impacted by my theoretical alignment with the emergent perspective: in order for these explanatory links to be convincing, I must be able to provide a theoretical rationale connecting the differences in affordances to learn to the differences in outcome.

In addition, non-differences are interesting non-results: on their face, the classes are quite different, and according to my theoretical perspective, differences in outcomes are to be expected. Non-differences should then be explained by identifying unexpected commonalities in the classrooms, or by identifying ways in which the opportunities to learn do not differ as much as would be expected. Again, these explanatory links must be theoretically grounded.

Chapter 4: Local Quantitative Comparisons

My interest in studying these four classes is fundamentally driven by my desire to improve student outcomes in Calculus I. I want to know if any of the four instructional approaches under investigation are better for students in any measurable way. My first research question is thus as follows: What impact do the four different instructional approaches have on students':

- persistence in STEM major tracks?
- attitudes and beliefs about mathematics?
- conceptual and procedural achievement in Calculus I?

In this chapter, I examine the local quantitative data that I collected to answer this question.

4.1. Grade distribution and DFW rates.

In my research questions, I chose several sets of student outcome variables to compare across the four classes. I considered including grade distribution or DFW rate (rate of D's, F's, or withdrawals; essentially, the rate of students not passing the class) as one of these outcome variables. However, grading practices might differ dramatically from instructor to instructor: in addition to the usual differences between the weight that different instructors assign to different components of the course (attendance, exams, homework, etc.), instructors might make larger adjustments. For instance, the instructor of the Inverted class offered "the C bargain" to his students: any student who earned a grade of C or better on the final exam and who had excellent attendance would be given no less than a C for the course, regardless of their performance on any other assessment. This adjustment could artificially suppress the DFW rate in the Inverted class relative to classes whose instructors do not have such a policy. (In practice, only four students utilized the C bargain; three had their grades raised from C- to C, and one from D to C.) Therefore, any test of differences between the classes would be liable to produce both false positives (significant differences that were due to instructors' grading policies, rather than something structural about the class) and false negatives (real differences hidden by instructor policies). Since any test would thus be unreliable, I elected not to include grade distribution or DFW rate as an outcome variable.

However, the impact of a student's grade on other outcome variables, particularly persistence, is non-negligible, so it was still important to examine the grade distribution in each of the four classes. This section thus serves as necessary background for the other analyses that will follow.

Grade distribution differed dramatically and statistically significantly ($\chi^2(9, n = 477) = 71.1, p < .001$) across the four classes, as shown in Table 4.1.1 and the stacked bar chart in Figure 4.1.2.

Table 4.1.1. Grade distribution and DFW rates

		Lecture	LD	LDT	Inv.	Total
A	Count (%)	46 (25.4%)	3 (4.4%)	10 (7.9%)	5 (4.9%)	64 (13.4%)
	Expected	24.3	9.1	24.3	13.7	
B	Count (%)	48 (26.5%)	10 (14.7%)	44 (34.9%)	26 (25.5%)	128 (26.8%)
	Expected	93.8	29.4	70.8	58.9	
C	Count (%)	43 (23.8%)	18 (26.5%)	40 (31.7%)	49 (48.0%)	150 (31.4%)
	Expected	56.9	21.4	39.6	32.1	
DFW	Count (%)	44 (24.3%)	37 (54.4%)	32 (25.4%)	22 (21.6%)	135 (28.3%)
	Expected	51.2	19.2	35.7	28.9	

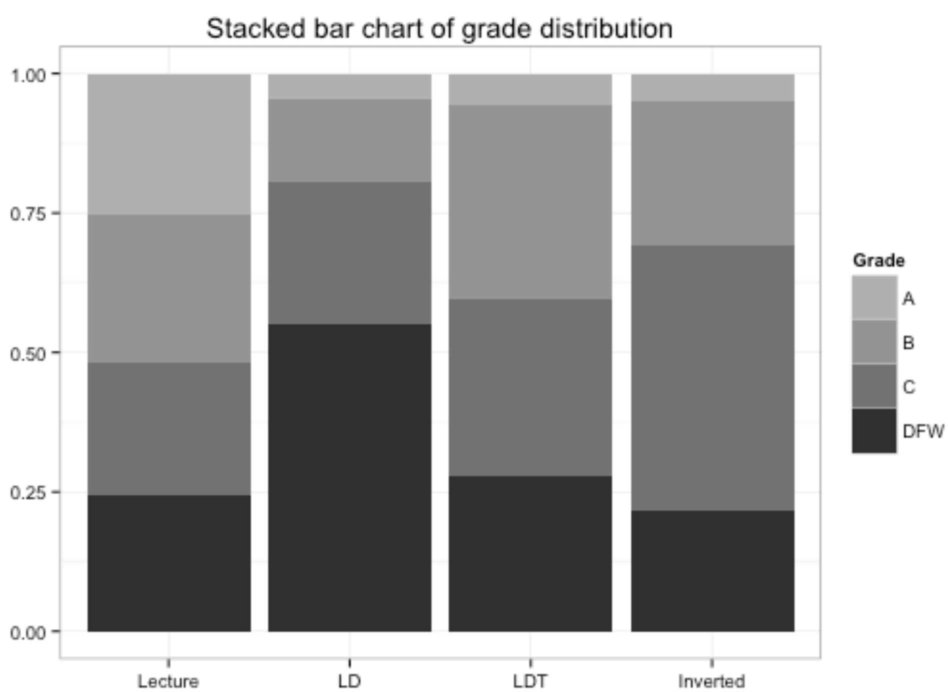

Figure 4.1.2. Stacked bar chart of grade distribution

Particularly noteworthy is the 54.4% DFW rate in the LD class; this is almost double the average rate for the four classes. Corbin, the instructor of the LD class, indicated that this was his highest-ever DFW rate in a Calculus I class. This was the earliest of the four classes, meeting at 7:30am on Tuesdays and Thursdays, and Corbin noted that attendance was poor; it was also the last of the four classes to be added to the schedule. Additionally, Corbin does not curve grades or allow exam corrections. Thus it appears that the difference in DFW rates is attributable to the circumstances of the class and the grading policies of the instructor.

4.2. Persistence

One of my research questions is to investigate the effects of the four classes on the rate at which students left STEM majors. As a proxy for STEM intention, which I define as intent to graduate in a STEM major track, I used student enrollment in Calculus II (or intention to take Calculus II). Intention to take Calculus II is a proxy for STEM intention because most STEM major tracks require Calculus II, while most non-STEM majors do not.

To examine this question, I conducted two analyses of student persistence, using two different methodologies and two different sets of data, to understand the impact Calculus I had on students' persistence in STEM major tracks. Students can be classified into four groups: switchers, persisters, culminators, and converters. Persisters are those who both began and ended the term intending to take Calculus II. Switchers are those who initially intended to take Calculus II but changed their minds during the course of the term. Culminators are those who both started and ended the term not intending to take

Calculus II. Finally, converters are those who initially did not intend to take Calculus II, then decided during the course of the term to take it. I am primarily interested in identifying switchers and persisters, since they are the students who were initially STEM-intending. In the following sections, I repeat the description of each methodology from section 3.4.1, discuss its strengths and weaknesses, and present the results of the tests I used to determine whether there were significant differences in persistence across the four classes.

4.2.1. CSPCC methodology

First, I identified switchers and persisters following the methodology used in the CSPCC study (Rasmussen & Ellis, 2013). I used several survey questions from both the STS and the ETS to determine students' intentions. The two most important questions were STS 26 and ETS 3, which both ask, "Do you intend to take Calculus II?" Students could respond to this question with "yes," "no," or "I don't know yet" on the STS or "I'm not sure" on the ETS. If a student marked "yes" on both STS 26 and ETS 3, they were classified as a persister; if they marked "yes" on STS 26 and "no" on ETS 3, they were classified as a switcher. I did not include in my analysis students who marked "no" on STS 26, because these students were not initially STEM-intending.

If students marked "I don't know yet" or "I'm not sure" for one of these two questions, they were more difficult to classify, and I examined their responses to several other survey questions. ETS 5 asked students, "When you started this class, did you intend to take Calculus II?" If a student answered "yes" to this question and marked "I don't know yet" or provided no response to STS 26, I classified them as initially-STEM-intending (as if they had answered "yes" to STS 26).

ETS 4 was particularly useful for classifying students who answered "I'm not sure" on ETS 3. The prompt for this question was "If you do not intend to take Calculus II, check all reasons that apply," with a list of checkboxes:

> I never intended to take Calculus II.
> (*) I changed my major and now do not need to take Calculus II.
> (*) My experience in Calculus I made me decide not to take Calculus II.
> I have too many other courses I need to complete.
> To do well in Calculus II, I would need to spend more time and effort than I can afford.
> (*) My grade in Calculus I was not good enough for me to continue to Calculus II.
> (*) I do not believe I understand the ideas of Calculus I well enough to take Calculus II.

The starred responses are particularly informative, because they say directly that something about a student's Calculus I experience made the student change their mind about taking Calculus II. If an initially-STEM-intending student ("yes" on STS 26 or ETS 5) marked one of these starred responses, I classified them as a switcher. If a case was still too ambiguous to code after examining these triangulation questions, I excluded it from analysis; there were very few such cases. This methodology is summarized in the flowchart below, reproduced from chapter 3.

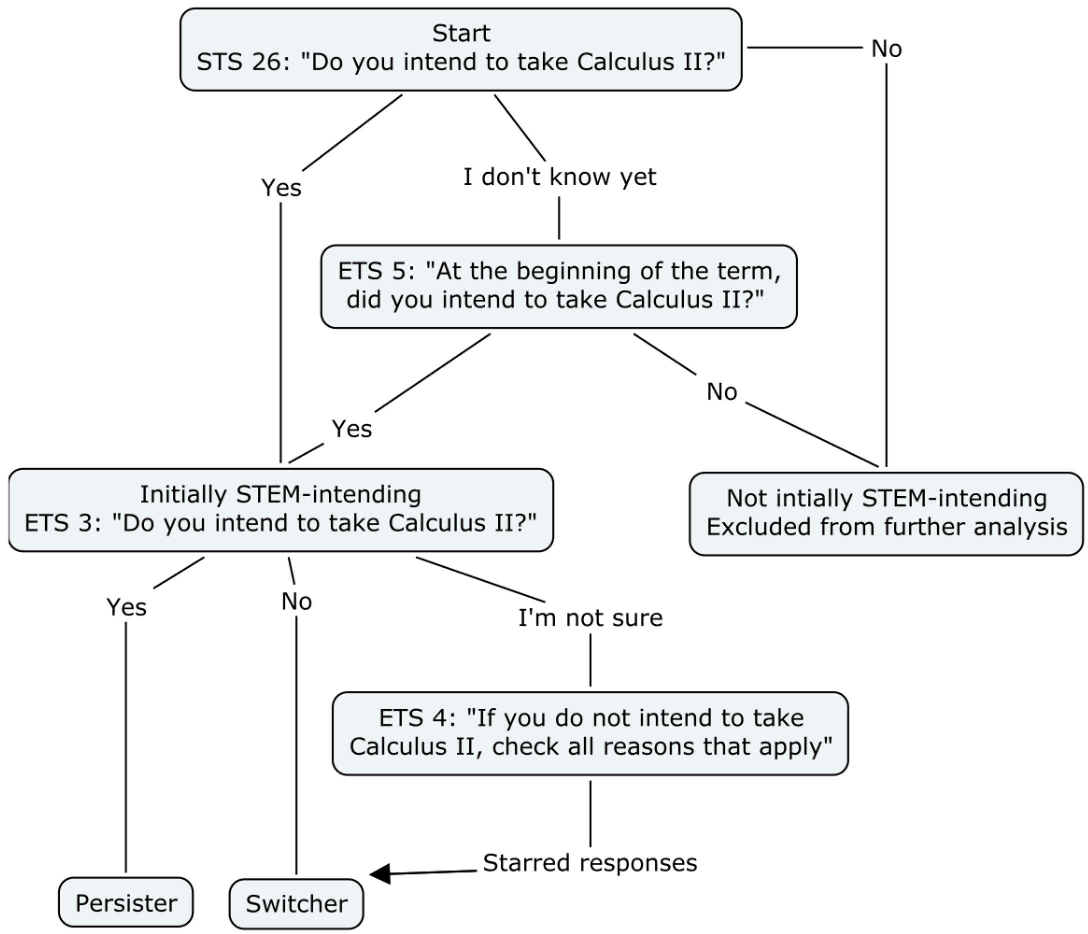

Figure 4.2.1.1. CSPCC methodology flowchart

Using the CSPCC methodology, I identified a total of 18 switchers and 260 persisters in the four classes. I conducted a chi-square analysis to determine if there were significant differences between the proportions of switchers and persisters in the four classes. The test reported no significant difference ($\chi^2(3, n = 278) = 0.914$, $p = .822$); I thus cannot reject the hypothesis that the differences in the four classes have no impact on students' persistence.

Table 4.2.1.2. Switchers and persisters, CSPCC methodology

	Lecture	LD	LDT	Inverted	Total
Switchers	7 (6.9%)	1 (3.0%)	6 (7.9%)	4 (6.3%)	18 (6.6%)
Expected	6.7	2.2	5.0	4.2	
Persisters	94 (93.1%)	32 (97.0%)	70 (92.1%)	59 (93.7%)	255 (93.4%)
Expected	94.3	30.8	71.0	58.8	

The 6.6% switcher rate reported here is much lower than either the 12.5% switcher rate reported by the CSPCC researchers (Rasmussen & Ellis, 2013) or the 60% switching rate reported by PCAST (2012). As will be discussed in the next section, the methodology used to calculate this switcher rate is likely to lead to an underestimate of the true switcher rate.

4.2.1.1 Strengths and weaknesses

This methodology is well-suited to answer the research question it addresses. The phenomenon under investigation is the impact of students' experiences in Calculus I on their persistence decisions -- in other words, I am interested in counting the number of students who switch out of STEM major tracks specifically because of their experience in their Calculus I class -- and this methodology isolates the effect of Calculus I. The data used in this methodology come from a survey administered in the context of a Calculus I class, and as discussed earlier, several of the questions used to classify students ask directly about students' experience in Calculus I. I will discuss later how the other methodology might include the impact of other classes on students' persistence decisions. Additionally, it is the methodology used in the CSPCC study, which allows me to make comparisons to a much larger national data set.

However, this methodology is not without its weaknesses. First, as will be explained in the following paragraphs, it probably under-reports the incidence of switching. Second, it uses intention to take Calculus II as a proxy for intention to complete a STEM major track; as discussed earlier, this is a reasonable proxy, but it is a proxy nonetheless. Finally, in order for a student to be classified using this methodology, they must have completed the ETS; thus, the number of students that can be classified depends on the survey response rate. In my sample, I was able to classify just 273 (57%) of the 478 students who took the final exam.

There are several factors that suggest the switcher rate obtained through this methodology is an underestimate. First, the ETS data used in this methodology was collected one or two weeks before the final exam, and thus before final grades were posted. It is likely that students overestimated their final grade, which would lead to an overestimate in persistence; in particular, recall from the grade distribution analysis presented earlier that 54.4% of the students in the LD class did not pass.

Second, the students classified in this table are particularly conscientious; they were the students who completed the ETS. This table does not include students who withdrew (either officially or unofficially) from the class before the end of the term, or those who simply did not bother to fill out the surveys. If these less-conscientious students could be captured, it is likely that the switcher rate reported here would rise.

4.2.2 Roster data methodology

The second methodology, instead of inferring students' STEM intention from their survey responses, determined students' STEM intention more directly. For each of the next four semesters (Spring, Summer, and Fall 2013, and Spring 2014), I obtained enrollment records for each section of Calculus I and Calculus II. This allowed me to determine which students actually enrolled in Calculus I and Calculus II. Furthermore, the enrollment data included each student's current major code, and I classified each of the majors as STEM or non-STEM. Majors with a substantial mathematics component, in particular those requiring Calculus II, were classified as STEM. I thus had a direct measurement of the STEM intention of every student that was enrolled in Calculus I in the Fall 2012 semester.

An initially STEM-intending student was classified as a switcher if they were never again enrolled in any calculus course, or if they were declared in a non-STEM major if they were ever again enrolled in Calculus I. An initially STEM-intending student was classified as a persister if they enrolled in Calculus II by Spring 2014 and remained declared in a STEM major throughout this period.

Using this methodology, I was able to classify a total of 405 students as switchers or persisters, in comparison to the 275 I was able to classify using the CSPCC methodology. Comparing proportions of switchers and persisters using this methodology yielded different results than using the CSPCC methodology; these results are summarized in the table below.

Table 4.2.2.1. Switchers and persisters, roster data methodology

	Lecture	LD	LDT	Inverted	Total
Switchers	19 (11.6%)	9 (17.0%)	25 (22.9%)	18 (22.8%)	71 (17.5%)
Expected	28.8	9.3	19.1	13.8	
Persisters	145 (88.4%)	44 (83.0%)	84 (77.1%)	61 (77.2%)	334 (82.5%)
Expected	135.2	43.7	89.9	65.2	

I conducted a chi-square analysis of this table to determine whether there was a significant difference in the proportions of switchers and persisters between the four classes. The test approached significance ($\chi^2 = 7.732$, $p = 0.052$). Noting that the switching rate was substantially lower in the Lecture class, but that the rates were fairly close in each of the other classes, I grouped together the LD and LDT classes and compared the proportions of switchers and persisters again. The proportions are given in Table 4.2.2.2 below.

Table 4.2.2.2. Switchers and persisters, LD and LDT classes grouped

		Lecture	LD/LDT	Inverted	Total
Switchers	Count	19 (11.6%)	34 (21.0%)	18 (22.8%)	71 (17.5%)
	Expected	28.8	28.4	13.8	
Persisters	Count	145 (88.4%)	128 (79.0%)	61 (77.2%)	334 (82.5%)
	Expected	135.2	133.6	65.2	

A chi-square analysis of this table revealed that the difference was statistically significant ($\chi^2 = 6.8571$, $p = 0.032$), indicating that students from the Lecture class switched out of STEM major tracks significantly less frequently than those in the other classes.

4.2.2.1 Strengths and weaknesses

This methodology has several particular strengths. First, it uses a direct measurement of STEM intention, which I define as intent to complete a STEM major track, whereas the CSPCC methodology measures students' intention to take Calculus II and infers students' STEM major-track intention from their Calculus II intention. Examining major codes, and thus students' actual declared majors, is the most direct way to assess their STEM intention. Secondly, this approach allows me to classify many more students; in this methodology, I was able to classify 405 (84.7%) of the 478 students who took the final exam. It is likely that the remaining students are culminators or converters.

However, this methodology suffers from its longitudinal approach to classifying students. Again, the phenomenon under investigation is the impact of students' experiences in Calculus I on their persistence decisions. This methodology has no way of isolating the contribution of Calculus I, and is thus likely to be an overestimate of the Calculus I-influenced switching rate. To illustrate, consider the following hypothetical situation: Isabella intends to major in chemistry, and thus enrolls in both Calculus I and general chemistry in the Fall 2012 semester. She enjoys her Calculus I class and achieves a good grade, but does not enjoy her general chemistry class. She changes her mind about pursuing a chemistry degree and instead changes her major to political science. Under this second methodology, she would be classified as a switcher, even though her decision to switch was not influenced by her Calculus I experience.

4.2.3 Discussion

It is noteworthy that the more innovative classes had higher switching rates (as measured by the second methodology) than the Lecture class. One possible explanation could be that students are used to a particular mode of instruction, and that when that mode is changed, students react poorly. This points to the need for innovative classes to explain themselves and achieve student buy-in. Instructors who wish to adopt innovative pedagogies have to spend a significant amount of time explaining the benefits of the changes being made. As will be seen in Chapter 5, students at this institution have less expert-like beliefs about how mathematics should be learned, so the need for instructors to justify innovative pedagogies may be even more marked than usual.

Viewed through the lens of the emergent perspective (Cobb & Yackel, 1996), innovative pedagogies disrupt the usual suite of classroom norms; in particular, they typically upset the roles of teacher and student, usually requiring the student to take a much more active role in their own education. However, classroom norms are negotiated by both the teacher and the students, rather than imposed by the instructor. In other words, classroom change is not an entirely top-down process; while it may be initiated by the teacher, it requires student buy-in to be successful. So, one explanation for the higher switching rates in the non-traditional classes is that they did not achieve student buy-in.

Given that the CSPCC methodology likely produces an underestimate of the Calculus I-influenced switcher rate, and the roster data methodology likely produces an overestimate, one can combine the two methodologies to provide an upper and lower bound for the actual rate. Overall, then, I can say with some confidence that the true Calculus I-influenced switcher rate lies between 6.6% (the estimate given by the CSPCC methodology) and 17.5% (given by the roster data methodology). This corroborates the 12.5% rate reported by the CSPCC researchers, though it is likely that this rate is also an underestimate.

Future work at the national level might incorporate a longitudinal analysis similar to the one conducted here. Institutions might be asked to provide anonymized enrollment data, perhaps with randomly-assigned unique identifiers for each student, for all sections of Calculus I and Calculus II for a period of two or three years. This would allow researchers to use the second methodology to provide an upper bound to an estimate of Calculus I-influenced switcher rate.

4.3. Beliefs and attitudes

4.3.1. ETS beliefs items

There were 15 belief items on the ETS; I used ANOVA to determine if there were significant differences between the classes in the responses to the ETS items. Five items were identified as differing significantly between classes. For those five items, I conducted post-hoc comparisons to determine the precise location of differences. The results are summarized in Table 4.3.1.1; unless otherwise indicated, higher scores indicate more favorable beliefs.

Table 4.3.1.1. Significant differences in ETS beliefs items

Item	ANOVA		Tukey:		
	F(3, 325)	p	Differences	MD	p
This course has increased my interest in taking more mathematics.	8.666	< .001	LDT > Inv. LD > Lecture LD > Inv.	.688 .849 1.259	.007 .003 < .001
I am good at computing derivatives and integrals.	4.953	.002	LDT > Inv. Lecture > Inv.	.564 .547	.008 .006
I am able to use ideas of calculus to solve word problems that I have not seen before.	2.718	.045	Lecture > Inv.	.502	.038
* My score on my mathematics exam is a measure of how well: (1 = I understand the covered material; 4 = I can do things the way the teacher wants)	3.766	.011	LDT < Inv. Lecture < Inv.	.410 .392	.021 .018
When studying mathematics in a textbook, I tend to: (1 = memorize it the way it is presented; 4 = make sense of the material so that I understand it)	4.235**	.007	Lecture > LDT Lecture > Inv.	.373 .413	.030† .044†

* Reverse-coded; ** Welch F(3, 137.994); † Tamhane's T2

One noteworthy commonality between all of these significant differences is that the Inverted classroom is on the unfavorable side of each. Indeed, of the ten significant contrasts revealed by Tukey post-hoc analysis, only two did not involve the Inverted classroom.

What might these differences mean? First, I will focus on the Inverted classroom, since that is the source of the majority of significant differences. When compared with students in other classes, an average student in the Inverted classroom is less confident in their ability to compute derivatives and integrals or use calculus to solve novel problems, less interested in taking more mathematics classes, more prone to memorization than sense-making, and more likely to think that mathematics must be done a specific way to earn a good score on an exam. In Chapter 6, I will show that these quantitative results are triangulated by the qualitative data from the focus group interviews; there are specific features of this Inverted classroom that are likely to promote these less-favorable views.

Backing away from the Inverted classroom, students in the LDT class were significantly more likely than students in the Lecture class to memorize instead of make sense of material. However, students in the Lecture class demonstrated less increase in their interest in mathematics than those in the LD classroom.

Given the differences between the way the four classes were taught, I was surprised that only five of the 15 beliefs items showed significant differences. This led me to hypothesize that the actual differences between the classes, in terms of the learning opportunities afforded to students, may have been smaller than I had originally anticipated. Chapter 6 will provide evidence in support of this revised hypothesis.

4.3.2. Conceptual groupings

As a dimension-reduction technique, I conducted a conceptual analysis of the beliefs items on both the STS and the ETS, with the intent of grouping them together. The items grouped together in two sets: the first set assessed affective beliefs about mathematics (that is, how one feels about mathematics, one's confidence in their mathematical abilities, etc.), and the second set assessed cognitive beliefs about the nature of mathematics (that is, what one thinks mathematics is, or how it is supposed to be done or learned). I checked these groupings with colleagues familiar with the survey instruments to ensure that these groupings were reasonable. Tables 4.3.2.1 and 4.3.2.2 summarize the groupings on the STS and ETS.

Table 4.3.2.1. Items in the affective grouping

Source	Item
STS	I believe I have the knowledge and abilities needed to succeed in this course.
	I am confident in my mathematical abilities.
	I understand the mathematics that I have studied.
	I enjoy doing mathematics.
	When experiencing a difficulty in my math class: (Alternatives: "I try hard to figure it out on my own" vs. "I quickly seek help or give up trying")
	My score on my mathematics exam is a measure of how well: (Alternatives: "I understand the covered material" vs. "I can do things the way the teacher wants")
	If I had a choice: (Alternatives: "I would never take another mathematics course" vs. "I would continue to take mathematics")
	The process of solving a problem that involves mathematical reasoning is a satisfying experience.
ETS	This course has increased my interest in taking more mathematics.
	I am confident in my mathematical abilities.
	I am good at computing derivatives and integrals.
	I am able to use ideas of calculus (e.g., differentiation, integration) to solve word problems that I have not seen before.
	My previous math courses prepared me to succeed in this course.
	I enjoy doing mathematics.
	When experiencing a difficulty in my math class: (Alternatives: "I try hard to figure it out on my own" vs. "I quickly seek help or give up trying")
	My score on my mathematics exam is a measure of how well: (Alternatives: "I understand the covered material" vs. "I can do things the way the teacher wants")
	If I had a choice: (Alternatives: "I would never take another mathematics course" vs. "I would continue to take mathematics")

Table 4.3.2.2. Items in the cognitive grouping

Source	Item
STS	For me, making unsuccessful attempts when solving a mathematics problem is: (Alternatives: "a natural part of solving the problem" vs. "an indication of my weakness in mathematics")
	My success in mathematics PRIMARILY relies on my ability to: (Alternatives: "solve specific kinds of problems" vs. "make connections and form logical arguments")
	When studying Calculus I in a textbook or in course materials, I tend to: (Alternatives: "memorize it the way it is presented" vs. "make sense of the material, so that I understand it")
	When solving mathematics problems, graphing calculators or computers help me to: (Alternatives: "understand underlying mathematical ideas" vs. "find answers to problems")
	The primary role of a mathematics instructor is to: (Alternatives: "work problems so students know how to do them" vs. "help students learn to reason through problems on their own")
	Mathematics instructors should show students how mathematics is relevant.
	If I am unable to solve a problem within a few minutes, it is an indication of my weakness in mathematics.
	Mathematics is about getting exact answers to specific problems.
ETS	Mathematics is about getting exact answers to specific problems.
	For me, making unsuccessful attempts when solving a mathematics problem is: (Alternatives: "a natural part of solving the problem" vs. "an indication of my weakness in mathematics")
	My success in mathematics PRIMARILY relies on my ability to: (Alternatives: "solve specific kinds of problems" vs. "make connections and form logical arguments")
	When studying Calculus I in a textbook or in course materials, I tend to: (Alternatives: "memorize it the way it is presented" vs. "make sense of the material, so that I understand it")
	When solving mathematics problems, graphing calculators or computers help me to: (Alternatives: "understand underlying mathematical ideas" vs. "find answers to problems")
	The primary role of a mathematics instructor is to: (Alternatives: "work problems so students know how to do them" vs. "help students learn to reason through problems on their own")

To confirm these conceptual groupings, I employed exploratory factor analysis separately on the STS and ETS beliefs items. High KMO values (.808 and .848, respectively) and significant results from Bartlett's test of sphericity indicated that the correlation matrices should be factorable. After examining the scree plots, I retained two components on both the STS and ETS. Oblimin rotation revealed that the groupings created by factor analysis closely paralleled the conceptual groupings reported above.

I then created aggregate scores which measured students' affective and cognitive beliefs about mathematics. First, I converted the Likert-scale responses to each of the beliefs items into z-scores, and negated the z-scores of reverse-coded items. I then calculated the aggregate affective beliefs score by taking the average of the z-scores on the items in the affective beliefs cluster, and calculated the aggregate cognitive beliefs score similarly.

Once these aggregate scores had been created, I compared the mean scores of the four classes using ANOVA. As expected, there were no significant differences between the classes on the STS affective or cognitive variables. There were also no significant differences between the classes on the ETS cognitive variable. However, ANOVA revealed a significant difference between class mean scores on the ETS affective variable, $F(3, 325) = 3.417$, $p = 0.018$. Tukey post-hoc analysis revealed two significant differences between class means: the mean score in the Inverted class was significantly lower than that in either the LDT class ($MD = 0.282$, $p = 0.023$) or the Lecture class ($MD = 0.260$, $p = 0.028$). This information is summarized in the means plot in Figure 4.3.2.3.

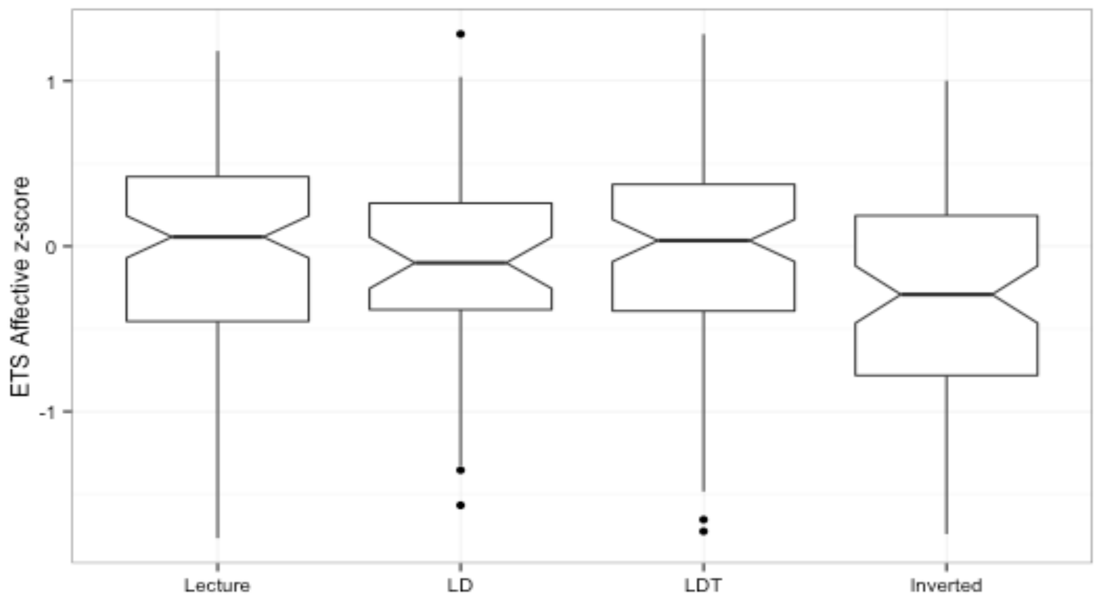

Figure 4.3.2.3. Means of ETS affective beliefs

Paralleling the earlier results obtained by comparing scores on each ETS belief item, students in the Inverted class had significantly less favorable affective beliefs about mathematics than students in either the LDT class or the Lecture class. In Chapter 6, I will discuss qualitative evidence that triangulates this quantitative result.

The lack of significant differences in the cognitive beliefs variable may indicate that despite differences on the face of the four classes, the kinds of things that students did to learn mathematics were not substantially different in each of the four classes. For instance, notwithstanding the LDT class's emphasis on developing conceptual understanding of calculus, students likely still prioritized memorization over sense-making, and used technology more to find answers to problems than to fuel understanding of concepts.

4.4. Conceptual achievement

To measure students' understanding of calculus concepts, I used the Calculus Concept Inventory (CCI; Epstein, 2006, 2013). This is a 22-item pencil-and-paper exam administered at the beginning and the end of the term. None of the items on the CCI require any computations to answer correctly.

Both pre-term ($r = 0.3048$, $p < .001$) and post-term ($r = 0.3066$, $p < .001$) CCI score were significantly correlated with final exam score, as well as with each other ($r = 0.5125$, $p < .001$).

As a measure of growth in student understanding of calculus concepts, following Hake (1998) with modifications made by Simon, Kohanfars, Lee, Tamayo, and Cutts (2010), I computed normalized gain on the CCI by taking the ratio of actual gain (post-term – pre-term) to possible gain (maximum – pre-term); for students whose post-term score was lower than their pre-term score, I divided by their pre-term score instead (since this is the maximum possible loss). Figure 4.4.1, below, displays the mean post-term CCI score and normalized gain in each of the four classes. While students in the LDT class performed slightly better than students in the other classes on both these measures, these differences were not significant. This non-result was surprising, given the stated emphasis of the LDT instructor on developing conceptual understanding. In section 4.6.2, I discuss differences in student achievement on conceptual items on the final exam, and in Chapter 7 I examine some results that call into question the psychometric properties of the CCI instrument.

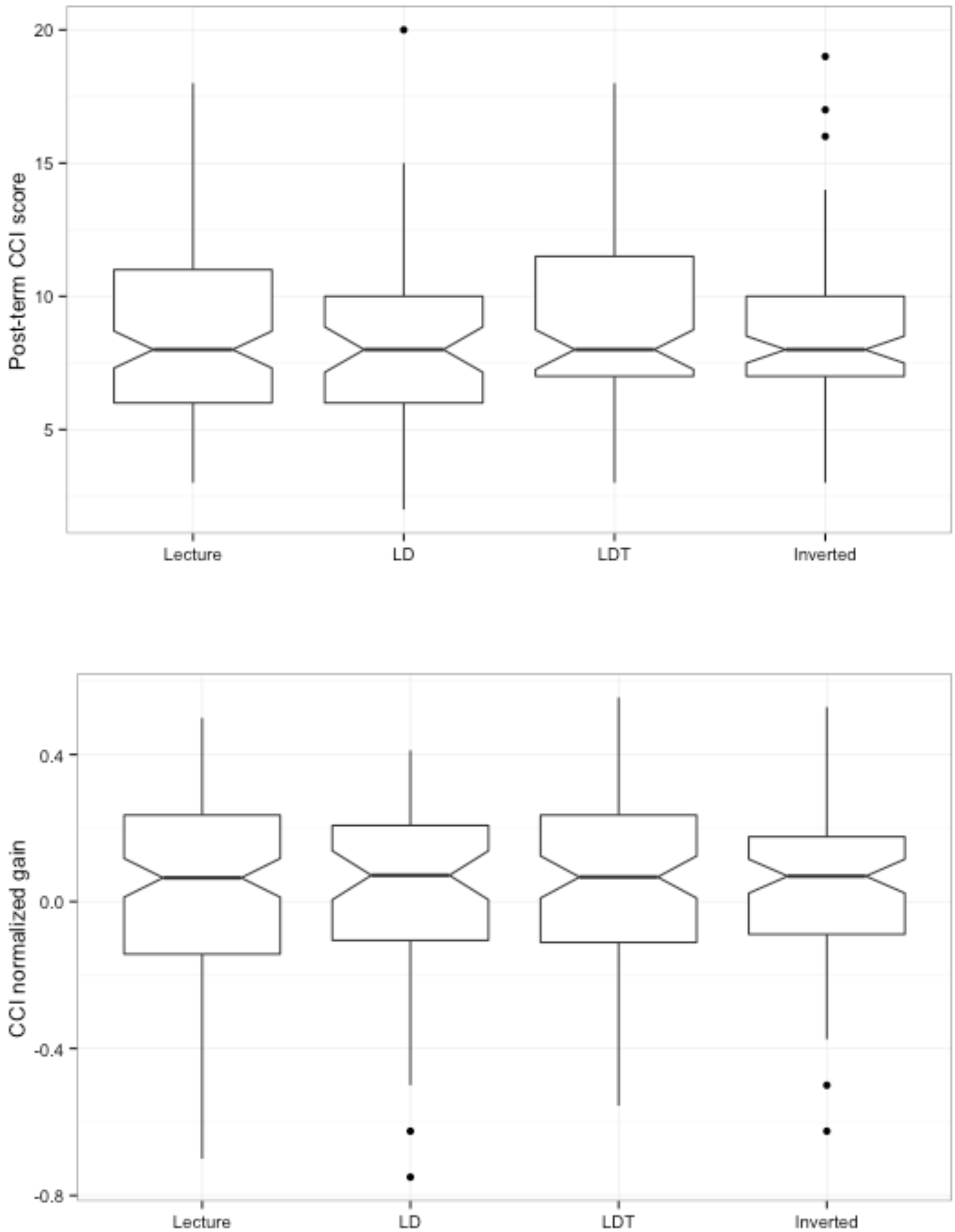

Figure 4.4.1. Mean post-term CCI score and normalized gain

I also conducted an item-level analysis of post-term CCI results. Chi-square analysis revealed that the proportion of correct answers differed significantly between the four classes on six of the 22 items, as displayed in Figure 4.4.2.

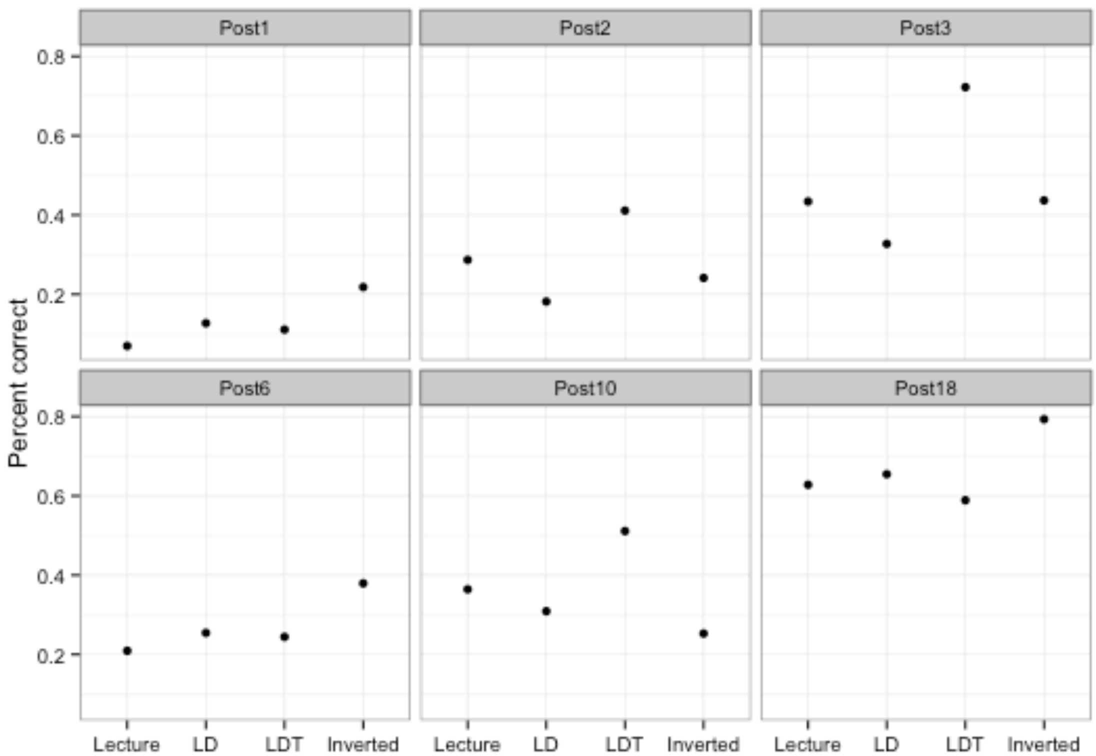

Figure 4.4.2. Post-term CCI items with significant differences

Items 2, 3, and 10 show the same pattern of differences: the LDT class outperformed the other three classes. These three items all assess conceptual understanding of the derivative; several of the applets developed by the instructor of the LDT class were specifically designed to scaffold intuition for the derivative, especially the meaning of the derivative as the slope of the tangent line.

Items 1, 6, and 18 also show a similar pattern, in which the Inverted class outperformed the other three. These items assess disparate concepts: Item 1 targets conceptions of limit, Item 6 is about an exponential function, and Item 18 involves an accumulation function. Preliminary IRT analysis revealed that Item 18 has poor discrimination; overall, 66% of students answered this item correctly.

4.5. Procedural achievement

ANOVA revealed significant differences between the classes in raw percentage scores on the common final exam ($F(3, 429) = 5.145$, $p = .002$). Post-hoc comparisons using Tukey's HSD test showed that Inverted-class students were outperformed by both LDT-class students ($MD = 7.09$, $p = .032$) and Lecture-class students ($MD = 7.58$, $p = .008$), and that Lecture-class students outperformed LD-class students ($MD = 7.25$, $p = .047$). The overall mean final exam score was 51.7% with a median of 53%.

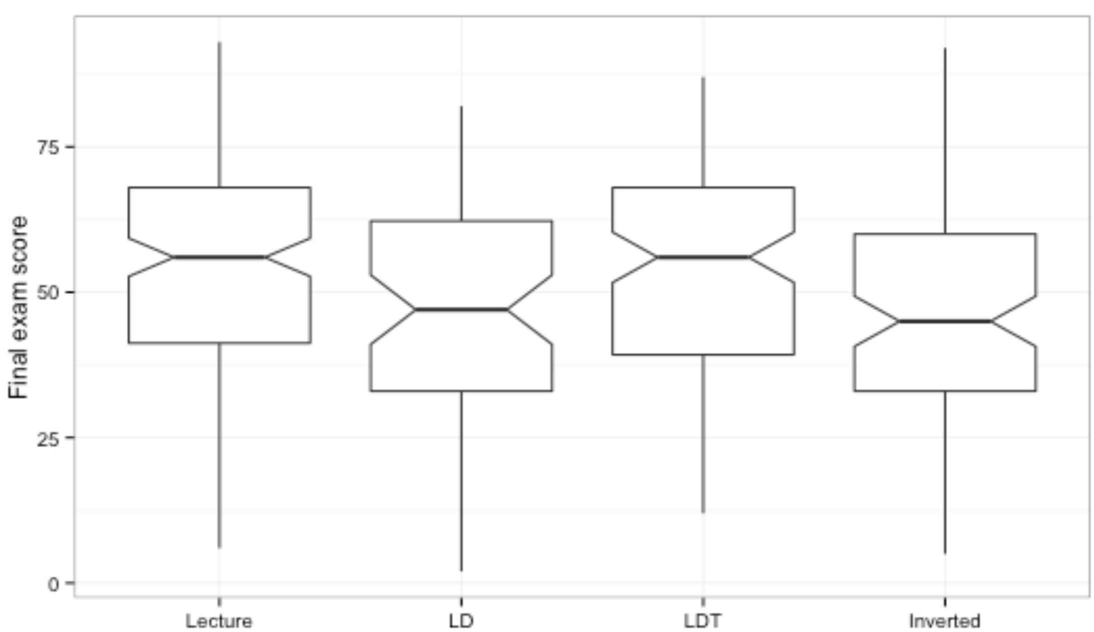

Figure 4.5.1. Mean final exam scores

The low scores on the final exam overall (recall that the overall mean score was 51.7%) were somewhat disconcerting. While this exam had a marked conceptual emphasis, due to the influence of the instructor of the LDT class, it was still a fairly standard Calculus I exam, and it would be reasonable to expect higher than a 51.7% average. To explain why the results were so uniform (and so uniformly low), it is necessary to examine the factors that were the same in the four classes.

One commonality between the four classes is the fact that this was a joint final. It could be argued that a joint final would artificially depress scores. The four different classes would inevitably place emphasis on different parts of the course; thus, students in one class may perform better than those in the other classes on one question, but worse on others. However, I find this explanation unpersuasive: since the final was jointly developed, the instructors would have caught things that unfairly favored the emphasis given in one class. Indeed, the instructor of the Inverted class felt that early drafts of the exam, developed by the LDT instructor, were too conceptual; since his class had less emphasis on conceptual understanding, he felt that his students would be at an unfair disadvantage on an overly-conceptual test. For all the instructors to agree to the final version of the exam, they must have felt that it was reasonably fair to all four classes.

A more persuasive argument can be mounted by considering the common curriculum mandated by the joint final and by departmental expectations of calculus classes. Perhaps final exam scores were so low because all four courses tried to cover too much material too fast, leading to poor learning outcomes across the entire curriculum. This is a common complaint (see, e.g., Seymour, 2006; Steen, 1987), and in section 6.7, I present evidence from the focus group interviews that students felt that their classes were rushed.

4.5.1. Controlling for incoming preparation

This analysis assumes that the incoming students are distributed more or less randomly between the four classes, and that the populations of the classes are more or less homogeneous. To allow testing of this assumption, I collected scores on the Calculus Concept Readiness instrument (CCR; Carlson, Madison, & West, 2010), which measures students' preparedness for calculus. CCR score was strongly correlated with final exam score (Pearson's $r = .408$, $p < .001$).

When comparing the mean scores on the CCR, I found significant differences (ANOVA $F(3, 424) = 10.134$, $p < .001$). Post-hoc comparisons using Tukey's HSD test showed that students in the Inverted class were significantly less prepared than students in the LDT class ($MD = 2.21$, $p < .001$) or in the Lecture class ($MD = 2.59$, $p < 0.01$). In addition, students in the LD class were significantly less prepared than those in the Lecture class ($MD = 2.13$, $p = .002$) or the LDT class ($MD = 2.21$, $p < .001$). These differences are summarized in figure 4.5.1.1.

Since there were significant differences in the incoming preparation of the students in the four classes, I used ANCOVA to compare final exam scores across the four classes while controlling for this variability, as measured by the CCR. The effect of class in this model was not significant ($F(3, 387) = 1.0224$, $p = .38$). This implies that more of the variance in students' final exam score was explained by their incoming preparation than by which class they were in.

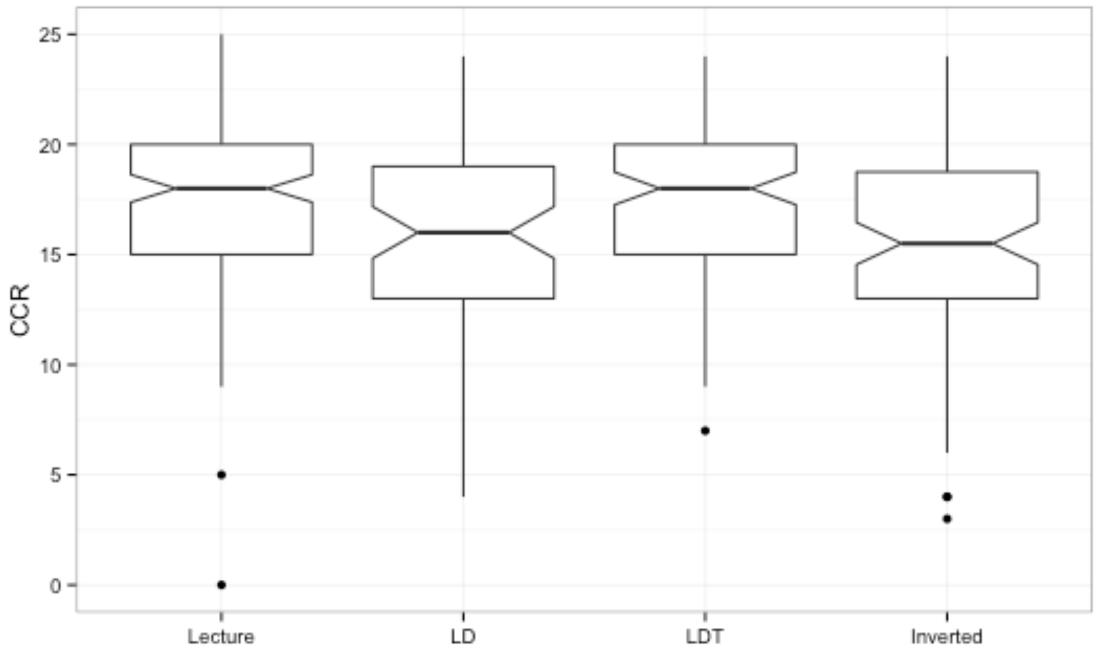

Figure 4.5.1.1. Incoming preparation, measured by CCR score

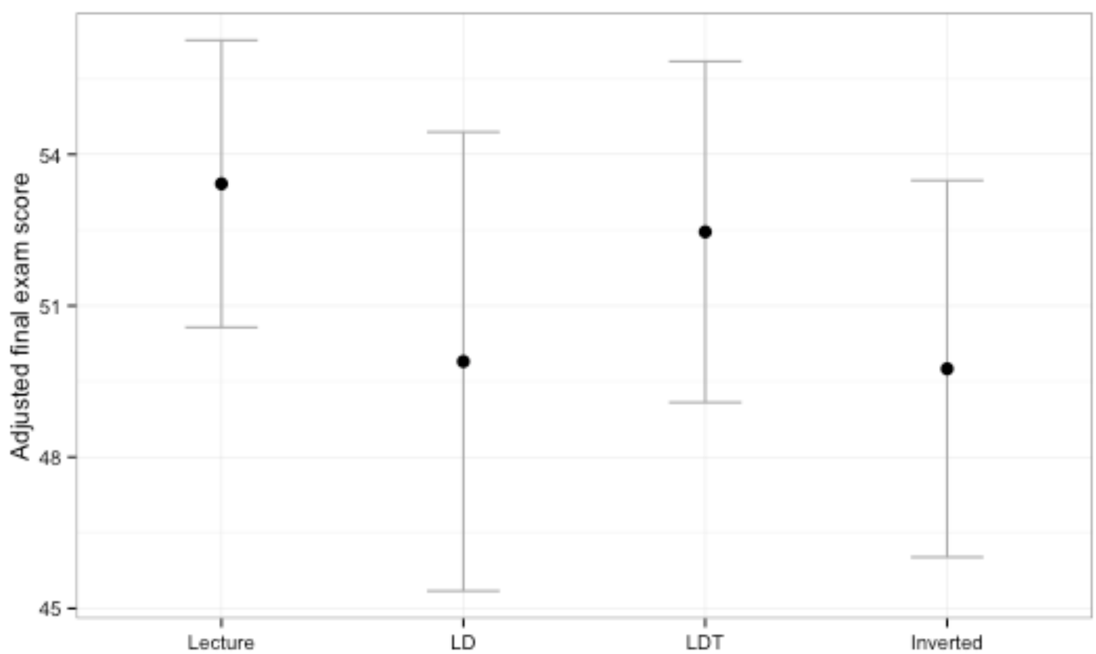

Figure 4.5.1.2. Adjusted mean final exam scores

After setting CCR to its mean score, the differences reported above between the mean final exam scores in the four classes remained; however, these differences failed to reach statistical significance. The plot in Figure 4.5.1.2 shows the adjusted means in each of the four classes, together with 95% confidence intervals around the adjusted means.

4.6. Conceptual and procedural items on final exam

Comparing students' raw final exam scores in total supposes that every final exam item measures approximately the same set of calculus skills. This is likely to be an oversimplification of the true picture; some items will assess procedural skill, while others will be more conceptually oriented. Following White & Mesa (2014), I classified each of the items on the final as "recall and apply" (7 items), "recognize and apply" (5 items), "understand" (5 items), or "apply understanding" (2 items). Another researcher independently coded the items, and we achieved 100% agreement after discussing items that were initially coded differently. Additionally, to increase the number of items in each category, I grouped together the "recall and apply" and "recognize and apply" items as procedural items (12 items, together worth 56 of the 100 points on the final), and the "understand" and "apply understanding" as conceptual items (7 items, together worth 44 of the 100 points on the final). Overall, students performed better on the procedural items (average score 54.3%) than the conceptual items (average score 49.7%). This difference was statistically significant, $t(845.817) = -3.1881$, $p = .001$. I then compared student performance in each of the four classes on these sub-scales. The results of these comparisons are summarized in sections 4.6.1 and 4.6.2, and are discussed together in section 4.6.3.

4.6.1. Procedural items

Taken together, the 12 procedural items were worth 56 of the 100 points available on the final. ANOVA revealed a significant effect of class on performance on the procedural items, $F(3, 425) = 7.377$, $p < .001$. Tukey post-hoc analysis revealed that the Lecture class outperformed both the Inverted class ($MD = 6.845$, $p < .001$) and the LDT class ($MD = 4.850$, $p = .006$). Breaking the procedural items into “recall and apply” and “recognize and apply” items did not provide additional information; the same pattern of differences emerged, so I do not present that analysis here. This analysis is summarized in the boxplot in Figure 4.6.1.1 below.

To control for students’ incoming preparation, I conducted an ANCOVA controlling for CCR score. The effect of class in this model was significant, $F(3, 383) = 5.5637$, $p < .001$. Tukey post-hoc tests of difference in adjusted means revealed a similar pattern of differences, with the Lecture class outperforming both the Inverted class (MD 4.627, $p = .022$) and the LDT class (MD 5.089, $p = .003$). Figure 4.6.1.2 shows the adjusted marginal means in the four classes, together with 95% confidence intervals.

It thus appears that the Lecture class was more effective in developing students’ purely procedural skills than either the LDT class or the Inverted class, even after controlling for students’ incoming preparation.

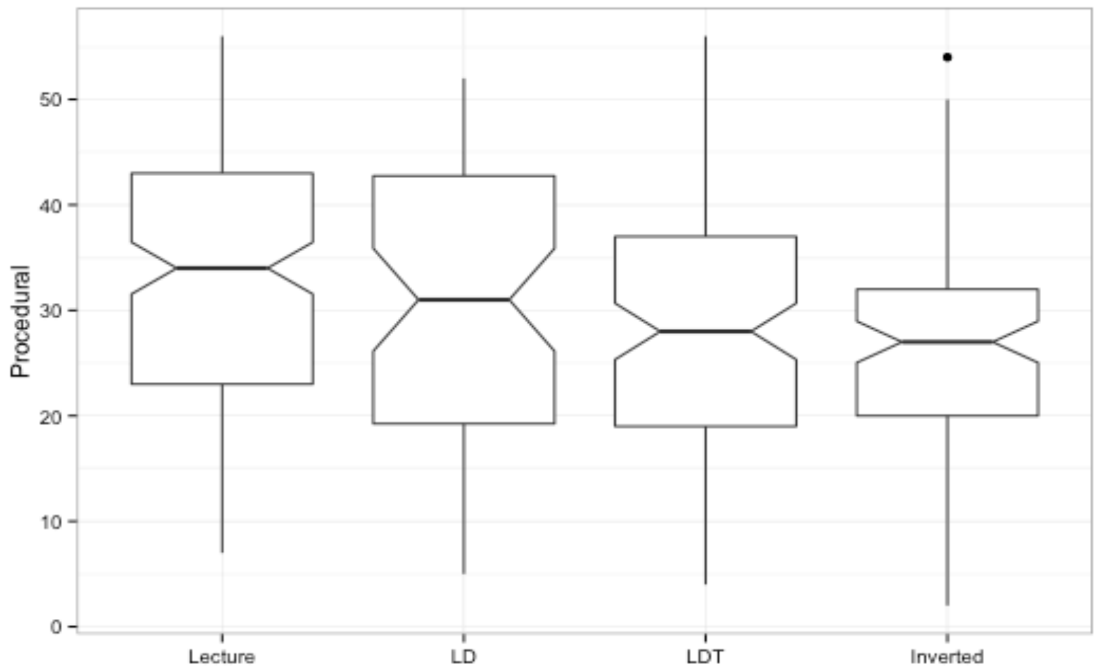

Figure 4.6.1.1. Boxplot of scores on procedural items

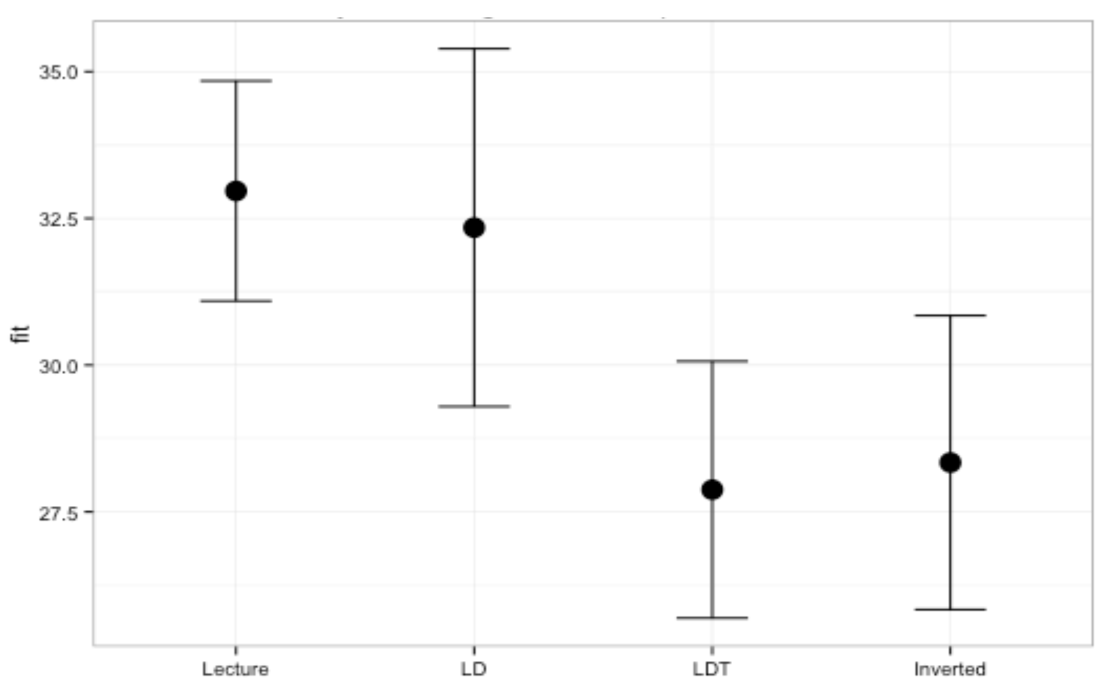

Figure 4.6.1.2. Adjusted marginal means on procedural items

4.6.2. Conceptual items

Taken together, the 7 conceptual items accounted for 44 of the 100 points on the final exam. ANOVA revealed a significant effect of class on conceptual score, $F(3, 425) = 7.9017$, $p < .001$. Tukey post-hoc analysis indicated that the LDT class outperformed all three other classes: the Lecture class (MD 3.626, $p = .003$), the LD class (MD 6.280, $p < .001$), and the Inverted class (MD 3.240, $p = .033$). The scores in each class are summarized in the boxplot in Figure 4.6.2.1.

I again used ANCOVA to control for students' incoming preparation, as measured by CCR score. The effect of class was significant once more, $F(3, 383) = 5.4929$, $p = .001$. Tukey post-hoc comparisons of adjusted marginal means revealed that the LDT class significantly outperformed both the Lecture class (MD 3.445, $p = .005$) and the LD class (MD 4.606, $p = .004$). While the adjusted mean score in the LDT class was higher than that in the Inverted class, this difference was not significant. This information is summarized in Figure 4.6.2.2.

In summary, the LDT class was more effective than any of the other classes in developing students' conceptual understanding, as measured by the conceptual items on the final exam. When controlling for incoming preparation, this was still the case for the Lecture and LD classes.

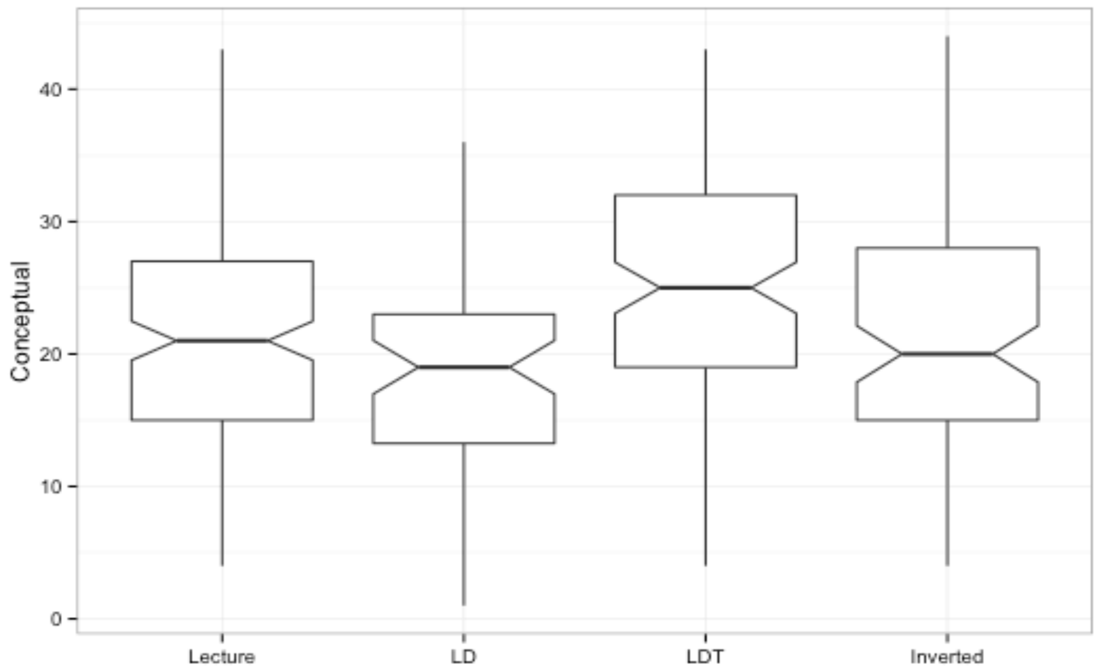

Figure 4.6.2.1. Boxplot of scores on conceptual items

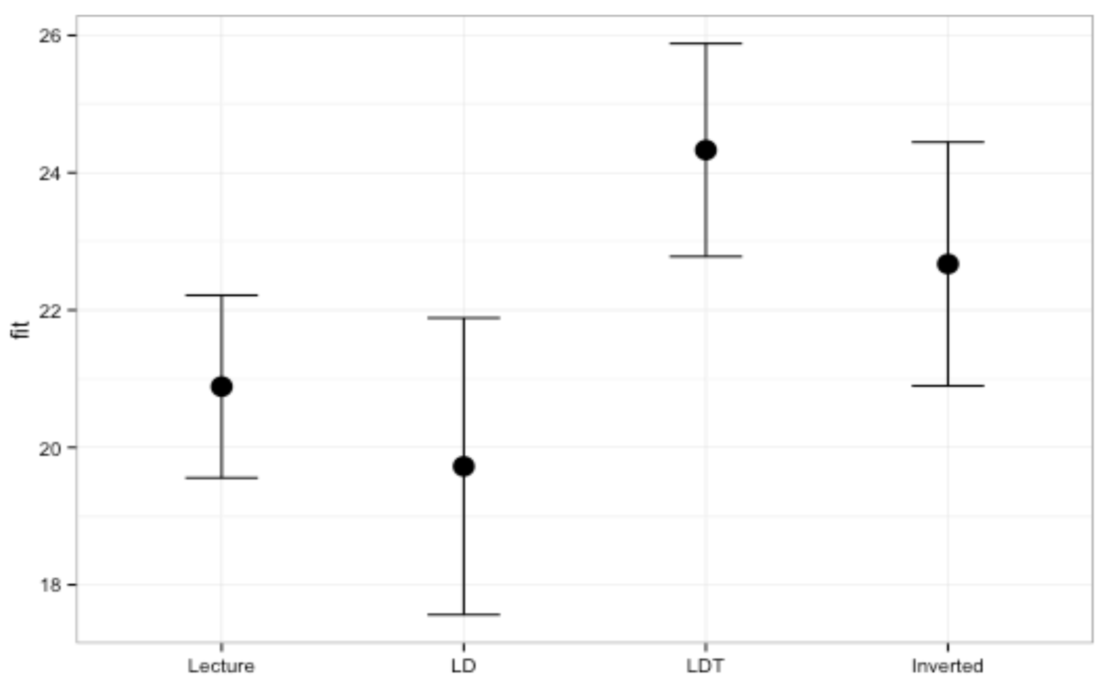

Figure 4.6.2.2. Adjusted marginal means on conceptual items

4.6.3. Discussion

As hypothesized, even when controlling for students' incoming preparation, students in the LDT section performed better on conceptual items than their counterparts in other classes. The instructor of the LDT section thus appeared to have attained her goal of promoting conceptual understanding by using Geometer's Sketchpad applets. However, this improvement in performance on conceptual items appeared to have come at the expense of performance on procedural items; students in the LDT section performed significantly worse on such items than students in the Lecture section. This could simply be because of the increase in time spent on conceptual understanding in the LDT class, and a concomitant decrease in time spent on developing mechanical skills. Section 6.2.1 presents qualitative evidence supporting this hypothesis.

Additionally, and similarly, students in the Inverted class performed significantly lower than those in the Lecture section on procedural items. When controlling for incoming preparation, this difference was erased for conceptual items; however, unlike the LDT section, Inverted students did not significantly outperform their peers on such items. One possible explanation for Inverted students' underperformance on procedural items is the lack of a unified viewpoint. For each class session, students had a wide variety of video and other resources to choose from, and thus were presented with a wide variety of procedural explanations. This lack of consistency between different presentations, especially in the absence of the strong central voice of the professor, may have inhibited the development of students' procedural skills.

Instead of framing these differences as the LDT or the Inverted section underperforming the Lecture section, it is useful to consider reasons why the Lecture section might outperform other sections. In particular, the instructor of the Lecture class tended to “hand-hold” through problems, showing each step in great detail, deep into the semester.

This result echoes earlier results on reform vs. traditional secondary school curricula. For instance, Huntley, Rasmussen, Villarubi, Sangtong, and Fey (2000) compared the effects of conventional and NCTM *Standards*-based curricula (in particular, the Core-Plus Mathematics Project; CPMP) on student performance in algebra. They found that students who learned from the conventional curriculum performed better on tasks assessing ability to carry out manipulations of symbolic expressions outside of contexts, while students in the CPMP classrooms had stronger skills in solving problems presented in realistic contexts.

Like the Huntley et al. (2000) study, the present study raises questions about what it means, or what it should mean, to understand mathematics, and what type of knowledge is or should be valued in the classroom. Proponents of conceptual knowledge, often aligned with reform movements, argue that because of the rapid proliferation and reduction in cost of computer algebra systems, it is less necessary now for students to develop strong symbolic-manipulation skills; those who value procedural knowledge, typically supporters of traditional curricula, “argue that automaticity of such skills is essential to problem solving and further mathematical learning" (Huntley et al., 2000, p. 357). Ultimately, this is a value question that cannot be settled by any one study.

Students' cognitive beliefs about mathematics are also relevant to this discussion. As will be seen in Chapter 5, students at this institution generally have less expert-like cognitive beliefs about mathematics than the national average. This means, for instance, that local students are more likely than average to believe that mathematics should be learned by memorization, and that the role of the instructor is to work problems so students know how to do them. So, the lecture-heavy, step-heavy style of the Lecture class is more in line with students' beliefs than the reasoning-heavy style of the LDT class. This helps explain why the LDT section's effect on students' conceptual-item scores does not translate to an effect on procedural-item scores: students at this institution are less likely to believe there is a link between conceptual understanding and procedural skill.

4.7. Differential impact

I conducted a number of two-way ANOVAs to assess the differential impact of the four classes on various different populations. I began by comparing the performance of males and females on the final exam in each of the four classes. The two-way ANOVA showed no statistically significant main effect of either gender or class; the interaction approached significance ($F(3, 236) = 1.97$, $p = .119$). A marginal means plot is displayed in Figure 4.7.1. Univariate contrasts revealed that, as suggested by the marginal means plot, males and females did not differ statistically significantly in any of the classes but the Inverted class, in which males outperformed females by 10.93 points ($t(36.97) = 1.77$, $p = .04$).

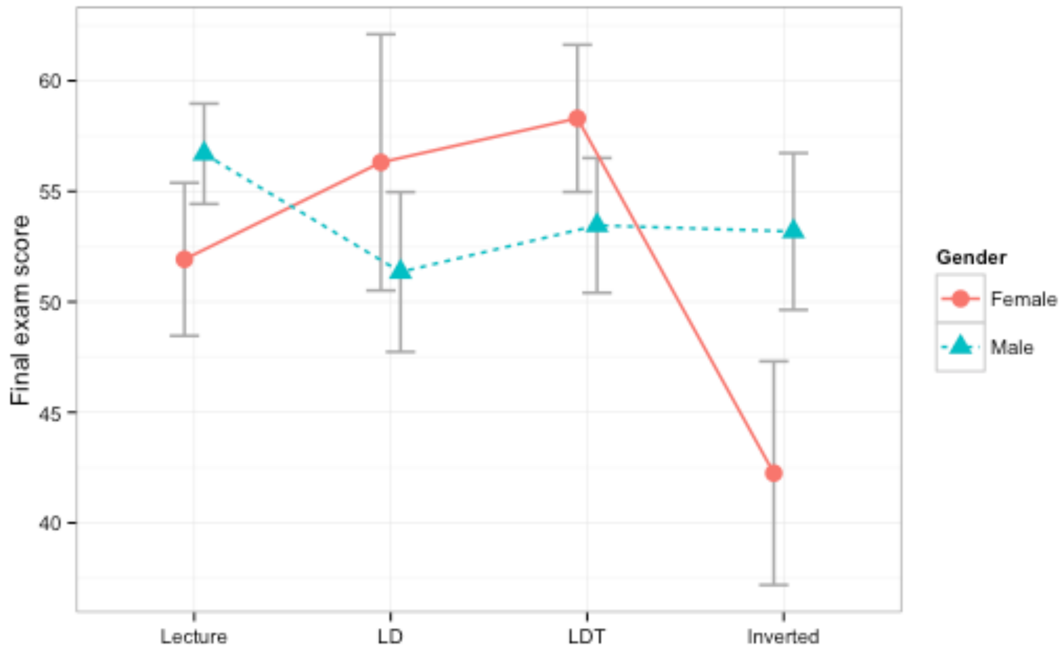

Figure 4.7.1. Comparison of final exam scores by gender

I also categorized students by their earliest prior calculus experience: high school (either AP or non-AP), college, or none. Two-way ANOVA revealed a significant main effect of prior calculus experience ($F(2, 233) = 20.67$, $p < .001$) and a significant interaction between prior calculus experience and class ($F(6, 233) = 2.63$, $p = .02$). The marginal means plot in Figure 4.7.2 suggests that students who took calculus in high school outperformed all other students in every class but the Inverted class; indeed, it is only in this class that the univariate contrast is not significant.

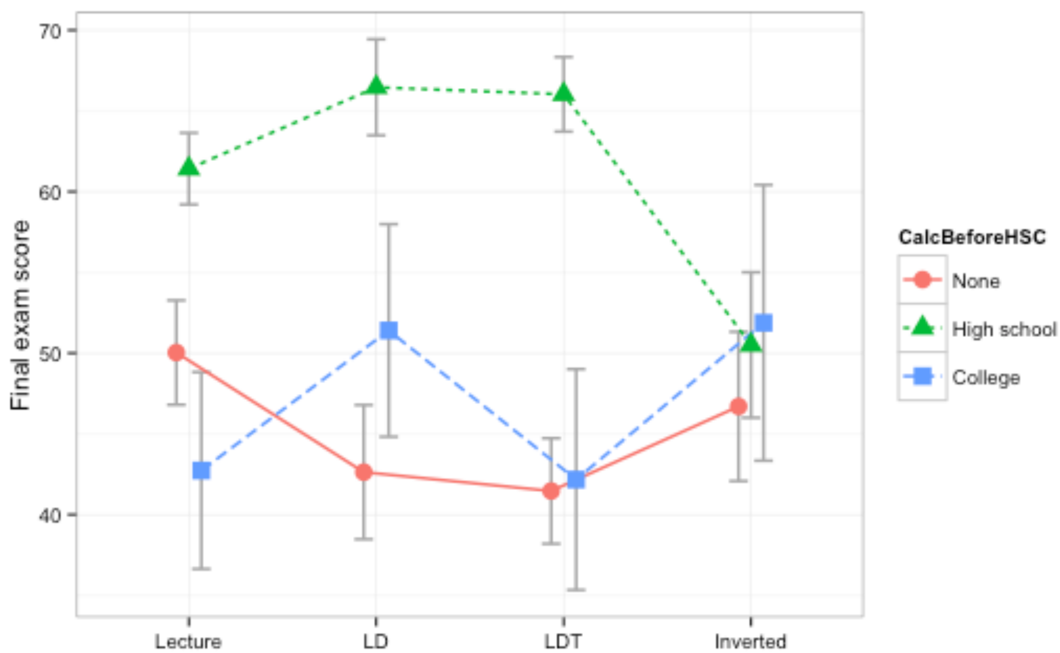

Figure 4.7.2. Comparison of final exam scores by prior calculus experience

Since the populations in the four classes were not homogeneous, I re-ran these comparisons while controlling for students' incoming preparation, as measured by CCR score. Figures 4.7.3 and 4.7.4, below, show the adjusted marginal means plot for prior calculus experience and gender, respectively.

As suggested by Figure 4.7.3, after controlling for CCR score, it is only in the Inverted class that students who took calculus in high school do not outperform students with other levels of prior calculus experience; ANOVA reports a significant difference between the groups in each of the other three classes. After controlling for CCR score, there is no significant difference in the performance of male and female students in any of the four classes, as suggested by the adjusted marginal means plot in Figure 4.7.4.

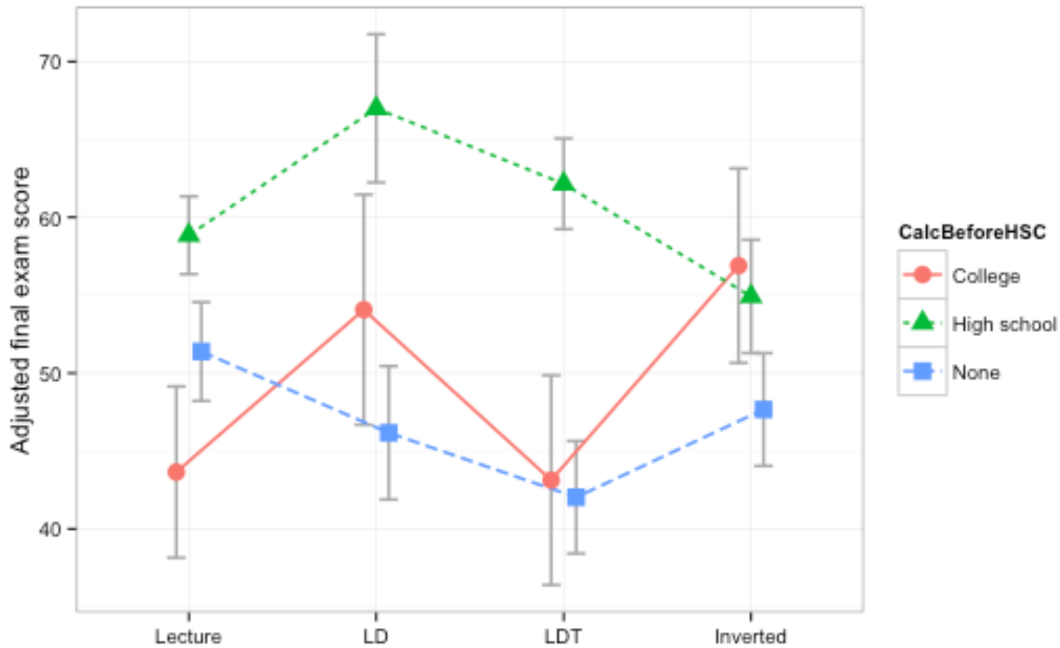

Figure 4.7.3. Comparison of adjusted marginal means by prior calculus experience

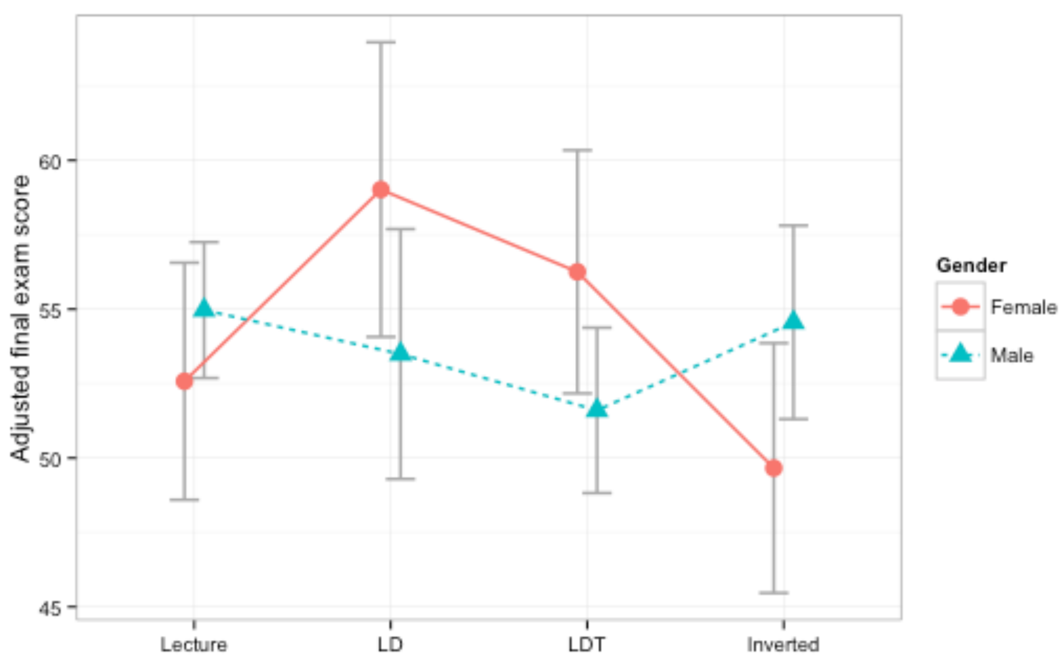

Figure 4.7.4. Comparison of adjusted marginal means by gender

4.7.1. Discussion

It is to be expected that students who took calculus in high school would outperform other students; in general, students who take calculus in high school are high-achieving students. An average student is more likely to finish their high school mathematical career with precalculus. Further, it is to be expected that students in the "college" group would not experience a marked advantage over students with no prior calculus experience, since these are students who are retaking calculus. It is thus likely that although they have been exposed to the ideas of calculus, they did not do well in their prior class.

So, given the above discussion, the picture in the three non-inverted classes is to be expected: students who took calculus in high school significantly outperform other students, and the other two groups are not statistically distinguishable. However, in the Inverted class, not only is the difference between the "high school" group and the other groups erased, but the advantage of the "high school" group is erased also. This is a surprising departure from the expected pattern demonstrated in the other three classes.

What is it about the Inverted class that erases the advantage of students with prior calculus experience? One explanation might be that students who took calculus in high school may feel that they do not need to complete pre-lecture activities, since they can bank on their prior knowledge. They thus perform lower than expected. The other students do not rely on their prior knowledge, and so complete pre-lecture activities; their performance is thus in line with the performance of comparable groups in other classes. In section 6.1.1, I present qualitative evidence supporting this hypothesis.

4.8. Discussion of local quantitative results

Synthesizing across the results presented in this chapter, two main themes present themselves. First, there are few statistically significant differences between the student outcomes. Second, when significant differences appear, the Inverted classroom most often finds itself negatively implicated.

4.8.1. Fewer differences than expected

Many of the tests I conducted found no significant difference between the four classes. There were no significant differences in switcher rate as calculated by the CSPCC methodology, ten of the fifteen ETS beliefs items, the end-of-term cognitive beliefs measure, end-of-term CCI score, normalized gain on the CCI, or final exam score after controlling for incoming preparation. Additionally, the classes performed uniformly poorly on the final exam.

These non-results are surprising because on the surface, the four classes seem to present entirely different learning opportunities to the students. If all the results are the same, then there must be some hidden similarities in the classes that inform the similarities in results. I am thus led to ask new questions: What are those similarities? In particular, why did all the classes perform poorly on the final exam?

4.8.2. The Inverted classroom

I present a brief summary of the results negatively implicating the Inverted classroom. First, using the roster data methodology, the Inverted classroom has higher than average switching rates. Second, students in the Inverted classroom demonstrate significantly less favorable beliefs about mathematics. Third, while the difference is not statistically significant, the mean final exam score was lower in the Inverted classroom than the LDT or the Lecture class, even when controlling for students' incoming preparation. Finally, in the Inverted class, students who took calculus in high school lose the advantage they have in other classes.

These negative results, and the lack of any positive results, about the Inverted class are surprising. Results in the literature about Inverted classrooms are uniformly positive, so the uniformly negative results here are a radical departure. In Inverted classrooms in the literature, students performed as well or better than in other classes, and reported enjoying their class more than traditional lecture-based classes they had taken before. Naturally, the literature is biased toward success reports, but this striking departure leads me to ask new questions: How does this unsuccessful implementation of the Inverted model differ from the successful implementations reported in the literature? What are the commonalities of successful models that are lacking in this implementation?

4.8.3. New questions

The analyses described here in chapter 4 left me with a number of new questions: What similarities existed between the four classes? Why did all four classes perform poorly on the final exam? How does this Inverted class differ from the successful ones in the literature? What do the successful implementations have in common that this implementation lacks? These are questions that cannot be answered with quantitative data. In chapter 6, I will use the qualitative data I gleaned from classroom observations and focus group interviews to propose answers to these questions.

Chapter 5: National Quantitative Comparisons

In this chapter, I examine how the local institution compares to other institutions across the country. This chapter focuses on my second research question: How do students in the local institution compare to students in the national database in:

a) their persistence in STEM major tracks, and
b) their attitudes and beliefs about mathematics?

To answer this question, I compared the local data to the data collected by the CSPCC researchers. I compared to four subsets of the national data: the overall national data considered in aggregate, the institutions selected for further study by the CSPCC researchers, and master's-granting and Ph.D.-granting institutions. I compared to both master's-granting and Ph.D.-granting institutions because the local institution shares many characteristics with both. The highest in-house terminal degrees in the mathematical sciences are master's degrees, and the university is listed in AMS's Group M, which means that it would be considered a master's-granting institution by the criteria used by the CSPCC researchers. However, while the local institution cannot offer a terminal Ph.D. on its own, due to limitations of its charter, it partners with several nearby universities to offer joint Ph.D.'s in various branches of the mathematical sciences. Further, with an enrollment of approximately 26,000, it compares in size with many Research I institutions. Thus, in many respects, it resembles a Ph.D.-granting institution.

I begin this chapter with a comparison of the student demographics in the local and the national institutions. First, I examine levels of parental education, a measure of socioeconomic status. On the STS, students were asked to indicate their father's (or male guardian's) and mother's (or female guardian's) highest level of education. I combined these two variables to obtain a measure of the highest level of parental education. Table 5.1 reports the percentage of students in each category reporting each level of parental education. As seen in Table 5.1 and Figure 5.2, the parents of students at the local institution had lower levels of education than in the national database.

Table 5.1. Comparison of highest level of parental education

	Local	National	Selected	Masters'	Ph.D.
Did not finish high school	8.70%	2.10%	2.40%	2.30%	1.70%
High school	18.20%	8.40%	8.40%	10.50%	7.30%
Some college	27.30%	17.60%	17.20%	22.70%	16.00%
Four years of college	29.20%	33.10%	30.30%	33.40%	33.30%
Graduate school	16.70%	38.70%	41.70%	31.10%	41.60%

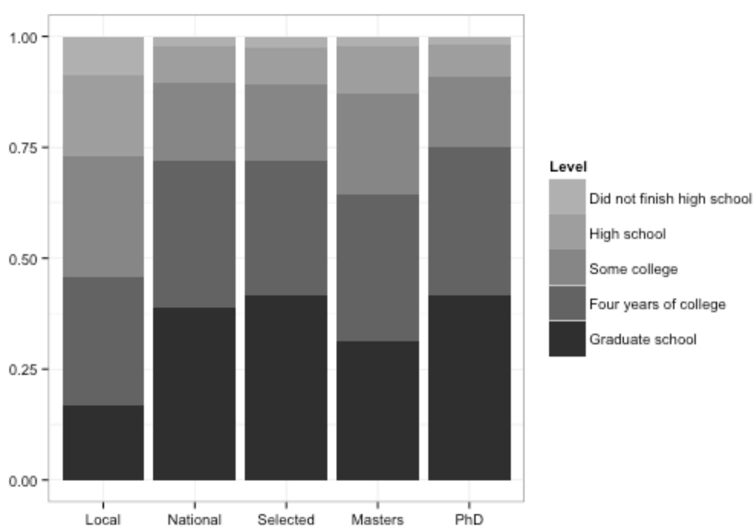

Figure 5.2. Stacked bar chart of highest level of parental education

As indicated in Table 5.3, there was a higher proportion of males in calculus at the local institution than in the national data.

Table 5.3. Comparison of gender

	Local	National	Selected	Masters'	Ph.D.
Male	67.8%	55.9%	55.5%	55.6%	54.4%
Female	32.2%	44.1%	44.5%	44.4%	45.6%

The ethnic makeup of the local student body was significantly more diverse than in the national database, as reported in Table 5.4. Notably, there was a much larger Hispanic population at the local institution; the local institution is a Hispanic-serving institution (HSI) and a member of the Hispanic Association of Colleges and Universities. Percentages were calculated as a fraction of the number of students who provided a response to any race/ethnicity-related survey question. Percentages may not sum to 100% due to students selecting multiple racial categories.

As reported in Table 5.5, students at the local institution had lower SAT math scores than students in the national database.

A similar proportion of students at the local institution and in the national database were taking calculus for the first time in the semester under study. These proportions are given in Table 5.6. To phrase it differently, these were the proportions of students who had not previously taken any course called calculus, whether in college or high school.

Table 5.4. Comparison of race/ethnicity

	Local	National	Selected	Masters'	Ph.D.
White	48.5%	75.9%	74.5%	77.5%	75.9%
Black	2.3%	5.4%	3.7%	7.3%	4.5%
Asian	27.0%	13.9%	16.0%	9.3%	15.5%
Pacific Islander	6.0%	0.9%	0.9%	0.6%	0.8%
American Indian	1.9%	1.5%	1.5%	3.5%	1.6%
Hispanic	32.3%	9.8%	9.7%	5.8%	9.3%

Table 5.5. Comparison of SAT math scores

	Local	National	Selected	Masters'	Ph.D.
Average SAT Math score	590	652	651	618	662

Table 5.6. Comparison of first-time calculus takers

	Local	National	Selected	Masters'	Ph.D.
Proportion of first-time calculus students	39.6%	35.2%	36.4%	45.5%	29.3%

5.1. Persistence

The primary reason I used the CSPCC methodology to determine switching rate was so that I could compare directly to switching rates in the national sample determined using the same methodology. As summarized in Table 5.1.1, the local switching rate was lower than the switching rate in each of the four groups examined in the national data.

Table 5.1.1. Comparison of switching rates in local and national data

	Local	National	Selected	Master's	Ph.D.
Switching rate	6.6%	12.5%	14.6%	17.3%	13.5%

I conducted four chi-square analyses to compare the switcher rate in the local data to that in each of the four subsets of the national data. In each case, the difference was statistically significant, indicating that the local switcher rate was significantly lower than the rate in the national data. These tests are summarized in Table 5.1.2.

Table 5.1.2. Chi-square tests of significance in difference in switcher rates

Other group	$\chi^2(1)$	p
Overall national data	7.8642	.005 *
Selected institutions	11.7972	< .001 **
Master's institutions	14.2102	< .001 **
Ph.D. institutions	10.1589	.001 *

5.1.2. Discussion

The large difference in switching rates between the national sample and the local institution requires explanation. This difference can be accounted for by examining the populations who enroll in mainstream calculus at the local institution and in the national sample. At the local institution, engineers and computer scientists make up a much larger proportion of the students than in the national database. Of the 267 students who provided a career goal, 172 (64.4%) marked engineering or computer science; in the national database, engineers and computer scientists made up 35.9% of the sample. These students switch at a much lower rate than students with other career goals in the national database. Table 5.1.2.1 shows the switching rates for engineers and computer scientists in the local and national data. The switching rates are almost identical, and indeed, there is no significant difference between the proportions of switchers and persisters ($\chi^2(1) = .1016$, $p = .750$).

Table 5.1.2.1. Comparison of engineer/CS switching rates in local and national data

	Local	National
Switching rate	5.8%	6.9%

Why, then, are engineers and computer scientists more persistent than students in other career tracks? While further research should be devoted to this question, one way to answer this question is to look for start-of-term differences between engineers and other students. Engineering is a mathematics-heavy discipline, and it is likely that students who choose to major in engineering have some affinity or proficiency in mathematics. In the national database, students in engineering tracks score significantly higher on start-of-term measures of both affective (STS.Affective: MD 0.227, $t(8830.358) = -17.997$, $p < .001$) and cognitive (STS.Cognitive: MD 0.122, $t(8095.724) = -11.594$, $p < .001$) beliefs about mathematics; more expert-like beliefs about mathematics are correlated with persistence into Calculus II (Rasmussen & Ellis, 2013). Further, while this was not the case in the national database, Zhang, Thorndyke, Carter, Anderson, and Ohland (2003) found that engineering students in a database from nine southeastern universities had significantly higher SAT math scores than other students.

Another factor decreasing the switching rate at the local institution is the existence of a somewhat unusual non-mainstream calculus class for life scientists. Most biology, pre-medical, and pre-nursing students at the local institution take a course entitled Calculus for the Life Sciences, rather than the mainstream calculus sequence for mathematics, engineering, and physical science majors. In the national database, students whose career goals are in the medical, health, or life sciences fields make up 29.8% of the sample, and these students switch at a rate of 23.1%. The removal of these students from mainstream calculus lowers the switching rate at the local institution.

5.2. Beliefs and attitudes about mathematics

To see how local students' beliefs about mathematics compared to those of students in the national database, I conducted a parallel analysis to that reported in section 4.3.2. Using the same conceptual groupings of STS and ETS items reported in that section, I created aggregate scores which measured students' affective and cognitive beliefs about mathematics. First, I converted the Likert-scale responses to each of the beliefs items into z-scores, and negated the z-scores of reverse-coded items. I then calculated the aggregate affective beliefs score by taking the average of the z-scores on the items in the affective beliefs cluster, and calculated the aggregate cognitive beliefs score similarly.

Once these scores had been created, I compared the mean scores in the local group to the mean scores in each of the other groups. As summarized in Table 5.2.1, the local group scored lower on each of the four measures than each of the other four groups, and all but three of those differences were statistically significant. The three differences that failed to reach statistical significance were all on the STS affective measure. The means of each group on each of the four measures are visualized in the boxplots in Figures 5.2.2 and 5.2.3.

Students at the local institution start the term with beliefs scores lower than the average in any of the comparison groups in the national data; their scores are significantly lower in the case of cognitive beliefs. This is likely indicative of demographic effects. Several possibilities for the source of this effect are discussed in section 5.2.3.

At the end of the semester, students at the local institution scored lower than students in any of the other samples. There are several possibilities: both local and national beliefs may have increased over the semester, with local beliefs increasing less; local beliefs may have decreased while national beliefs increased; or both local and national beliefs may have decreased, with local beliefs decreasing more. Since the variables for start-of-term and end-of-term beliefs include different questions, they cannot be compared directly to determine which of these possibilities was the case. In section 5.2.2, I analyze pairs of items that were worded identically on the STS and ETS to assess how beliefs changed over the semester.

Table 5.2.1. Summary of t-tests comparing local scores to national scores

Dependent variable	Other group	MD: local – other	Welch two-sample t-test	p
STS.Affective	Overall national data	-0.0672	t(279.796) = -1.6570	.099
	Selected institutions	-0.0550	t(326.468) = -1.3032	.193
	Master’s institutions	-0.1330	t(519.146) = -2.7512	.006 *
	Ph.D. institutions	-0.0744	t(285.468) = -1.8237	.069
STS.Cognitive	Overall national data	-0.1402	t(279.320) = -4.2512	< .001 **
	Selected institutions	-0.1468	t(325.314) = -4.2855	< .001 **
	Master’s institutions	-0.1875	t(496.429) = -4.8402	< .001 **
	Ph.D. institutions	-0.1299	t(285.375) = -3.9186	< .001 **
ETS.Affective	Overall national data	-0.1224	t(369.354) = -3.3352	< .001 **
	Selected institutions	-0.1597	t(530.194) = -3.9553	< .001 **
	Master’s institutions	-0.1841	t(696.513) = -3.6266	< .001 **
	Ph.D. institutions	-0.1151	t(395.677) = -3.0820	.002 *
ETS.Cognitive	Overall national data	-0.1318	t(363.883) = -4.1874	< .001 **
	Selected institutions	-0.1532	t(504.516) = -4.4678	< .001 **
	Master’s institutions	-0.1880	t(694.703) = -4.3525	< .001 **
	Ph.D. institutions	-0.1095	t(388.521) = -3.4221	< .001 **

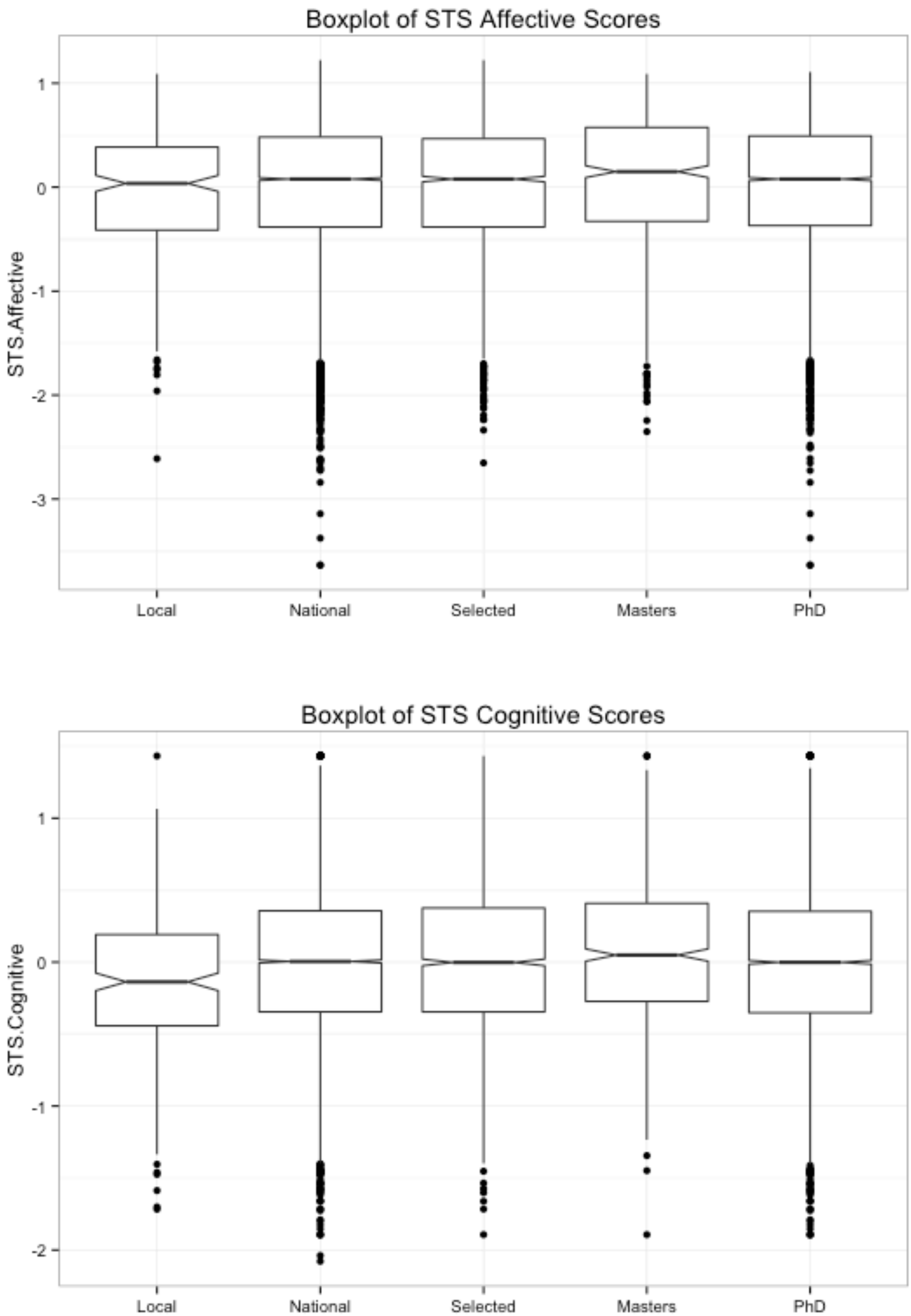

Figure 5.2.2. Boxplots of STS affective and cognitive scores

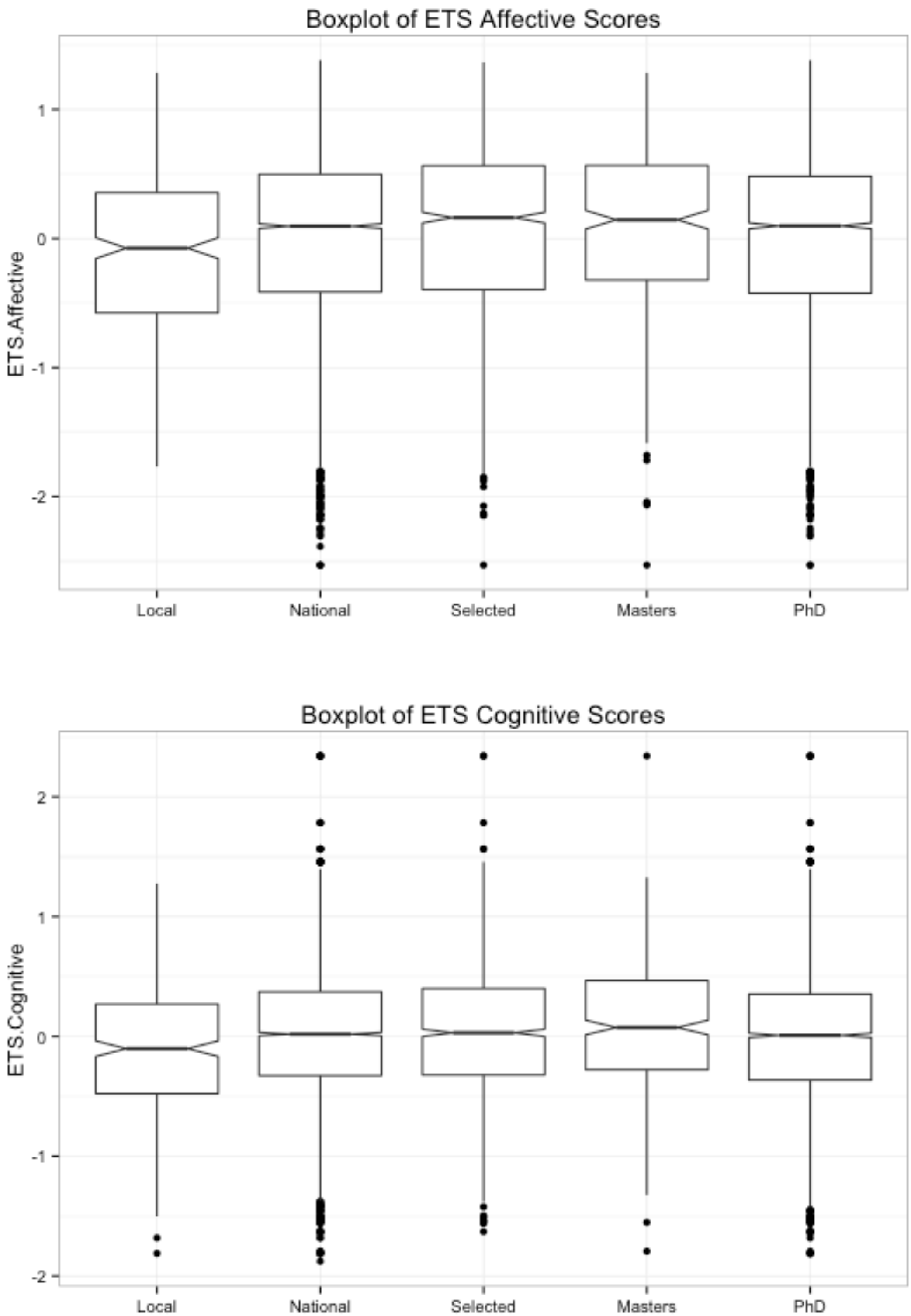

Figure 5.2.3. Boxplots of ETS affective and cognitive scores

5.2.1. Controlling for incoming beliefs

I used ANCOVA to control for incoming beliefs while comparing beliefs scores from the local institution to those from each of the other four groups of interest. As summarized in Table 5.2.1.1, both affective and cognitive scores at the local institution were significantly lower than those in each of the other groups, except for affective beliefs at the Ph.D. institutions (and this difference approached significance).

Table 5.2.1.1. Summary of ANCOVA comparisons of local and national groups

Dependent variable	Other group	MD: local – other	F statistic	p
Affective	Overall national data	-0.0950	F(1, 3673) = 6.8148	.009 *
	Selected institutions	-0.0910	F(1, 1035) = 4.9321	.027 *
	Master's institutions	-0.1582	F(1, 442) = 10.494	.001 *
	Ph.D. institutions	-0.0693	F(1, 2240) = 3.3934	.066
Cognitive	Overall national data	-0.0857	F(1, 3670) = 7.9428	.005 *
	Selected institutions	-0.0731	F(1, 1035) = 4.8837	.027 *
	Master's institutions	-0.1395	F(1, 441) = 10.485	.001 *
	Ph.D. institutions	-0.0643	F(1, 2237) = 4.4239	.036 *

These results suggest that even when controlling for students' incoming beliefs, local students still score lower on measures of beliefs and attitudes about mathematics than students in any of the four national comparison groups. Combining this result with the result from Chapter 4 that the only significant between-class effects at the local institution involved the inverted classroom, this implies that there are likely to be either demographic effects that were not controlled for when controlling for start-of-term-beliefs, or effects at the institutional level that are less successful in terms of promoting student beliefs. Possibilities for these factors are discussed in section 5.2.3.

5.2.2. Paired items

To develop a more detailed understanding of how beliefs shifted over the semester, I examined students' responses to the items that were identical on the STS and ETS, and compared the change in these items for each of the five groups of interest. Pre-term and post-term scores for the paired items are displayed in Figure 5.2.2.1.

Care should be taken when examining this figure. Several of these variables were measured on a four-point scale, while others were measured on a six-point scale, so it did not make sense to use the same scale for each facet. I thus chose to focus each facet on the range of data in that variable, so the scales differ in each facet. In each facet, however, higher numbers indicate more favorable or more expert-like beliefs about mathematics.

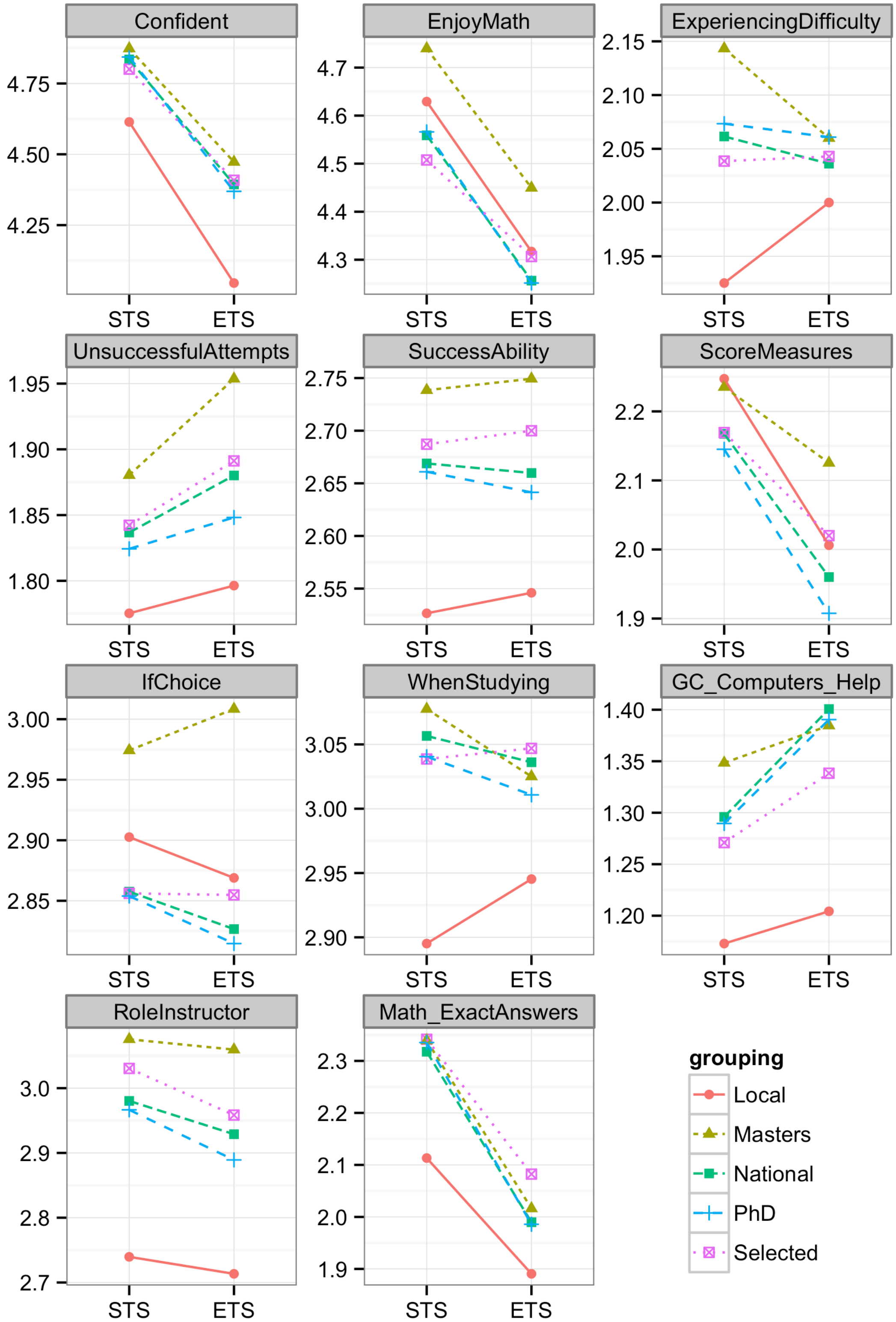

Figure 5.2.2.1. Comparison of trends on paired items

In all but two instances (WhenStudying and ExperiencingDifficulty), the local data displays roughly the same trend as each group of the national data. In the two exceptional instances, I used a t-test to determine if the difference in STS and ETS scores was statistically significant; neither one was (ExperiencingDifficulty: $MD = .075$, $t(577.501) = -1.1351$, $p = .257$; WhenStudying: $MD = .050$, $t(576.711) = -0.6164$, $p = .538$). In other words, the slope of the line connecting the STS mean and the ETS mean did not differ significantly from zero.

The overall implication of this plot is that the trends in the paired items were quite similar at the local institution and in the national database. In other words, beliefs at the local institution did not shift appreciably more or less than those at institutions in the national database. However, triangulating the results from section 5.2, beliefs at the local institution were lower in most instances than those in any of the other groups of interest, at both the start and end of the term.

5.2.3. Discussion

Overall, students at the local institution had somewhat lower beliefs than the national average, even at the beginning of the term. This suggests demographic effects: there may be particular features of the population of students who enroll in calculus at the local institution that are associated with less expert-like beliefs and attitudes about mathematics. As reported at the beginning of this chapter, the population of students in the local classes was more heavily male and more diverse than the national average, and showed lower levels of parental education, a marker of lower socioeconomic status. Additionally, the average SAT math score was lower at the local institution. The analysis presented in section 5.2.1 used ANCOVA to control for students' incoming beliefs when examining students' end-of-term beliefs. However, incoming beliefs are just one plausible predictor of outgoing beliefs; the demographic effects discussed in the paragraphs above are also likely predictors. Future work could use multiple regression techniques to examine the links between these demographic factors and beliefs and attitudes about mathematics.

There may also be institution-level factors that impact student beliefs. The local institution is located in a large city, is a Hispanic-serving institution (HSI) and a member of the Hispanic Association of Colleges and Universities. Additionally, the local institution is a member of the California State University (CSU) system; according to the California Master Plan for Higher Education, the University of California (UC) system selects from among the top one-eighth of college-bound high school graduates, while the CSU system selects from among the top one-third. Thus, since the top students in California are more likely to enroll at a UC than at a CSU, it is likely that the average student at the local institution is less academically prepared than the average student at a similar institution in a different state. This hypothesis is supported by the lower average SAT math score at the local institution. Further work could compare the local institution to institutions with similar institutional characteristics, including other large urban institutions, other HSIs, and other large CSUs to determine the association of various institutional characteristics with students' incoming beliefs scores.

The institutional factors discussed in the prior paragraph are ones that interact with demographic factors: for instance, because the local institution is an HSI, the ethnic makeup of the student body is skewed toward Hispanic students. In surveying almost twenty-five years' worth of research on how college affects students, Pascarella and Terenzini (1991) found that the differential impacts of attending different kinds of colleges were substantially less pronounced than the net effects of attending college at all. While they did not examine beliefs and attitudes about mathematics specifically, this general result likely holds true in this case. Pascarella and Terenzini speculated that this lack of impact might be due to the "essentially conservative" estimates of institutional impact produced by controlling for student characteristics: "any variance *jointly* due to the effects of students' backgrounds and institutional characteristics is attributed entirely to student differences" (p. 589, emphasis in original). In other words, it is difficult to disentangle the effect of students' backgrounds from the effect of other institutional characteristics.

Another avenue for future work is to employ different methods to measure students' beliefs and attitudes about mathematics. Ambrose, Clement, Philipp, and Chauvot (2004) argued that measuring beliefs is difficult, since beliefs are held with different intensities, are context-specific, and must be inferred. They pointed out three problems with using Likert scales, such as those used in the present study, to measure beliefs: it is difficult to know how respondents interpret the form of words used in items; they provide no information for determining the strength or importance of the issue to the respondents; and they do not provide contexts. This third point is particularly important: since beliefs are context-bound, asking the same question framed in two different contexts might elicit two very different answers. Ambrose et al. thus developed a survey in which respondents constructed responses instead of choosing from provided options, and developed rubrics for quantifying these responses. While this approach is much more labor-intensive than an approach involving Likert scales, and thus may not be appropriate for a large-scale national study, it yields a more valid measure of respondents' beliefs.

Pursuing this line of inquiry further, future work might conduct one-on-one interviews with students that focus on the effect of their calculus classes on their beliefs and attitudes about mathematics. While the focus group interviews employed in this study (discussed in chapter 6) provide some insight into these effects, it is not their primary focus, and the amount of information that can be gleaned from these interviews about students' beliefs is limited.

Chapter 6: Analysis of Qualitative Data

When I began this study, I hypothesized that the active learning pedagogies would outperform the more traditional Lecture class. This hypothesis was largely not borne out by the quantitative data. I thus turned to the qualitative data to develop explanations for the quantitative data. In this chapter I present my analyses of the focus group interviews, course evaluations, and classroom observations.

I conducted four focus group interviews: one interview with a convenience sample of students from each of the four classes. Each was a semi-structured interview following a protocol developed by CSPCC researchers (Appendix B), and was approximately one hour long. The Inverted class interview had 5 participants; the LDT class had 7; the LD class, 4; and the Lecture class, 10.

I also observed the class sessions of each of the four classes in which related rates and the fundamental theorem of calculus were discussed. I observed four class sessions of the Lecture and LDT classes, and two class sessions of the LD and Inverted classes; in total, I observed approximately 200 minutes of class time in each class.

All names in this chapter, both of students and of instructors, are pseudonyms.

6.1. Problems with this implementation of the inverted model

After surveying the literature on the inverted model, my hypothesis was that the outcomes of students in the Inverted classroom would be better than those in other classes (and, in particular, better than the Lecture classroom). As discussed in Chapter 4, this was not the case; in general, the students in the Inverted classroom performed no better than those in other classes, and, in fact, certain groups of students performed worse. This surprising quantitative finding drove me to seek explanations in the qualitative data, particularly the focus group interviews. I found that there were substantial differences between the successful inverted classes in the literature and the Inverted class in my study.

I also identified three categories of student concerns and dissatisfaction with this implementation of the inverted model, and found that these categories intersected substantially with the set of departures from the reports in the literature. I thus identified these departures as "lethal mutations" (Brown & Campione, 1996) and propose the beginnings of a list of best practices for implementing the inverted model.

To understand the comments the students made in the focus group interviews, it is important to first draw a picture of a typical day in the Inverted classroom. The professor in the Inverted class summarized his approach to the inverted model in general terms on his syllabus: "In the inverted model, students begin their learning at home via a variety of resources, then complete their learning in class by training on exercises." To complement this general description with a detailed account of the daily activity of this class, I examined the syllabus and course website, conducted classroom observations, spoke with the three teaching assistants (TAs) assigned to the course, and asked students in the focus group to describe a typical day in class.

Several days before each class session, the professor posted links on his website to videos and other resources discussing the material that would be covered in class. The videos came from a variety of sources, including Hippocampus, Khan Academy, PatrickJMT, and MIT's OpenCourseWare collection. Additionally, the professor commonly provided some text-based resources from Wikipedia and online textbooks such as Strang (1991).

Students were expected to prepare for class by watching videos or reading text materials. Students were free to choose which resources to use; in general, in order to accommodate a wide variety of student learning styles, more resources were provided for a given day's lesson than any individual student would use. There was no specific mechanism to check whether or not students had watched videos or read materials before class.

As students entered the classroom, they signed in on an attendance sheet, and received a worksheet described in the syllabus as containing "a sequence of increasingly challenging exercises." There were usually between 10 and 20 exercises on a worksheet. The entirety of class time, a 100-minute block, was spent working on the problems on the worksheet. Most students chose to work in self-assigned groups of four to six, while a few generally preferred to work by themselves.

Except for the first day and the days on which exams were administered, the professor did not attend class. Instead, three TAs were assigned to attend class and answer student questions. The TAs did not feel that they could stop class to hold a brief mini-lecture, even if they noticed that a substantial number of students all had the same question. Thus, the atmosphere of the classroom was more like a tutoring lab than a classroom with one central authority.

Near the end of each class session, the teaching assistants would announce which of the problems on the worksheet would be collected and graded. Students would recreate their work on that problem and turn it in before leaving. The problem would be graded and returned to them in a later class session.

In focus group interviews, the students in the inverted class were uniformly and vociferously dissatisfied with the implementation of the inverted model. Data from the course evaluations triangulated the general dissatisfaction expressed in the focus groups: of the 36 students who responded to the open-ended comment prompt, 33 left comments negatively evaluating the inverted model. Students' comments in the focus group and the course evaluations clustered into three main categories: problems with the pre-class videos, problems with the in-class activities, and a feeling of disconnect from the professor.

6.1.1. Problems with the pre-class videos

Early in the focus group interview, a student named Sarah said, "I feel like that's the biggest problem in this class, is the videos are not applicable to the work." This was a common complaint in both the focus group and the course evaluations. Five of the 36 open-ended comments on the course evaluations addressed this concern: for instance, one student said that "the videos did not always match the class problems", and another said that "the videos that we used to learn the material had little to nothing to do with the material being taught in class."

An overlapping category of concern, found in five of the course evaluations, was that the instructor did not create the class videos himself. One student wrote that "it would have been really good if the material came from the professor," and recommended that "next time the inverted method is used to teach the class again, the teacher should make his videos for each chapter as well."

Data from the focus group interview triangulates these concerns. One student named Bob was in a unique position to assess the efficacy of the Inverted class, because his high school statistics class used a similar inverted model where "we watched videos prior to coming to class, and we did worksheets during class." This seemed to have been a positive experience. However, in his high school class, "it was nice because the teacher devoted a lot of time, because he created the videos himself, and ... went over the videos in class and stuff," which was not the case in the Inverted class. Bob continued, "I mean, [the instructor of the Inverted class] is a professor, he has a lot on his plate, and maybe he doesn't have the time to do that. But if he were to make the time in terms of creating the videos himself and shaping the videos towards his class, I think it would be more beneficial than just pulling random [internet videos.]" When I asked the students what other resources they would have liked to have in their class (a standard question in the focus group interview protocol), Bob replied, "I wish he had videos that he made himself, or that had more direct correlation to the class we were taking, ... covering all the concepts on the worksheets that he put. ... I mean, he can put random problems, but have the video apply to that thing so that we can watch the video [and] come to class feeling like we're actually prepared."

Bob felt that the videos did not adequately prepare students for the work in the class, and that it was thus necessary to have or find some extra knowledge:

> Bob: So I mean, what he's giving us [on the worksheet] one day is different than what the videos are, and there's no relation -- I mean, there's some relation of course, because we're able to do some of them, but there's

> still some of them we've never even learned of, and the videos are just not clear. So we either have to know the information from a prior class, [or] you have to do a Google search yourself and find some help for it, or you just ask the TA or go in for help. ... In order to succeed in the class, you have to take the extra step. But if ... you don't have time to go to the extra thing, you have a full schedule, what are you going to do?

Bob was not the only focus group interviewee to express these concerns. Phoebe made similar comments about the lack of cohesion between the videos and the worksheets: "The way this professor puts his problems doesn't correspond with the way we're taught on the videos [from] Hippocampus, or Wikipedia [articles] he puts on there, just like random stuff. It just isn't cohesive."

The feeling that the videos did not prepare students for the in-class worksheets had adverse effects on students' confidence:

> Melissa: I watched the videos and I understand it going in, I feel very confident, and then I get that paper [the in-class worksheet] and I'm like, well, I give up already.
> Sarah: I feel the same.
> Int: So you feel like you understand things after watching the videos, but then...
> Melissa: It just doesn't relate to the worksheets, yep.
> Phoebe: We're getting a good understanding of calculus watching the videos, but just not the way he wants it done. That's where it gets confusing.

Earlier in the interview, Sarah expressed similar concerns: "So I sit there, like I do all the easy ones, and I'm just like, well, I could ask the TA how to do every single problem, because I've never seen it. But then you just kind of sit there and you give up a little bit. It's just like I don't know what I'm doing." Additionally, one student on the course evaluations said, "I felt that I was all right at math before I took this class."

All these problems with the pre-class videos led to a general feeling that they were not useful, and some students abandoned them altogether.

> I: Do you feel like on average, the videos have been useful for you to learn calculus?
> Sarah: No.
> Melissa: No.
> Bob: To be honest, I haven't watched very many of the videos.
> Sarah: I've watched videos before every class, and they haven't been helpful.
> Bob: I've watched videos maybe two to three times out of the whole semester. I mean, most of what I remember is from high school.
> Phoebe: Yeah, what's keeping me going in this class is ... math classes in high school I took.

This corroborates the finding reported in Chapter 4 that in the Inverted class, unlike in the other classes, students who took calculus in high school did not perform better than other students; it seems that these students were less likely to watch the videos, relying instead on the knowledge they obtained in their high school classes. Recall also that there was no specific mechanism for holding students accountable for watching the videos.

When I asked students what they did to prepare for tests, none of the students reported watching the videos again. This is in stark contrast to the findings in Moravec et al. (2010), whose students reported that the learn-before-lecture activities were particularly useful for reviewing before tests (see also Lage et al., 2000). However, when I asked students in the focus group interview what they liked about videos, they gave answers that were consistent with those given by students in the literature (see, e.g., Bergmann & Samms, 2008; Gannod, 2007; Gannod et al., 2008):

> Bob: You can do it on your own time.
> Sarah: You can pause and rewind, that's what I like. It's the best thing for me.

Additionally, one of the course evaluations (indeed, the only one to give a specific positive comment about the videos) said, "I liked that we were able to learn the mathematical processes in a variety of ways."

6.1.1.1. Discussion

The general thread of students' comments about the pre-class videos is that because the pre-class videos were not made by the professor, they were not applicable to the in-class work the professor required. Thus, students felt that the videos did not adequately prepare them to complete the in-class worksheets. This was a source of frustration, because the videos failed in their express purpose. Students commonly directed this frustration toward the professor, who they seemed to regard as having abrogated his responsibility to prepare them for the work in the class. These latter concerns are examined further in section 6.1.3, which discusses students' feelings of disconnect from the instructor.

Brousseau's (1997; see also Herbst & Kilpatrick, 1999) construct of the didactical contract is a useful way to understand the students' frustration. The didactical contract in a typical classroom would likely include the following: the instructor's responsibility to the students is to teach them the material, and the students' responsibility to the instructor is to do the work assigned by the instructor. Students can expect that if they complete the assigned work, they will be well-prepared for the assessments that will follow. In the Inverted class, students felt that they were ill-prepared for the worksheets despite having watched the assigned videos; they thus felt that the instructor had breached the didactical contract by failing to create or select adequate videos.

It does not appear that the students felt it was required to have all the videos made by the professor. As Phoebe observed in the focus group interview, "at the beginning of the year, the videos corresponded well, because it was just a lot of simple stuff."

Variation in presentation seems to be less of an issue for foundational early material in a course; further work could turn empirical attention to the question of which topics are specific enough to how a particular instructor teaches to require purpose-built materials.

6.1.2. Problems with in-class activities

In addition to the failure of the in-class worksheets and videos to articulate well, students identified several other concerns with the in-class activities. One concern was the lack of structure in class time:

> Sarah: An hour and forty minutes straight of doing word problems is kind of a lot, at least for me. I don't know, I can't just...
> Phoebe: I get a headache.
> Melissa: Me too.
> Bob: I get distracted all the time.

Another concern was that the problems on the worksheet became too difficult too rapidly:

> Ben: I feel like he has too many problems that go way too deep into the concept. I mean, he'll start out basic, like let's just say it was 2 + 2 = 4. And then by the end it'll be, if all you're trying to do is learn addition, by the end, he'll have 2 times this times this plus this, just so you can get the concept of adding. He'll have sine squared, squared, to the third, or something like that. It's just too far, and it just takes up too much time.
> Bob: I mean maybe that's the level of calculus for college that we need to be at, and that's completely understandable, but let's work to it [laughs] rather than just going from, hey, simple sine is cos, and then jumping to what's sine cos sine to the fifth or something.
> Sarah: Zero to a hundred.

These concerns are likely related to each other (and to the absence of the instructor, as discussed in section 6.1.3). With stronger scaffolding from the instructor or TAs, and thus more structure provided to class time, exercises increasing in difficulty would be less problematic. This was borne out in the next line of transcript:

> Melissa: And then when we get to those problems, we all need help, because all two hundred of us or whatever don't understand that level, but there's only three TAs. So we'll sit there and they'll [say,] "I'm gonna start the problem, but then I gotta walk away and start it for somebody else."

Giving TAs the authority to conduct mini-lectures, and thus impose more structure on class time, would likely have ameliorated this concern. If the TAs saw or expected that many students would have the same question about the same problem, allowing them to explain the question to the entire class at once would have been more efficient and likely more effective.

One further concern that was expressed in both the course evaluations and in the focus group interviews was a perceived disconnect between the in-class worksheets and the test. One student wrote that "the tests pertained only vaguely to the homework we practiced in class, and curve balls were always put into the problem on the test, things that we had not focused on in our homework." Another wrote that "the tests were so much harder than what the homework showed."

The focus group interview protocol contained a question designed to assess the frequency of novel problems. When I asked this question of the Inverted students, Bob indicated that he felt the tests were entirely like this.

> Int: How often are you asked to solve problems that you've never seen before?
> Bob: You mean like the tests? [laughs]

Later, he elaborated:

> Bob: I was with a TA earlier today, working on the past test, and there was a problem that we'd never seen before, in terms of the format of the problem. We'd seen things like finding the tangent line of something given a point. Well, we knew the concept of how to get a tangent line, but we didn't know, we'd never seen a problem where it's within a range of equations. ... I guess how he does his things is he takes concepts and ideas

> that we've done, and then applies it to a new way and gives it to us. It's a good way to test what we know, but it's still kind of a little screwed up that we've never seen problems in that format.

6.1.2.1. Discussion

One of the affordances of the inverted model identified by Talbert (2014) is that students who have difficulty managing their time outside of class are at a disadvantage in the traditional classroom, since "higher-level cognitive tasks often require extensive periods of time for work and reflection; these segments of time are often mismanaged or are simply unavailable to many students." He argues that the inverted model, in which high-level tasks are done in class "where the instructor is present to guide students in efficient and effective work," removes this disadvantage (p. 362). In this Inverted class, students were presented with 100-minute blocks of time designated for working on problems; however, these blocks were not further structured by the instructor or TAs. Thus, while students in the Inverted class at least had time set aside for working on problems, the difficulties of managing that time effectively were still present. Again, this connects to the next category of student concern, that of feeling disconnected from the professor, as will be discussed in section 6.1.3.

This is one of the most obvious departures of this Inverted class from the inverted models reported in the literature; approaches in the literature are much more structured by the involvement of the professor. For instance, Lage et al. (2000) engaged their students in highly-structured economic experiments, while Moravec et al. (2010) used clicker questions and class demonstrations to structure their class time. It is likely that these highly-structured activities contributed to students reporting that the inverted approaches were more enjoyable than traditional lecture formats (Lage et al., 2000). A large block of unstructured problem-working time, as found in the Inverted class, is likely less interesting and less motivating for students.

One student, Sarah, complained further that "there's so many people around." One common point of skepticism about the inverted class is that not all students prefer group work to working solo. Indeed, as reported by the students in the focus group, there were students in this Inverted class who chose to solve the problems on their own instead of working with a group of students around them. However, in the case of the Inverted class, the feeling that there were "so many people around" is likely to be a symptom of the lack of structure in class time. Studies on inverted classrooms in which the researchers asked students how they felt about their class experience consistently report that students enjoyed working in groups (e.g., Lage et al., 2000; Talbert, 2013).

6.1.3. Disconnect from the professor

The most common complaint by far on the course evaluations, mentioned by 24 of the 36 students who provided an open-ended comment, was that the instructor did not come to class sessions. The tone of these comments was usually angry, and there was often a perception that the instructor did not do anything, and thus did not hold up his end of the didactical contract. For instance, one student wrote, "I understand this was an experiment course where tutors would be in the classroom helping, but I probably saw the professor 3 times the whole semester. No one in the class could even remember his name. I believe if we are paying so much for tuition, we deserve an actual professor in the classroom more than 3 times in 4 months." Overlapping with concerns about the pre-class videos, another student wrote that the professor "managed to not teach us a single thing and rely purely on online lectures that weren't even taught by him to teach the class." Another student asked, "What is he getting paid to do?" Perhaps the most positive comment came from a student who wrote, "The inverted learning system is effective but I think we can't completely disregard the value of the professor teaching the class. Maybe this is what office hours are for, but I recommend that at the very least, [the professor] uses the review days before exams to work out problems on the board that students have. This would have helped me a lot."

These concerns were echoed by students in the focus group interviews. The first time this theme came up was when I asked a follow-up question about the videos:

> Int: So most of the videos that you see, are they made by your instructor, or just chosen by him from other sources?
> Bob: No, he chooses all of them.
> Phoebe: He never makes them.
> Ben: He doesn't make any of them.
> Bob: He just references them out to different online sources.
> Phoebe: We've only seen him once.
> Sarah: Twice.
> Bob: Yeah, we've seen him on the first day, we've seen him on the test. The first test. I don't think he was there the second test though.
> Ben: No, he was.
> Bob: He was? Okay. Three times, he's showed up. I mean -- [laughs] I know he's busy, but --
> Sarah: Yeah, it's kind of ridiculous, to be honest. I mean, the TAs are there to help us, but it would be nice to talk to a professor. Like during class, if he was there.

Overlapping with concerns about the in-class worksheets, the students reported that sometimes TAs were unable to answer questions about problems on the worksheet:

> Ben: I was talking with one of the TAs once when we were doing homework, and [sometimes] ... they'll be like, "I don't necessarily know how to do this."
> Phoebe: A lot of the time, if I ask them questions, they're like, "I actually

> don't know how to do that, let me ask another TA." Because they look at it and they're confused themselves.
> Ben: It's not that they're uneducated, they just --
> Phoebe: It's hard.
> Ben: Yeah, he just puts random stuff on there a lot of the time.
> Sarah: I went in to have help with a specific problem once, and the TA did it like five times and we still didn't get the right answer, and then I -- he just never figured it out.

Melissa also complained that there were questions about the professor's expectations that the TAs were unable to answer: "Since he's not there, when we ask the TAs what does he want with this problem, they say 'I don't know,' because no one knows what he's thinking. And so we're like, is this type of problem -- how should we set this up, or do this? And they're just like, 'I don't know what to tell you.'"

The fact that the instructor did not come to class was a violation of students' expectations founded in their high school experience:

> Bob: I'm a freshman coming in, first semester, it seems a little weird. I know college is entirely different, professors are doing research, and they have to do a certain amount of things to stay on campus, but it seems a little strange going from a teacher where every day [they're] helping you out, there for help, and then you have no professor. [laughs]

One of the standard questions on the focus group interview protocol assessed students' use of instructors' office hours. Sarah responded, "I don't know my professor's name." Later in the interview, Phoebe added, "I feel like it would be weird if we went to them, because we don't know him." Melissa agreed, and said, "I feel like it would be really awkward."

Students also felt that there was a disconnect between the grading on the in-class worksheets, which was done by the TAs, and on the exams, which was done by the professor:

> Sarah: But then there's the problem with there's different grading between the homework and the tests.
> Bob: I know the homework is graded by the TAs, but when I've asked the TA about a question on the test, they weren't sure. ... When I brought up a, I was like, "why did I get a 7 on this instead of a 9," they were looking through it and it seemed logical to them, but because maybe I was missing a piece of work.
> Sarah: See that's also another flaw of the class is, I get my homework back and I see the corrections or whatever, and I -- or, you know, I know that I did it right, and so that's -- I'm going by how the TAs are grading me, and then I'm not being graded by the same people.
> Phoebe: But then on the test you'll get like a 6, on something you would get like a 10 on during class.

Later, Bob said that this concern could be ameliorated with greater involvement from the instructor:

> Bob: What would be nice, though -- since the exams are the level of calculus that we need to be at -- if he comes in for a half an hour, and goes through the test himself and how he graded it and what he was looking for, then maybe we would actually understand okay, what he's looking for.

Students felt like this approach negatively impacted their understanding of calculus:

> Sarah: I could have done so much better in a different class.
> Phoebe: I have friends in another [calculus class] and they have a professor who teaches them and like...
> Ben: They say they're really good, too.
> Phoebe: Yeah, really good professors, really understand, and then there's me, and I'm like, I get it to an extent, but then I feel ... like in a way, behind. Like I'm nervous for Calc 2.

Later in the interview, Phoebe said that not having the professor present decreased her enjoyment of calculus as well:

> Phoebe: Math is my favorite subject. Since I was in elementary school, I was like, I love math! But now like this semester, I kinda sit there sometimes and I'm like, why don't I get this? Because math is my class where I get this, it's easy to me, but when we don't have a professor, it's -- I kinda sit there and [say], oh no. It's kind of discouraging, I guess.

Near the end of the focus group interview protocol, I asked several standard questions about the attitude of the instructor toward students. The responses I received in the Inverted interview were generally negative:

> Int: What would you say is your instructor's attitude towards students?
> Ben: I feel like we're kind of a nuisance to him. Just because of the fact that he only comes for tests, and when he does, he's really short.
> Bob: I don't know. He didn't leave a very good impression the first day, like he was -- It almost seemed like he was being rude to people.

A later question pursued the same theme and received similar responses:

> Int: Does your teacher seem to care about your learning in this class?
> Phoebe: Well, we don't know --
> Bob: I don't know this guy. I mean he might care.
> Sarah: I grew up around professors, and a lot of the time you know they're here for the research aspect. So I feel - I don't know if he does research, but I feel like that's quite possible, that he doesn't actually care about teaching.
> Melissa: That's what I feel like. I feel like he's at this school just to use its technologies and get ahead in his own personal research.
> Sarah: Yeah, that he's not here for the students.

These views parallel those reported by Seymour (2006), who conducted exit interviews with students completing STEM degrees as well as those who had changed their major. Both groups of students reported taking classes from unavailable, disinterested faculty with an implicit or explicit dislike for teaching.

Further, as examined later in Chapter 6, these views were in stark contrast to the views found in the other focus groups. In each of the other three classes, students all reported feeling that their instructors cared about their learning.

6.1.3.1. Discussion

Again, Brousseau's (1997) construct of the didactical contract is a useful way to discuss students' frustration with their absent instructor. Since the instructor did not create the videos, and since he did not attend class sessions, the perception of the students was that he did not do anything. In other words, he did not hold up his end of the didactical contract. Several students in the focus group pointed out that they felt underprepared for Calculus II, or that they enjoyed mathematics less after having taken this class than they had before, and laid the blame for these feelings at the professor's feet. Many had attended class and watched videos as assigned, but did not feel adequately prepared since the professor was not involved.

Additionally, the TAs, who assumed the instructor's role as authority figures in the classroom, were not given enough authority or enough training to fill that role effectively. They were unable to answer some of the students' content-related questions because they had not seen the worksheets prior to class. Some of these concerns could be resolved by the instructor meeting with the TAs prior to class to discuss each of the problems, or having the TAs work through the worksheet prior to class. Also, as discussed in the previous section, giving the TAs more authority to structure class sessions would likely help. Widening the scope, TA training programs have been shown to increase student success (Ellis, 2014); meetings with the instructor before each class session could be the nucleus of a more comprehensive training program.

However, there are some questions, particularly those having to do with the expectations of the professor, that TAs are inevitably unable to answer, no matter how proficient in calculus or how well-trained they are. For these questions, answers must come from the instructor of record; his absence, and the accompanying unanswerable questions, was thus a source of frustration for both the TAs and the students.

The lack of an instructor in the classroom is a radical departure from students' experience in high school. This disruption of the typical classroom structure is particularly evident to incoming freshmen. While it is unreasonable to expect students' experiences in college to be identical to their prior experience in high school, students are not prepared for a shift in structure of this magnitude.

The physical absence of the professor led to a feeling of disconnect. Students did not utilize the professor's office hours, because they felt that they did not know him well enough; "it would be really awkward," one student said. They doubted that he cared about their learning, viewed him as disinterested and unavailable, and felt that he saw them as "a nuisance." This led to a profusion of negative feelings about the instructor and the class in general.

When I began this study, having read many success reports in the literature, my hypothesis was that the Inverted class would perform well in various measures of student success. When this hypothesis was not borne out by the quantitative data, I turned to the qualitative data to seek an explanation for the lack of results. The instructor's absence from the class, another large departure from studies in the literature, is a plausible explanation. One of the main objectives of the inverted model is to get students in the same room as the more-knowledgeable other (Vygotsky, 1976) when working on the tasks with the greatest cognitive demand, so that assistance can be provided when it is most needed (Gannod, Burge, & Helmick, 2008; Talbert, 2014). If the instructor is not in the room with the students, this strength of the inverted model is lost.

To summarize, to achieve the full benefit of the inverted model, the instructor must be present and actively involved. Indeed, as Talbert (2014) writes, "open lines of communication between the instructor and the students are critical to the success of the inverted classroom" (p. 365). The failure of this iteration of the inverted model is an example of what can happen when the lines of communication are closed off.

6.2. Features of the LDT class

I begin this section by providing a description of a typical day in Julie's classroom. There were three one-hour sessions of class per week. As students walked into the class, a one-question poll would be projected on the screen. The content of the poll question varied from day to day; it might be a simple conceptual question, a poll on the difficulty of the homework, or an assessment of how well students felt they understood the material covered in the prior session. The poll would run for a few minutes after the official beginning of class time. Students were free to work with their classmates to determine their answer to the poll. Answers were submitted via text message or, for those students without cellphones, via small notecards. The poll functioned both as a warm-up question and as a method of checking attendance.

Class time typically began with Julie lecturing briefly to introduce the day's topic. During this portion of class, Julie commonly asked students questions, and welcomed questions from students. Students seemed to feel comfortable asking questions. This lecture portion might also include Julie working several examples on the board to demonstrate procedures that students would attempt later. Julie commonly used Geometer's Sketchpad applets to illustrate the geometric aspects of concepts underlying various problems. Some students followed along with these Geometer’s Sketchpad illustrations on their own iPads or laptops.

After the brief introductory lecture, students were given time to work on one or two problems. For instance, during one class session I observed on related rates, Julie helped the students set up a problem with a ladder sliding down a wall during the introductory portion, then had students work together to solve the problem. Julie encouraged students to work together to solve these problems. During this problem-solving time, both Julie and her TA would wander the room answering questions or providing advice. These problem-solving segments generally lasted between five and fifteen minutes.

After a problem-solving segment, Julie would pull the whole class together again to discuss the problems students had been working, provide new information, collect students' ideas on how to solve the problem, or demonstrate more examples. This cycle might repeat a number of times during the class session.

In addition to the three weekly class sessions, there was also a mandatory one-hour discussion section led by the TA that met once per week. During this session, students commonly asked for explanations of homework problems or examples presented during class. The TA would solve problems on the board and answer general questions about class material.

6.2.1. Geometer's Sketchpad

While LDT students outperformed other students on the conceptual questions on the final exam, there was no significant difference in overall CCI score. One potential explanation for this result is that the Geometer's Sketchpad (GSP) applets developed by Julie were not as successful as anticipated in helping students develop conceptual understanding of calculus. Why might this be the case? Were there problems with the applets themselves, or were there problems with the way they were integrated into the classroom?

A consistent feature of the reports on using GSP and similar programs in the literature (e.g., Jiang & McClintock, 1997; Olive, 1998; Purdy, 2000; McClintock et al., 2002) is that the benefits of using GSP generally accrue only to the person whose hand is on the mouse. It does not seem to be effective to simply show students GSP sketches; the power of GSP appears to be in students themselves interacting with the drawings. In other words, similar to other kinds of mathematical tasks, applets alone do not guarantee engagement; the implementation and teacher practices help determine the strength of the affordance.

Gresalfi (2009; see also Gresalfi, Barnes, & Cross, 2012) discussed the *forcefulness* of affordances, defined as "the imperative a student was likely to feel to comply" (p. 341). An affordance is *strong* when students are required to respond, or when not responding would violate rules or social norms; an affordance is *weak* when responses are not necessary, or when failing to take up the affordance does not have negative consequences. In the LDT class, the use of GSP was framed as an affordance for developing conceptual understanding. The affordance here was two-fold: GSP could be used in-class, to explore applets at the same time as the instructor demonstrated them, or after class, while studying, to enhance understanding of material covered in class. Throughout the focus group interview, students spoke of GSP in ways that called into question the strength of these affordances; it became clear that students were largely not the ones with their hand on the mouse, either during class time or outside of class.

While GSP was listed as a required course material, and while it was inexpensive (approximately $10 for a student license), not every student purchased it. Three of the seven focus group students did not purchase a license. When I asked the focus group students to estimate how many students bought GSP, the consensus was a little less than half. It is difficult to assess the accuracy of this consensus, but it is indicative that a substantial number of students did not buy GSP. These students certainly could not have had their hand on the mouse, either to follow along in class with Julie's demonstrations or to work with the applets later to strengthen their understanding.

There are dynamic geometry programs similar to GSP that are free, most notably GeoGebra; moving to a free program would remove the cost barrier to participating. However, adoption rates would likely remain low in the absence of accountability measures; in the language of Gresalfi (2009), without specific requirements to use GSP, the affordance would remain weak. However, if students were required to complete assignments that required GSP, perhaps by submitting applets or screenshots, the affordance would become much stronger.

A student named Michael compared his experience using GSP in class on an iPad to the experience of the students around him. Having the iPad and interacting with the applets made it much more useful for him.

> Michael: I have an iPad so I can actually work with it myself as opposed to just watch her work with it. So for example, I sat with two other kids who didn't have access to computers or iPads, so they'd just watch her and they were like, "I really don't like this," because they were following her pace. But with me, I could follow along with it and I could actually, if she did something that I didn't understand, I could keep looking through it. So if you have access to it, I think it's really helpful, but for students that didn't have access to it, I think it was an unfair disadvantage.

Since Michael had his hand on the mouse, he was able to interact with the GSP applets in a more substantial way than students who did not. He thus felt that it was more useful for him than it was for the others.

Several students said that it was difficult for them to see the relationship between the geometric representations presented with GSP and the algebraic representations.

> Michael: I felt GSP, how she kind of showed us this graph and she's like this is what this change does and this is what this change does, but she didn't really show the algebraic approach to it, which is what we're gonna need more on tests for example. ... I just didn't think it was a good mix of the two.
> David: Yeah. It made more sense when you kind of understood the algebra before she showed the graph. And then you can kind of figure out why the algebra works that way. But she showed the graph and then showed the algebra and I have no clue what she did on the graph.

Later, both Michael and David said that they felt that there needed to be a better balance between GSP and other methods.

> Michael: She needed to not use it and use it, but she didn't really blend that well. She leaned more towards the technology.
> David: I think she did fine with the online videos, but the GSP wasn't really blended, like he said, very well.

Michael also said that he felt that class lectures involved "too much of her showing things on the computer."

The perceived lack of balance between GSP with its conceptual orientation to calculus and other methods which might place a greater emphasis on procedural understanding helps explain an earlier quantitative result. In chapter 4, I reported that students in the LDT class performed better than other students on conceptual items on the final exam, but were outperformed by other classes on procedural items. I conjectured that this could be because of an increase of time spent on developing conceptual understanding in the LDT class, and a concomitant decrease in time spent on developing mechanical skills. Michael's and David's comments about the balance between GSP and other methods triangulate this hypothesis.

Another problem with the implementation of GSP in the class was that students were not trained on how to use it. Thus, they were unsure how to create their own applets, or significantly modify the ones created by Julie. Rose said, "I liked it because it helped me understand more, but then when I went home and like, I got to like play with it or whatever, I didn't understand it at all. Or like, how do I work this?" This theme recurred later in the interview when I asked how students used GSP outside of class.

> Int: Did anybody use GSP outside of the applets that Julie had?
> Rose: I tried.
> Vincent: Yeah, I had no idea how to create an applet for myself. I just download hers because it's really confusing.
> Rose: I tried to like work other problems, but since I didn't know how to use it, I was just like okay, I give up.
> Cesar: I didn't try.

Students would likely have benefited, and the in-class and out-of-class affordance would have been strengthened, if they had had instruction on using GSP.

The students had reasonably clear ideas about Julie's aims in using GSP. They said that it was intended to help them draw connections between the procedures and the geometry in a clear way.

> Fred: She wanted it to show us the geometry behind it, that she really wanted us to understand both algebra and the geometry behind the problem.
> David: Yeah she really wanted to give us like a visual... how it's done. You know, why you're doing it, kind of.
> Michael: I kind of also got that she likes how clean it is, and she says, it looks better than her handwriting and stuff.

I asked students if they felt GSP was useful for their learning. While this prompted Michael's concerns about balance, reported earlier, the responses were positive. Fred said that it helped him make connections and use dynamic images: "I liked it, because I saw things moving. Like, I connected everything together, like the cosine and sine graphs, I had no idea where they come from, I still kinda don't, but I know a little bit better."

These qualitative results help provide explanations for the quantitative results presented in chapter 4. There were not statistically significant differences between the four classes in overall performance on the CCI, a measure of conceptual understanding. However, item-level analysis of the CCI revealed that students in the LDT class performed better on three items assessing understanding of derivative; further, students in the LDT class had a higher average score on the conceptual items on the final exam than those in other classes. So, several different measurements of conceptual understanding yielded mixed results: while there was no difference in overall CCI score, there was some measurable impact on students' conceptual understanding, as seen in the item-level analysis of the CCI and the analysis of conceptual items on the final exam.

While the psychometric properties of the CCI have been called into question, as discussed in section 7.4.4, it is worth investigating explanations for the face result that there was no difference in overall CCI scores, and only a few differences in an item-level analysis. A plausible explanation for these results is that while GSP was useful for students, the affordance for using it was weakened by logistical implementation issues. Since there was functionally no requirement to use GSP regularly, a substantial number of students did not buy the program. Those who did received little training in its use, and thus may not have used it to its full potential. These issues in the implementation of GSP in the LDT class, and the concomitant weakening of the affordance, might be mitigated by requiring students to turn in assignments involving the use of GSP and providing some basic “bootstrapping” training at the beginning of the semester.

Whether the applets themselves were well-designed is another question, and one that is difficult to answer with this set of data, though students in the focus group generally felt that the GSP applets were valuable for their learning. Zazkis (2013) studied the use of an applet called the Tangent Intuition Applet (TIA), which was used in both the LDT class and in a calculus class Julie had taught the previous year. He found that several students who had used the applet in class continued to reason in ways inspired by the applet, even in paper-and-pencil settings where the applet was not available. Future work could extend this line of research by conducting task-based individual interviews focused on tasks that would be expected to evoke applet-inspired reasoning, or by conducting interviews with students around their use of GSP applets to see how the applets strengthened students’ conceptual understanding of calculus.

6.2.2. Level of interactivity

I framed this class as an interactive class and thus expected, in line with the meta-analysis conducted by Freeman et al. (2014) that found that active learning leads to better student outcomes, that students in this class would outperform students in less interactive classes like the Lecture class. However, as reported in chapter 4, there were fewer significant differences in learning outcomes than I had expected. This led me to wonder how strong the affordance of interaction was for the students in the LDT class. In other words, how interactive, really, was the LDT class?

Near the beginning of the focus group interview, I asked the students this question directly.

> Int: How interactive would you say her average lecture is?
> David: Pretty interactive. She tries to get everybody involved.
> Vincent: She always asks us for the answer. She doesn't just give it to us.

Students recognized the instructor's focus on keeping students involved in the intellectual work of the classroom. They felt that students could not be passive receivers in the class; as a whole, the students could not sit back and wait for everything to be done by Julie. However, they did not think that all students engaged equally, identifying a group of seven to ten students who most frequently asked and answered questions.

> Int: Do you feel like most students tended to be involved and stay involved?
> Michael: It was basically the first two, maybe three, rows that were actually paying attention and involved, and it seemed like the back -- I forgot many times existed. I found out that another member who I saw every day was actually in that same class, weeks into it, and I didn't know because they sat in the back.
> Jessica: Yeah, I feel like the same group of people answer the questions every time.
> Int: How big would you say that group of people was?
> Jessica: Seven?
> Michael: Seven to ten people.
> Int: Would everybody agree with that?
> David: That's about right.
> Fred: And I also think it was the same group of people asking the questions.
> Michael: And I think she should have some sort of system where she picked random people as opposed to whoever just wanted to say the answer, raising their hands.
> Vincent: Yeah, she would ask a question and they'd blurt out the answer. But I mean, that might just be because people don't know what the answer is, or how to do it, so. Those seven to ten people might be the greatest, the

> A+ people in the class, and the other people just might be like, B, B- people, who don't really know it.

One way that Julie created affordances for engagement and opened spaces for students to participate was by asking them questions during the course of the lecture. Following Gresalfi’s (2009) descriptions of the forcefulness of affordances, this affordance was voluntary in that it would not be a violation of the norms of the classroom for any one student not to respond. The perception of the students in the focus group interview was that only a small group of students took up this affordance. During my classroom observations, I did not keep close track of which students responded to questions, so I cannot say definitively whether or not this perception was accurate. Additionally, there is no way to observe students’ internal level of engagement with such questions; it is probable that many students thought about the questions that Julie asked without ever voicing an opinion.

Vincent speculated that it was the high-achieving students who chose to engage with these questions by voicing opinions. It could also be that these were the students who were most outgoing and therefore most comfortable speaking up in class. Michael's suggestion to call on students at random can be read as a way to strengthen this affordance; this would make it more mandatory for students to answer these questions.

Another way that Julie created affordances for engagement was by giving students problems to work during class time and encouraging them to work together to solve them; she and the TA would then wander the class helping students. This happened in all four of the class sessions I observed. Julie indicated that this was common practice, occurring nearly every day. When I asked her to estimate the proportion of students who actually engaged with these problems, she said that over 85% of students engaged every time she gave such problems; this figure is in line with the number of students I saw participating during my classroom observations. In three of the four class sessions that I observed, I noted that most students chose to work with other students sitting around them; on the other occasion, most students chose to work individually, though there were still several students working in small groups.

At one point in the focus group interview, I asked students how often they worked in groups in their class.

> Int: How often do you work in groups in your class?
> Rose: Almost never.
> David: Very rarely. Only when she puts that problem up there, something like that, and says, "Okay, take five minutes to try to solve it." You kind of just can work with the person next to you, that's about it.
> Fred: Well, it's hard to work in groups, because you're kind of in an awkward position.
> David: Yeah, it's a really small space.
> Fred: So you can't turn around.
> Michael: So the only people you can really work with are the people to your left and your right, and that's about it.

This response was surprising; to my eyes, what was happening during those periods was group work. I thus wondered why students said that group work happened "very rarely" or "almost never," or, in other words, how students interpreted the meaning of “group work.” It seems that the students did not view these problem-solving sessions as group work, perhaps because the groups were informal rather than well-defined, or perhaps because of the difficulties of working with students around them imposed by the physical setup of the classroom. Perhaps students define group work exclusively as work with assigned groups and group submissions; that is, perhaps turning something in together is required for students to think of an activity as group work.

Further research could examine the level of engagement of different students in opportunities to work and discuss their ideas with their classmates. This may be a complicated question; while they did not examine group work directly, Ellis, Kelton, and Rasmussen (2014) found that different groups of students perceive the frequency of use of important instructional techniques pertaining to active learning in different ways. For instance, they found that even when switchers and persisters were in the same classrooms, switchers reported lower levels of whole-class discussion than persisters. They argued that instructors are more likely to engage the students who most clearly articulate their thinking and who are most likely to advance the discussion; the students who were not explicitly involved in the discussion may thus not perceive the discussion as involving the whole class. Thus, instructors and articulate students may perceive the activity as a whole-class discussion, while quieter or less-engaged students perceive the same activity differently. Similar phenomena may occur in the context of group work: for those students who choose to work together, or for an outside observer, an activity may be perceived as group work, but for other students, hindrances to working together in groups may be more pertinent, and the same activity might be perceived differently. Perhaps there are specific qualities of students that might influence them to engage less frequently or less substantively in opportunities to work with their classmates; teachers could then be on the lookout for students showing those qualities, and could support their engagement more directly.

At the beginning of each class, Julie held a poll to which students could respond by text message. These polls were usually simple conceptual questions that did not require a substantial amount of work to answer. Fred described them as "very simple math problems, but you had to understand this big overlaying idea as to why it works like that." Students could work together to determine their answers to the text polls as well; during my classroom observations, I noted that most people discussed the poll question with the students sitting around them. This was corroborated by students in the focus group interview:

> Int: Would she usually have you solve [the poll questions] on your own, or could you work with people around you?
> Rose: You could work with people.
> Michael: She had it open for discussion so anybody could talk about it.

This is another example of something I would classify as group work that the students apparently did not, likely for similar reasons to those discussed above.

To sum, students felt that Julie's classes were interactive, but that the affordance to participate could have been stronger; Michael suggested that the affordance might be made more forceful by calling on students randomly to answer questions. Although there was ample time in each class session for students to work together on problems, they said that group work happened "very rarely" or "almost never," perhaps because the groups were informal, there were hindrances to working in groups in the physical setup of the classroom, and there was no requirement to turn in work done as a group. This time was successful in allowing students to engage with the material, with 85% or more of the students consistently working on problems in each class session.

Overall, despite the fact that there were areas for improvement and ways that the affordances for engagement could have been strengthened, classifying Julie's class as "interactive" is reasonable. Freeman et al. (2014) stipulate the following definition for active learning: "Active learning engages students in the process of learning through activities and/or discussion in class, as opposed to passively listening to an expert. It emphasizes higher-order thinking and often involves group work" (p. 4-5). By this definition, Julie's class certainly qualifies.

It is curious, then, that the differences in student outcomes were not as marked as in many of the studies reviewed by Freeman et al. (2014). Perhaps this is because the level of engagement of individual students was not as high as might be expected. Further research could attempt to track how each student engages or does not engage with their nearby classmates during periods of time similar to those in the LDT class, where the instructor gives students problems and suggests that classmates team together to work on them.

6.2.3. Online videos

Like the Inverted class, the LDT class made use of online videos, including Khan Academy and videos made by the instructor. However, the ways in which online video were used were different from those in the Inverted class, and students' experience with the online videos was much more positive. In this section, I examine how the uses of online video differed, and why students in the LDT class found it useful.

The primary difference between the use of online video in the LDT class and in the Inverted class was that Julie framed online videos as a supplement instead of the sole means of content delivery. Students found this approach helpful.

> Int: Did [Julie] say anything about why she liked to use online video?
> Rose: Because it was another, like if you don't understand it her way, you could learn it somebody else's way.
> Vincent: A different perspective.
> Int: Do you think it was helpful for you?

> David: Yeah. It kind of helped to see it from different viewpoint or somebody else explaining it. Like if you didn't understand her explanation, you could understand theirs.
> Fred: Many times I found myself confused as I left her class, and then the videos were like a little bit clearer.

Students in the LDT class used online video to provide different viewpoints and explanations for concepts taught in class, or to help clarify confusing points. This is almost a complete inversion of the situation in the Inverted class, in which online videos were the sole source of explanations. The LDT students appreciated having supplements to clarify things their professor said, while the Inverted students were frustrated that their professor was not available to clarify things the videos said.

The affordances of online video pointed out by students in the focus group were consistent with those reported in the literature on inverted classrooms: videos can be paused and rewound as many times as necessary, and can be watched at students' leisure.

> Fred: You can slow it down -- well, not slow it down, but go back and look at it again.
> Michael: And the cool thing with the videos, it's actually something that Khan even says himself, is that you can just play it, pause it, rewind it, and you don't have to keep asking the same question over and over again. It's at your own leisure, and you don't feel that need to bug somebody to ask again and again.
> Cesar: You can go back and listen to -- because sometimes I didn't get something, so I kept rewinding the video until I got it. Or it got clearer.

These affordances made online videos a useful way for students to study. When I asked the focus group students what resources they used to prepare for tests, three of them (Vincent, David, and Fred) mentioned using online video. Daniel even watched online videos multiple times: "It kinda helps to watch videos beforehand too, and then once you kind of learn how to do it a little bit and then watch the video again, you understand it differently, is what I've found out before tests."

Julie made a number of online videos herself, to help students understand things that she felt were not made clear during class time, or to provide tutoring for difficult homework problems. Additionally, when the class did not perform well on one of the exams, she made a suite of videos to help students create test corrections. The videos she made went over very well; students found them very useful, and as reported in section 6.2.4, they saw the videos as a sign of her caring.

> David: If her classes were online too, she'd be a pretty good online teacher, because those explanations she did on the homeworks -- If she just did her whole class session online, I think I'd understand it a lot better than actually going to class.
> Michael: Her online videos were pretty helpful.
> Vincent: Yeah, her online videos were really good.

David felt that Julie's online videos were even more clear than her in-class explanations. He later said that the ability to edit videos contributes to their clarity in comparison to a live presentation: "The thing with recording a video is that if you mess up, you just delete that part and re-record that section, as opposed to a class, if there's an error or a mistake you gotta fix it in that moment, and [the students] have already seen the mistake happen."

So, why did students in the LDT class find online videos so much more helpful than did the students in the Inverted class? Primarily, online videos in the LDT class were framed as supplements to the presentation during class time. Julie put a great deal of time and effort into creating videos tailored to the needs of the LDT students; these videos were seen as helpful resources, as well as a sign of caring. I elaborate on this point in section 6.2.4.

In contrast to the Inverted class, in which the instructor's reliance on outside videos was seen as a sign of disinterest, the videos provided in the LDT class strengthened the relationship between the students and the instructor. Once more, the didactical contract is a useful way to understand this difference. Students in the LDT class saw their instructor as fulfilling her end of the didactical contract, then going beyond her responsibilities to find or create extra videos. Students in the Inverted class, on the other hand, felt that their instructor breached the didactical contract by not providing satisfactory videos.

6.2.4. Instructor’s level of caring

In Seymour's (2006) report before Congress on why students leave STEM disciplines, many students implicated unavailable, disinterested faculty, who took little responsibility for student learning and exhibited dislike or disinterest in teaching. Julie, on the other hand, was none of these things for the students in her class. Students felt a strong sense of connection to Julie. They felt that she was invested in their success and would go the extra mile to help them be successful. In this portion of transcript, I asked students a series of questions about their instructor's attitudes toward students.

> Int: What would you say is Julie's attitude towards students?
> Vincent: She loves us. [Other students agreeing.] She wants us to do really well, and she really cares, and she gets mad at herself and us when we don't do well.
> David: Yeah, like I said, she genuinely actually does care. She's not like, "I'm going to try and put really hard problems on the test and make you mess up."
> Int: So what kinds of things does she do that that make you feel this way?
> Michael: She spends a lot of time working on external resources for us.
> Vincent: Yeah, like she made all those videos for the test corrections, and I totally appreciated that one. And she is available by appointment for

> office hours, and she does office hours.
> David: That and when you do go to appointments with her, when you leave she's like, if you need any more help, just email me and -- she's always really adamant about wanting to help you out if you need help.
> Fred: And I think I felt like she was really easy to talk to. That helps out too, when you're - when your teacher's approachable.
> Int: Does [Julie] seem to think that students are capable of understanding calculus?
> Fred: I think that's the main reason why it frustrates her when we don't do well. Because she thinks we should.
> David: Yeah, she knows that you can understand it, but everybody kind of needs a different way for it to be explained. So that's what she wants to do is, if you go to office hours she'd show you three different ways, slowly until you've figured out which one would work best for you.

There are several themes that emerge in this portion of transcript. First, Julie cared deeply about the success of all her students. Second, Julie was willing to create extra resources for her students. Third, she was available and helpful outside of class time. These themes recur in other portions of the focus group interview.

The feeling that Julie was invested in her students' success was brought up by several students. Michael said, "She was very, very concerned about making sure that you passed. And you could really tell that she did want us all to pass." David said, "She like actually wants you to do good." In contrast to the picture of "poor teaching" painted by the Seymour report, Julie took responsibility for her students' success.

Students appreciated Julie's willingness to create extra resources to help her students learn.

> Michael: The main thing I loved is that she would go out of her way to make study guides and extra work for us to go through. And I think a lot of people didn't really take advantage of it but I was very grateful for it.
> David: If somebody asked for help she would make a video.
> Fred: She's very good at that. If somebody asked for an explanation she would go out of her way.
> David: She like actually wants you to do good. That's why it kinda helped with her branching off and letting you see other explanations of how to do

> it. Because a lot of other teachers are kinda like, "Well, this is how I do it, this is the way I do it, you have to do it this way."

David felt that Julie was less concerned with having students memorize particular procedures without understanding, and more concerned about helping students find understanding using whichever procedure worked best for them. In general, students recognized the extra time and effort that Julie put into creating resources, and this served to strengthen the connection between them. This is in contrast to the Seymour report in which students described faculty as distanced and disinterested in teaching.

Students praised the availability and usefulness of Julie's office hours. They were the chief outside resources that students in the focus group reported using outside of class.

> Int: Did you use resources like the tutoring center, TA office hours, professor office hours? What kinds of outside resources did you use?
> Vincent: I went to Julie's office hours. She was good during her office hours.
> Jessica: I went to Julie too, for office hours. She was really good. And she was more outgoing in office hours than in her class. Like, she'll break it down more, if you don't understand it.
> David: Yeah, I think [both] the TA's and Julie's tutoring sessions were pretty helpful.

Again, students noted Julie's willingness to "go the extra mile" in presenting multiple explanations and "break[ing] it down more" in her office hours. They appreciated her availability outside of class time. There was no sense of Julie being insufficiently available, as there was in the Seymour report.

A theme that did not emerge in the extended portion of transcript presented at the beginning of this section was Julie's enthusiasm for calculus. Even though students did not necessarily share her enthusiasm, they appreciated that she enjoyed her subject.

> Fred: I liked how she was so enthusiastic about calculus.
> Vincent: Like she cares, a lot.
> Fred: I don't understand why, but she loves calculus.
> Vincent (quoting Julie): "This is the beautiful thing about calculus... Calculus is a beautiful thing!"
> Michael: She did things that would make us laugh, and you know, she'd have fun with it. I liked that.

Again, this is reflective of Julie's interest and enjoyment in teaching, in contrast to Seymour's reports of teachers manifesting implicit or explicit dislike for teaching or for their subject.

6.3. Features of the LD class

I begin as usual with a description of a typical day in the LD class. Corbin's class met twice a week for 100-minute blocks. Class time cycled between periods of lecture and periods of student problem-solving time. During lecture periods, Corbin introduced new material and worked examples. While working examples, he very frequently asked students what they thought the next step should be. He encouraged and was responsive to student questions during lecture periods.

Problem-solving time usually occurred after Corbin had demonstrated an example of a similar problem on the board with student input. He would write a new problem on the board and invite students to solve it on their own. For example, during the lecture I observed on the fundamental theorem of calculus, he solved $\int_0^2 2x\, dx$ on the board, then invited students to solve $\int_{-1}^{3}(-x^3 + x^2 + x\ )\, dx$. Periodically, he would provide scaffolding, or quietly work the next step on the board. Students sometimes raised their hands to call him over to ask individual questions. Students worked individually to solve these problems.

When cycling back to lecture from problem-solving time, Corbin might collect students' final answers to a problem, or ask if there were questions about how to solve the problem. Then he usually worked the problem in its entirety, soliciting and incorporating student suggestions on how to do it. He might then invite students to work another problem, or introduce the next concept.

I framed the LD class as an interactive lecture: the instructor posed problems to the students and gave them time to work on the problems and ask questions. He encouraged and was responsive to student questions during lecture. However, similar to the case in the LDT class, student responses in the focus group interview seemed to indicate that they perceived a lower level of interaction in the LD class than I expected; there was a perception throughout that the onus was on students to engage.

Early in the interview, I asked students to describe a typical day in class, and assess the usual level of interactivity.

> Int: Describe for me what a typical day in class looks like. Does your instructor lecture most of the time, do you do stuff? How interactive is the lecture?
> Carmen: He lectures most of the time, I would say. He does try to encourage people to ask questions if they do anything, and yeah, he helps people solve stuff. But I still feel like most people don't really get a chance to ask questions.
> Steven: I think it's as interactive as how the students make it. Like, you can ask questions if you want. I think some people just don't speak up.

Steven's comment seemed to set the tone for the rest of the discussions about the interactivity of the class: "it's as interactive as the students make it." In the language of Gresalfi (2009), while affordances for engagement were present, they were in general not very forceful. Again, since there is no way to observe students' internal level of engagement, and since I did not collect data on which students asked or answered questions during the lectures I observed, I cannot assess the accuracy of this perception.

However, during my classroom observations, I noted that students generally seemed to feel comfortable asking questions. This observation was triangulated by comments in the focus group interview. Michelle compared her prior experience in a large-enrollment calculus class to her experience in the small LD class, and said she felt more comfortable asking questions in the smaller class:

> Michelle: I took calculus last semester, in the spring, and it was this huge class, and then I just stopped going, because I was like I don't get it, and I don't wanna be the person in the 300-person class that's asking questions, you know. Now I at least feel comfortable asking questions in class. It's such a small classroom, and that really makes such a difference.

Later, she also pointed out that Corbin was approachable during class and did not demean students for asking questions:

> Michelle: I don't think he ever makes people feel stupid. There are people in our class that will ask really simple questions, and he's never like, "they just asked, I can't believe they're asking this." He's very enthusiastic about it, and that's nice, because if that was me asking it, then I hope that you would act that way to me.

In a review of literature on student participation (Rocca, 2010), smaller class sizes and supportive instructors were found by several reports to be associated with higher levels of student participation. On the other hand, students who perceive their instructors as sarcastic, demeaning, or condescending were less likely to participate; these are traits that Michelle specifically said that Corbin did not exhibit.

One notable strategy Corbin used to create spaces for student engagement was to invite them to work on problems after he demonstrated a few examples on the board.

However, students in the focus group felt that they were not given enough time to work on these problems.

> Becky: He gives you an example, then he gives you probably 15 seconds, 20 seconds before he introduces the next step to you. So he says, "okay work on it for a minute," and then he'll probably give you maybe that minute, and then he'll start working on it. And I think because students are so used to knowing -- they know that he's going to start working on it, that they probably don't even continue to work themselves. We just wait for him to finish it. I'm pretty sure that's how it works.
> Int: Is that everybody's experience?
> Michelle: Yeah.
> Carmen: Yeah, I agree with that. I don't feel like I have enough time to finish my problem.
> Steven: I don't even try to do the problems sometimes, I just write it down exactly how it is, and actually my notes are pretty easy to read after, so I don't have any problem with that. But I don't do the problems [during class].

This consensus was surprising to me. When I observed Corbin's class sessions, I noticed this pattern of giving problems to students and allowing them time to work on them individually. I wrote in my field notes that I felt Corbin allowed ample time for students to work through problems themselves, and also noted that several students provided numerical answers to a problem when Corbin asked what they got after a problem-working period.

It appears that that Corbin's strategy of working through the problems impacted how some students engaged with them. Henningsen and Stein (1997) found that the cognitive demand of a task was often inadvertently reduced by the ways in which the teacher introduced or supported it. They found that strategies that preserved the level of cognitive demand included using tasks that built on students' prior knowledge and devoting appropriate amounts of time to the task (neither too short nor too long); strategies that caused a decline in cognitive demand included too much or too little time and a lack of accountability. Corbin's strategy of periodically doing the next step on the board appears to have caused some students to disengage from the problem entirely. Additionally, all four students in the focus group reported feeling that there was not enough time to complete the problems.

Students reported that group work did not occur in their class, which aligned with my observations. During the times when Corbin gave the students problems to work on, I observed that the students worked on the problems individually.

So, was the label of "interactive lecture" appropriate for the LD class? On the whole, I found that it was: students were comfortable asking questions, Corbin asked students to provide input on how to solve problems, time was devoted to student work during each class session, and I observed many students actively engaged in solving problems during these times. Similar to the LDT section, there were certainly ways that the affordances for interaction could have been made stronger. In addition to increasing the level of accountability for engaging in problem-solving segments of the class by perhaps requiring students to hand in their work on one or two problems, Corbin might revise the strategies he employs during the problem-solving segments to ensure that the cognitive demand remains high. In particular, allowing students more uninterrupted time before showing partial or complete solutions, as well as allowing or encouraging students to work in groups, would likely increase the level of cognitive demand experienced by students during these segments (Henningsen & Stein, 1997).

6.4. Features of the Lecture class

Finally, I examine the Lecture class, taught by Rachel. I begin with a description of a typical class session in the Lecture class. The Lecture class met three times a week for 50-minute class sessions in a large lecture hall. The entirety of class time was spent in lecture. Rachel, the instructor, usually came to class with several pages of notes prepared in advance. Rachel used a document camera to project these notes onto a large screen at the front of the room, moving a blank piece of paper down the notes to gradually reveal each successive line. These notes usually included definitions, theorems, and a few proofs, reproduced from the textbook, as well as examples that had already been worked out.

Rachel very frequently asked simple questions with definite answers, such as "Since it's a product, we're going to have to do what?" or "tan θ equals what?". These were reminiscent of the typical IRE pattern characteristic of many classrooms (Mehan, 1979), although they frequently did not involve an explicitly-voiced evaluation. Students also commonly asked clarifying questions such as "do you really need to find out what θ equals?" or "what happened to the x^2 on the bottom?".

In addition to the large lecture sections, students were assigned to one of five discussion sections led by the TA that met once per week for 50 minutes. These discussion sections were very similar to those in the LDT class. During these sessions, students commonly asked the TA for help with homework problems, or for further explanations of concepts presented during class.

Students in the Lecture class did not enjoy the lectures delivered in each class session. Throughout the focus group interview, their concerns paralleled those reported by Seymour (2009) in her report to Congress on why students switched out of STEM major tracks.

Students in the Lecture class found the presentation of the material to be dry and technical, and too much like the presentation in the textbook. Frank described the lectures as being "like copying out the book." These concerns were voiced consistently throughout the focus group interview.

> Daniel: I thought that the way that we were taught, it was like exactly if we were reading the actual book.
> Frank: She was being a little bit too technical I think.
> Daniel: Yeah. I felt like if she would have explained the theorems, and then actually put it in her own words, and gave us an example, that would have helped out a lot.

Later in the interview, Lindsay expressed similar concerns:

> Lindsay: I think she did a lot of, she just reflected what the book said. That's not really helpful. You try reading a math book, it's kind of difficult. She's more like that, she's more notational, she likes teaching the exact notations, but it's kind of hard when you're just trying to grasp the conceptual aspects of a lesson. And you're just putting it into dry, mathematical -- and just writing out basically definitions from the book. Not really helpful.

Students felt that it was not helpful for them to see concepts presented in lecture in essentially the same way as they were presented in the book. For these students, the didactical contract likely included an expectation that the instructor would elaborate and clarify the material found in the textbook; the absence of “her own words” or examples to make the theorems and definitions more accessible was frustrating for these students.

The dry, technical presentation of the material led students to find class sessions boring and uninspiring.

> Frank: Yeah, her teaching methods were way too technical, and really dry. I like math, but the way she taught it, I [thought,] this is so boring.
> Paige: I thought so too, I thought it was boring.

Frank indicated that the presentation of the material negatively impacted his enjoyment of mathematics. This triangulates the quantitative result reported in section 4.3.1: students in the Lecture class demonstrated less increase in their interest in mathematics than those in other classes.

The perception that they had to copy the material as it was presented led some students to disengage from the intellectual work of the class.

> Isaac: At times, I did feel like I was copying the lecture and then I forgot what she was saying because I couldn't do both at the same time.
> Bianca: Oh yeah, I feel like I was so focused on catching up with what she was writing that I would not catch at all what she was explaining.

Frank echoed this concern at another point in the interview:

> Frank: You know, if you're just sitting there and you're copying stuff down, then you're focusing on writing down the notes and not even listening. My brain was wandering, [more] like what I'm going to do in two hours than paying attention.

Students felt they had to scramble to copy down the definitions, theorems, and examples as they were projected, and that they did not have enough time to engage with the conceptual material. This parallels Seymour's (2009) findings: students reported that their courses were delivered at too fast a pace for comprehension or engagement.

In another parallel to Seymour's (2009) report, students in the Lecture class reported that there was little application or illustration of material.

> Liam: Yeah, you don't get much breakdown of the problem, because she has such an agenda with the textbook. She doesn't stop and say, you know, "This is what you can do with this problem." And you can use those breakdowns to do other problems, even if they're not completely similar. But she was just plowing through the material.
> Sam: Yeah, when she tries to teach us how to do something, if-then

> statements don't really help. A lot of the class is just like, "if f(x) is differentiable, then... something." And it's like okay, show me how, I don't know.

Sam wanted there to be more examples in class to illustrate the theorems ("if-then statements") Rachel presented. While Rachel showed examples in each of the class sessions I visited, they were worked out in their entirety on the sheets she projected. Sam felt that "examples that are already worked out" were not helpful for showing him how to do something. Anna agreed with this concern:

> Anna: I didn't really like how in the lectures all the notes were already pre-written up. And I felt like she kind of went over it too fast, and didn't really give us enough time to copy it from the stuff. A lot of times not being able to finish all the notes.

Liam's concern about "plowing through the material," and Anna's comment here, connect to concerns shared by students in each of the three non-inverted classes about the pace and timing of their classes. I examine these concerns further in section 6.7.

On several occasions, Rachel forgot to bring the notes she had prepared for class, and would instead work problems out on the board or the document camera. Bianca said that these classes were more useful: "And at times I guess she would forget her notes. It'd be, I feel like I would learn stuff, because she would go along with us and write it, and explain it while she was writing it down."

Sam made a similar observation later in the interview: "She would have pre-- examples that were already worked out. And like [Bianca] said, when she'd forget her notes, she'd actually work the problems out, and that would help a lot."

Why did students find the classes in which Rachel forgot her notes more helpful than the classes in which she progressively uncovered a pre-worked example? Atkinson, Derry, Renkl, and Wortham (2000) conducted a review of the literature on worked examples, and concluded that students with more robust self-explanation behavior appeared to learn more effectively from worked examples. In particular, the self-explanation behaviors exhibited by successful students fell roughly into two clusters: anticipative reasoning, in which students predicted the next step and then checked their prediction, and principle-based explaining, in which students self-explained the overall conceptual structure of the problem and identified subgoals. Less successful students engaged in passive or superficial self-explanation, spending little time studying examples and generating explanations.

These categories of successful and unsuccessful self-explanation strategies are informative in the present context. If students feel rushed to copy down an example as it is revealed, lacking time to engage on more than a superficial level, they will not be able to generate substantial self-explanations. However, when Rachel worked out examples in real time, and "explain it while she was writing it down," students could engage with these explanations and produce self-explanations to go along with the ones given by the teacher.

6.4. Online homework systems

I now turn attention to several themes that cut across the focus group interviews. To increase clarity during the discussion of these themes, which includes portions of transcript from all four focus group interviews, I include the name of the class in parentheses behind the name of the student in all quotations from the focus group transcripts.

The first of these themes is student perceptions of the affordances and constraints of online homework systems in the classes. The LD class used an online homework system called MathXL; both the Lecture class and the LDT class used a similar system called Wiley+. I asked students in the focus group interviews to identify features they liked and disliked about whichever online homework system was used in their class.

Features of online homework systems that students enjoyed were instant feedback and the availability of a variety of problems for practice. Instant feedback was mentioned positively by students in both the LD and Lecture classes.

> Becky (LD): I like that if you enter it correctly it tells you it's correct, and if it's wrong it tells you it's wrong. Because then you know, and you can go back and see what you did wrong.
> Isaac (Lecture): I like that they show that if you're right or not.
> Tricia (Lecture): I like that too, that you had immediate feedback if it was right or wrong.

Instant feedback is the primary feature that distinguishes online homework from traditional pencil and paper homework. It was also mentioned as a major affordance of online homework by the students surveyed by Roth, Ivanchenko, and Record (2008). Students like Becky (LD) used the instant feedback as a check that allowed them to go back and find errors they made in their solution.

Students also appreciated the variety of problems presented in their online homework.

> Steven (LD): What I like about MathXL is that it gives you different types of problems, like different scenarios of where you have to do this or that. It just gives you a variety of problems.
> Lindsay (Lecture): It presents you problems that makes you think differently, and you're forced to ask about how to solve these, so it forces you to learn different tricks.
> Bianca (Lecture): Yeah, that's what I liked about it, it challenges you. For

> me, Wiley+ provided all those examples that I wanted the professor to have for us in class.

As discussed in section 6.4, a common complaint of students in the Lecture class was that Rachel did not present enough examples, or that the way in which examples was usually presented was not useful. The ready availability of examples on the online homework sets helped ameliorate for Bianca this shortcoming of her instructor's presentation of the material in class.

Because of these features, students found online homework to be a useful resource for studying for exams, especially in the Lecture section.

> Int: What kinds of stuff do you to do study for exams?
> Paige (Lecture): I use Wiley a lot.
> Isaac (Lecture): Yeah, me too.
> Int: So do you redo the problem sets?
> Paige (Lecture): Yeah - I look through them.
> Bianca (Lecture): Yeah, that's what I would do. Once I had them — well, some of them — done, I would go back and try to review them.

This was another category of positive comments made by the students surveyed by Roth, Ivanchenko, and Record (2008) about the use of WeBWorK in their course. Additionally, a Likert scale question assessing student perception of the effectiveness of WeBWorK for preparing for exams had average responses well above 3 on a 1-5 scale.

Students also identified several features of the online homework system that they did not like: the need to input answers in a particular way, and the all-or-nothing grading. Students in all three classes where online homework was used commented on the need to input answers in a specific way.

> Isaac (Lecture): Thing I didn't like is that it didn't accept some of the options -- like you have to simplify it in a specific way. It was annoying.
> Frank (Lecture): You had to cater to Wiley.
> Becky (LD): What I don't like is that sometimes you have the right

> answer, and you didn't read it right, and you're supposed to put it in decimal or simplest form, and it's wrong.
> Michael (LDT): And I also didn't like it because there were times where I got the question wrong because it was written a different way, but when I presented it to [Julie] she'd say, "oh, well that was correct." And I pretty much get docked because of that.

This was a source of frustration for students particularly because their instructors or TAs were able to see that their work was correct, but due to technical limitations of the online homework system, their right answers were incorrectly marked wrong. This was perceived as unfair; another likely provision of the didactical contract is that students expect that their work will be graded fairly.

This concern about fair grading tied into the next concern: since problems were graded as either correct or incorrect, small mistakes could have disproportionate impact.

> David (LDT): And then the problems that were really long, like there was like A through F on it, and if you got one of them wrong —
> Rose (LDT): Oh, then the whole problem was wrong.
> David (LDT): That was really annoying.

Some students said that these features caused them to worry more about getting the problem correct than about understanding the procedure:

> Tricia (Lecture): I feel like with online homework, sometimes I was a little bit too concerned with the grade I was getting, and not so much with how do I do this problem. And I think that kind of hurt my overall grade. I was so focused on just getting the answer right that I was looking for book problems, basically copy what it said in the book, for the answer.
> Bianca (Lecture): Me too. But I wouldn't understand how to do it.
> Tricia (Lecture): And I would get it right, but I didn't really know how to solve that problem on my own.

Seymour (2006) reported to Congress that many students developed instrumental attitudes to learning characterized by an overemphasis on their grades and an underemphasis on content mastery. It is possible that the kinds of grading embedded in online homework systems might help contribute to these instrumental attitudes; further research could attempt to assess the existence or strength of this link.

These student comments on the affordances and limitations of online homework systems are consistent with those reported in other studies. Roth, Ivanchenko, and Record (2008) surveyed students on what they liked and disliked about WeBWorK, another online homework system, and found that students enjoyed the immediate feedback and help for exam preparation, while disliking the difficulty of inputting answers and the lack of hints when a submission is close to the correct answer. Similarly, Hauk and Segalla (2005) found that many students disliked the difficulty in communicating effectively with the online homework system.

The feedback provided by online homework systems like Wiley+, WeBWorK, and MathXL is usually limited to binary answer-checking, correct or incorrect, though some provide small hints. Based on Tricia's (Lecture) and Bianca's (Lecture) comments cited above, I hypothesize that because online homework systems generally only provide binary feedback, though some provide small hints, they might serve to strengthen students' belief that mathematics is about getting correct answers to specific problems. Hauk and Segalla (2005) argued that some student concerns with WeBWorK could be traced back to a challenge to their belief that mathematics is computation, but they did not attempt to determine whether or how students' beliefs about mathematics shifted due to their work with online homework systems. Further work could examine this important question; in chapter 7, I propose several ideas for potential research designs.

6.5. Calculus as a prerequisite for calculus

The second cross-cutting theme that emerged from focus group interviews was that students felt like their calculus classes were taught in such a way that calculus was a prerequisite for calculus. My demographic analysis, presented at the beginning of chapter 4, indicated that approximately 60% of students in these classes had taken some class called calculus before, whether at the high school or the college level. Many students felt that it would be much more difficult for the other 40%, the first-time calculus students, to succeed.

Rose, from the LDT class, felt that “Julie teaches the class for people who have already taken calculus. Like, she teaches it at that point of view.” David (LDT) was retaking calculus; thinking back to his prior college calculus class, he observed the same problem: "I noticed last semester the teacher I had was teaching it like everybody took calculus in high school too.”

Sam, from the Lecture class, expressed similar sentiments: “I think Rachel kinda assumed that we already have knowledge and – I mean, some people do, but those who don't, it's really kind of overwhelming.” Both Sam and Lindsay felt that not having prior calculus experience would make the class more difficult.

> Lindsay (Lecture): I was lucky enough to take calculus before, so it wasn't that difficult, but I just put myself in the shoes for people who are just learning it for the first time, and I was like wow. Really tough.

Lindsay felt that her previous calculus experience made the class easier, but thought that it would be difficult for students without prior experience.

A lack of prior experience was doubly problematic in the Inverted class, because of the problems with the pre-class materials documented in section 6.1.1. Both Bob and

Phoebe from the Inverted class had taken calculus in high school, and commented that it made their experience in class much better.

> Bob (Inverted): I've watched videos maybe two to three times out of the whole like out of the whole semester. I mean, most of what I remember is from high school.
> Phoebe (Inverted): Yeah, what's keeping me going in this class is high school, like math classes in high school I took, I just remember.
> Sarah (Inverted): Which is putting the people who haven't taken these classes at such a high disadvantage.

Sarah had not taken calculus in high school, and felt that she and other students like her were at a disadvantage. In a later segment, Bob agreed with this sentiment:

> Bob (Inverted): I honestly feel bad for people who haven't taken a calculus class before. So going in, I know what a derivative, I know antiderivatives, so I know the different things, like chain rule, product rule stuff. And I've been able to be like, "oh, okay, that's the rule," and then remember how to do it, and just have to walk through a problem and you fully understand it. But going in, if you don't know that and you have no background, that's a little difficult.

On the STS, students were asked how much they agreed with the following statement: “In order to succeed in calculus at a college or university, I must have taken it before.” The average score on this question was 3.345 on a six-point scale, which is just to the disagreeing side of neutral. This question was not asked on the ETS; future work could examine how students’ perceptions of this question shift over the course of a semester.

I conjecture that this theme is linked to the concerns about pace that I discuss in section 6.7, and thus leave further discussion of student perceptions that calculus is a prerequisite for calculus to be discussed together with the concerns about pace.

6.7. Pace of class sessions

The third cross-cutting theme was that students in each of the lecture-based classes shared concerns about the pace of daily class sessions. These concerns often paralleled those reported by Seymour (2006): students reported over-stuffed courses that were delivered at a clip that did not allow enough time for comprehension.

Jessica, in the LDT section, felt that the pace at which Julie presented examples did not allow her to check understanding and ensure that her students were following along: "I feel like in large lecture, she goes too quickly through examples, and she just expects us to follow it." Relatedly, Rose (LDT) commented that Julie “never really finished the lesson;” there was always more material than could be covered in the time allotted.

Michelle, in the LD class, felt that she did not have enough time to formulate questions when the instructor would ask the class if they understood something:

> Michelle (LD): Sometimes I need — if you're writing, and trying to look [at the board] at the same time — I'm trying to write down the step, and I'm trying to take it in at the same time, and so I need to look back, I'm working to see if I have questions. And then Corbin will say, “okay,” [moving on.] Which, I get why, because we only meet twice a week, and it's really not that much [time].

While this concern could be ameliorated with more wait time from the instructor, Michelle attributed the lack of wait time to the lack of time in class. Again, this is reflective of a feeling, similar to those reported by Seymour (2006), that there was too much material to cover during the class sessions.

A particularly telling quote that brought together several strands of the Seymour (2006) report came from Liam, in the Lecture class:

> Liam (Lecture): You don't get much breakdown of the problem because she has such an agenda with the textbook that she doesn't stop and say,

> you know, this is what you can do with this problem and, because you can use those breakdowns to like do other problems, even if they're not completely similar. But she was just plowing through the material.

In Liam's view, the curriculum and the textbook imposed such a rushed "agenda" on the class that Rachel was unable to provide "breakdowns" of problems that might have allowed for connections to be drawn across the course. In other words, the over-stuffed curriculum caused the lack of application or connection across the class. Similarly, Liam attributed a lack of step-by-step explication of examples to a rush to get through material:

> Liam (Lecture): Rachel would show us something on the board, and it would just be three lines, and okay, how did she get from this line to that line? Like what did she do? She would say, "Did you see that?" And she asked that a lot. "See that? See what I did there?" No, I don't see anything. You wrote the simplified version, but can I see the step-by-step please? But I think it's just time maybe. She's trying to cram everything.

Students also felt that the pressure to cover material precluded the use of active learning strategies. When I asked students in the Lecture class if there was ever group work in their class, Tricia replied that "there really wasn't enough time." Lindsay agreed, saying, "I think Rachel was just trying to rush through it because of time."

Even in the interactive LD and LDT classes, students felt that the pace of class sessions was often rushed, or that their instructors were often unable to cover all the material they had planned. They felt that this pressure for time caused their instructors to adopt strategies that sacrificed understanding for coverage. Students in all three classes felt that examples went too quickly for comprehension or reflection. These concerns are very similar to those reported by Seymour (2006). One clear solution to this problem is to scale back the set of concepts included in the curriculum. This recommendation is not new: Steen (1987) called for a "lean and lively calculus" nearly thirty years ago. In chapter 7, I discuss potential directions for creating leaner, livelier calculus courses.

Additionally, the perception discussed in section 6.6, that calculus was a prerequisite for calculus, is likely linked to these concerns about the curriculum forcing too fast a pace for understanding. With more modest curricular goals, more time could be spent on ensuring that students have a solid understanding of the foundational ideas of calculus, no matter how much or how little prior exposure to calculus the students have.

Chapter 7: Summary and Discussion

I begin this chapter with a brief summary of the results presented in chapters 4 through 6. Then, I discuss some broader conclusions and future directions.

7.1. Summary of results

7.1.1. Summary of quantitative results in chapter 4

In chapter 4, I presented the results of quantitative comparisons between the four classes. I measured persistence using two different methodologies: the methodology used by the CSPCC researchers, and a methodology relying on roster data. By the CSPCC methodology, there was no significant difference in switching rate between the four classes; by the roster data methodology, students from the Lecture class switched out of STEM major tracks at a lower rate than students in other classes.

Of the 15 beliefs items on the ETS, only five items were identified as differing significantly between classes. The Inverted class was on the unfavorable side of each of these differences; additionally, students in the LDT class were significantly more likely than students in the Lecture class to memorize instead of make sense of material, while students in the Lecture class demonstrated less increase in their interest in mathematics than those in the LD classroom.

I grouped together items assessing affective beliefs about mathematics and cognitive beliefs about mathematics, and created composite scores measuring these two variables at the start and end of the term. Differences at the start of term were not significant; at the end of term, differences in cognitive beliefs were not significant, but students in the Inverted class demonstrated lower affective beliefs than those in the LDT class or the Lecture class.

I compared mean post-term scores and normalized gains on the CCI, an instrument measuring conceptual understanding. Students in the LDT class performed slightly better than those in other classes, but these differences were not significant. Students in the LDT class performed better on three individual items measuring understanding of derivative.

Grade distribution and DFW rate differed dramatically between the four classes, but this can mostly be attributed to grading policies differing among the four instructors: in particular, the instructor of the LD class did not curve exam scores, and the DFW rate was especially high in this class. LDT-class and Lecture-class students outperformed Inverted-class students on the final exam, and Lecture-class students outperformed LD-class students. The overall mean final exam score was 51.7%. However, when controlling for incoming preparation, as measured by the CCR, there were no significant differences in overall final exam scores. Students in the Lecture class outperformed those in the LDT class and the Inverted class on procedural items on the final, while students in the LDT class outperformed all other students on conceptual items. These patterns held when controlling for incoming preparation, except that the difference between LDT and Inverted students on conceptual items was no longer significant.

In the Inverted class, male students had higher final exam scores than female students; however, when controlling for incoming preparation, this difference vanished. In the three non-Inverted classes, students who took a calculus course in high school outperformed those who had no prior experience and those who had taken a prior calculus course in college; in the Inverted class, the difference was not significant. This pattern remained when controlling for incoming preparation.

7.1.2. Summary of quantitative results in chapter 5

In chapter 5, I presented the results of quantitative comparisons between students at the local institution and students in the national database, as well as to students in three subsets of interest of the national data: those at institutions selected for case studies by the CSPCC researchers, those at master's-granting institutions, and those at Ph.D.-granting institutions. There were substantial differences in demographics, with students at the local institution having lower levels of parental education, greater racial/ethnic diversity, and lower SAT Math scores. Additionally, the proportion of male students was higher at the local institution, while the proportion of first-time calculus students was similar across the local and national data.

The switching rate at the local institution was significantly lower than the national rate. This is attributable to the higher proportion of engineering students (who switch at lower rates than those in other career tracks), and the sidelining of pre-medical and life science students into a non-mainstream calculus class, at the local institution. Comparing just the populations of engineers, there was no significant difference in switching rates.

I next compared composite scores measuring affective and cognitive beliefs at the start and end of term. At the start of term, local students had similar affective beliefs scores to those in each of the national groups, except for students at master's-granting institutions, who had higher affective beliefs scores. Local students had lower cognitive beliefs scores than students in any of the national groups. At the end of the term, local students' affective and cognitive beliefs scores were lower than those in any of the national groups. When controlling for incoming beliefs scores, local students still had lower affective and cognitive beliefs scores than those in any other group, except when comparing affective scores to those at Ph.D.-granting institutions (and this difference approached significance). When comparing the patterns of change in scores on the beliefs items that were identical on the STS and ETS, the local data displayed roughly the same trend as each group of the national data, though most scores were lower overall.

7.1.3. Summary of qualitative results in chapter 6

In chapter 6, I examined the qualitative data collected in focus groups and classroom observations in each of the four classes. Students in the Inverted class identified three categories of concerns with the way the inverted model was implemented: the online videos were not made by the instructor and thus did not align well with the objectives of the course; the in-class activities were not well-structured; and the students felt disconnected from the instructor, who they perceived as abrogating many of the responsibilities in the didactical contract.

In the focus group for the LDT class, it was revealed that although the instructor provided Geometer's Sketchpad as an affordance to promote conceptual understanding, the affordance was not very forceful, due to a lack of accountability and a lack of training. Students also indicated that the LDT class was interactive, although the students' understanding of what constitutes "group work" may have differed from mine. The use of online videos in the LDT class was much more successful than that in the Inverted class, because they were seen as supplements rather than as the primary vehicle for delivering content, and because the instructor developed many of the videos herself. The students took this, as well as her ready availability for office hours and her willingness to make extra resources to help her students succeed, as a sign of caring for her students.

In the LD class, students indicated that although the instructor provided time in class to work on problems after showing examples, they often did not work on the problems, because the instructor did not allow enough time before showing the next step. By employing this strategy, the instructor inadvertently reduced the cognitive demand of the examples. However, the instructor's demeanor and the small class size helped encourage students to feel comfortable asking questions, and his strategy of asking students to provide feedback on how to solve problems helped students stay involved in the class.

Students in the Lecture class voiced a variety of concerns paralleling those found in Seymour's (2006) report on why students leave STEM major tracks. They felt that the presentation of the material in lecture was dry, technical, lacking in applications, and too similar to that in the textbook. They often resorted to mechanical note-taking, without attempting to make sense of examples, to keep up with the fast pace of the class. They disliked the instructor's strategy of gradually revealing pre-worked examples, preferring the class sessions when she forgot her notes and worked out examples in real time.

Each of the three non-Inverted classes used online homework systems. Students enjoyed the instant feedback delivered by these systems, as well as the variety of problems available for practice, and found online homework systems a useful way to help prepare for exams. However, they did not like the all-or-nothing grading and the difficulty in inputting answers in a form that the system would accept. There was some evidence indicating that the kinds of grading embedded in online homework systems may have contributed to instrumental attitudes to learning.

Another theme cutting across classes was that students felt that calculus was a prerequisite for calculus. They indicated that the way their instructors presented material seemed to assume some level of prior calculus knowledge. Students without prior calculus experience were seen as being at a disadvantage. A lack of prior experience was doubly problematic in the Inverted class, because of the problems with pre-class videos.

Students in all three non-Inverted classes expressed concerns about the pace of daily class sessions. They felt that examples were delivered too quickly, that lessons were often left unfinished, that active learning strategies were precluded, and that they did not have enough time to formulate questions or understand concepts. Students speculated that the rush to get through each day's material was caused by the "agenda" set by the curriculum and the textbook.

7.2. Discussion

I now turn my attention to broader discussions and implications of the findings summarized above. The primary story told by this dissertation is a story of shrinking differences. On their face, the four pedagogical strategies were quite different, so it was surprising that there were so few statistically significant differences in the quantitative data. The consistent story of chapter 6 was that several implementation issues, some the result of pragmatic constraints, others the result of design choice, weakened affordances provided by innovative features and shrunk the differences between classes. The Inverted class did not follow best practices for implementing the inverted model; the LD class did not mandate participation and may not have allowed enough wait time; in the LDT class, not every student had their hand on the mouse with GSP, whether inside or outside the classroom. Therefore, while the affordances I hypothesized each innovative feature would provide were still present, they were diminished by the way the innovative features were implemented.

Some of these issues in implementation were the result of pragmatic constraints. For instance, in the LD class, allowing enough wait time for every student to finish working the problem might severely limit the amount of material that could be discussed in a class session; additionally, it may not have been clear to the instructor that some students wanted more wait time. In the LDT class, both the number of students in the class and the fixed arrangement of chairs limited the ways in which group work could be approached; the limited number of students who purchased GSP restricted the ability to have students' hands on the mouse in or out of class.

Revisiting the discussion of my theoretical framework, the emergent perspective, participation in classroom activity is seen to "constitute the conditions for the possibility of learning" (Cobb & Yackel, 1996, p. 185), and an individual's psychological development is seen as enabled and constrained by their participation in classroom activities. Therefore, classes that present more affordances for students' engagement and participation in classroom activities are seen as presenting more opportunities for student learning. The affordances for engagement presented to students differed in each class: the Lecture class offered frequent IRE-type questions (Mehan, 1979); the LD class offered questions moving past IRE patterns, as well as opportunities to work problems in class;

the LDT class offered opportunities to work with other students in class, as well as opportunities to use GSP applets to scaffold conceptual understanding; the Inverted class also offered opportunities for students to work together, and to be exposed to content before class time. However, in contrast to other inverted classrooms discussed in the literature, the Inverted class did not offer the opportunity to work with the instructor, which was a primary reason motivating inverted classroom design (Lage, Platt, & Treglia, 2000).

The affordances presented to students might vary not only in kind but also in forcefulness, or the imperative a student is likely to feel to take up the affordance (Gresalfi, 2009), and the forcefulness of an affordance may be modified by the way in which it is presented to students. In the LD class, the affordance to solve problems during class was weakened for some students by the instructor's practice of periodically working the next step on the board; in the LDT class, the affordance of using GSP was weakened by the fact that there were no assignments that required the use of GSP to complete.

Further, while increasing the forcefulness of an affordance increases the likelihood that it will be taken up by more students, students must still choose to engage. Students' choices are influenced both by the norms of the classroom and by students' beliefs, attitudes, and dispositions, the psychological correlates of the classroom norms. Innovative practices, especially those that modify the usual didactical contract prescribing the mutual obligations of student and instructor (Brousseau, 1997), require shifts in classroom social and sociomathematical norms. These norms are negotiated by both the students and the instructor, as members of the classroom community, and they are influenced by the beliefs all parties have about their roles and about mathematics

(Yackel & Rasmussen, 2002). Thus, the ways in which instructors implement affordances, and the ways in which students understand and take up those affordances, are influenced by student expectations. For instance, some students in the LD class chose not to work on problems, because of their expectation that the instructor would eventually provide answers; the instructor's choices in presenting solutions to advance the mathematical agenda of the class was likely a response to a tacit understanding that not every student could be expected to work every problem individually.

At the end of chapter 4, I listed several questions raised by the quantitative findings: What similarities existed between the four classes? Why did all four classes perform poorly on the final exam? How does this Inverted class differ from the successful ones in the literature? The qualitative data presented in chapter 6 helped answer these questions, and I summarize those answers below.

7.2.1. Best practices for the inverted model

The reports of successful inverted classes in the literature share several commonalities: the pre-lecture activities were tailored to the particular class, often personally created by the teachers or researchers; students were held accountable for completing the pre-lecture activities; and time formerly occupied by lecture was replaced with active-learning exercises led by the instructor of the class. The concerns reported by students in the Inverted class, both in the focus group interviews and on course evaluations, implicate the failure of the Inverted class to replicate these important commonalities. I thus propose these features as the beginning of a list of best practices that should be adopted for an inverted class to be successful. Here, I summarize the discussions given above on why each of these features is necessary for success.

The first feature is that pre-lecture activities are closely tailored to the class. The students in the Inverted class read the articulation failure between the pre-class videos and the in-class worksheets as a breach in the didactical contract: the videos were meant to prepare the students for the in-class work, but they did not, so they failed in their purpose. This caused some students to disengage entirely from watching the videos. While instructors in the literature typically create pre-lecture activities themselves, this does not seem strictly necessary, particularly for early foundational material. Perhaps this is because for this material, variations in presentation are less impactful; a lesson on the power rule, for instance, likely looks much the same no matter who delivers it. Future work could attempt to find criteria for when pre-lecture activities can be borrowed from other sources and when they must be developed in-house.

The second feature is accountability for completing pre-lecture activities. This can take many forms, from handing in a filled-in worksheet to completing a clicker quiz at the beginning of class; it may even be as simple as using a content management system to ensure that students clicked the link to a video. Accountability measures can be effective in motivating students to complete activities; Moravec et al. (2010) reported that each of their “learn-before-lecture” activities was completed by over 90% of their students. Without accountability, however, there is no guarantee that students will complete the activities; it is no surprise that students often do not do things they are not accountable for.

The third common feature is the use of active-learning activities in class, led by the instructor. One key motivation for the inverted model was to allow students the opportunity to engage with challenging material with the instructor physically present to provide scaffolding and support. The inverted model thus contrasts with traditional models, which assign students to complete challenging tasks at home without the instructor's help. Without engaging, well-structured activities, or without the instructor present, this affordance of the inverted model is lost.

7.2.2. "Lean and lively calculus"

One obvious similarity between the classes is the shared curriculum mandated by the common final. Given the poor performance on the common final exam (the overall mean was 51.7%, despite the fact that the majority of students had taken calculus before), these results appear to confirm Seymour's (2006) previous findings that introductory courses including Calculus I are often "over-stuffed" and taught too quickly. My data appear to be one more piece of evidence supporting the ongoing push for a "lean and lively calculus" (see, e.g., Steen, 1988). Many studies over the past thirty years have supported this conclusion; perhaps we in the mathematics education community need to find new ways to communicate to administrators and instructional designers that the current curriculum is "too much, too fast" for students to master.

Given that the majority of the students in this study and in the national database had taken a calculus class before, why shouldn't college calculus be able to proceed at a rapid clip? In other words, why were mean final exam scores so low even though most students had taken calculus before? This is an interesting question worthy of further study. Perhaps the problem of over-stuffed, too-rapid calculus classes is not unique to college; the mere fact that a student has taken calculus in high school is no guarantee that they have a solid foundation on which a college calculus class might build. Additionally, many students coming to a college calculus class with high-school calculus experience are freshmen, faced not just with the problem of learning calculus, but also the problem of negotiating a new set of institutional norms and expectations. Whatever the explanation, a more lean and lively calculus will likely benefit all students who take it, regardless of their prior calculus experience.

Future work could assess the efficacy of various different approaches to making the calculus curriculum more “lean and lively.” The removal, or at least de-emphasis, of several topics is likely to be necessary to create a “lean” curriculum; common targets include the epsilon-delta definition of the limit, l'Hôpital's rule, and Riemann sums (Douglas, 1986). However, the problem of which topics to excise is contentious, due to the variety of stakeholders in calculus. For instance, many mathematicians object to the removal of the epsilon-delta definition of the limit, since it reduces the mathematical rigor of the course.

To reduce the effective number of stakeholders, we could consider further fragmenting calculus courses. It is already the practice of many universities to have separate calculus courses for business majors and STEM majors, and some universities even break up STEM major tracks; in section 7.5, I reflect on my experience teaching a calculus class for life scientists. Having separate calculus courses for separate disciplines allows the courses to specifically target the calculus skills each discipline values most highly and ground the calculus concepts in the discipline-specific situation from which they emerge. However, especially at institutions with smaller enrollments, over-fragmentation could drive class sizes down to a point at which they are not economically viable.

Other approaches to changing calculus instruction might focus more on the “lively” than on the “lean.” The inverted model is a plausible way to increase the liveliness of a calculus course without sacrificing (too much) coverage. Moving content delivery outside of class time allows more time for “lively” active learning exercises, which have been shown to improve student outcomes in a wide variety of fields (Freeman et al., 2014). The particular implementation of this model in the Inverted course should not dissuade further research on inverted calculus courses.

Moving outside of class time, it might be possible to increase the effectiveness of recitation sections. As I discuss further in section 7.2.3, the usual model of a graduate TA working example problems and answering student questions increases the availability of opportunities for discussion. However, other models, such as workshop or laboratory models driven by students working together to solve challenging problems with minimal scaffolding from the more knowledgeable other, might be more effective. Talbert (2014) demonstrated that coupling a workshop model with inverted classroom design principles resulted in an effective and satisfactory learning experience in linear algebra; similar approaches could be studied in the calculus classroom.

The problem of making a “lean and lively calculus” is one that has persisted for almost thirty years, and it is not an easy problem to solve. It is likely that some combination of the approaches discussed above, as well as other innovative approaches not yet invented, will be necessary to address it.

7.2.3. Recitation sections

On the final exam, the classes clustered into two groups: the Lecture and LDT classes scored higher than the LD and Inverted classes. The classes with higher scores were the classes that had recitation sections in addition to whole-class lecture sessions, and I hypothesize that this may be a causal relationship deserving of further study. Rocca's (2010) survey of the literature on student participation found that larger class sizes were a deterrent to participation in the classroom, and recommended that large classes be divided into smaller groups to facilitate discussion. Recitation sections are a way to provide students with some time in smaller class sections, which, according to Rocca, increases their opportunities for discussion. Indeed, students in both the Lecture class and the LDT class noted that they were more comfortable asking questions and participating during recitation sections than during whole-class lecture. Increased opportunities for discussion translate, under the emergent perspective, into increased opportunities for student learning, since students' learning is enabled and constrained by their participation in classroom activities (Cobb & Yackel, 1996).

A rival hypothesis for the greater success of the Lecture and LDT classes is that recitation sections simply gave these classes more time on task. However, this is not the case, since the LD and Inverted classes had longer whole-class lecture sessions; the number of contact hours per week was the same in all classes. Thus, the more likely explanation rests on the differences in the *nature* of time on task, rather than the *amount* of time on task.

7.3. Limitations

In general, in this study, I traded fine-grained understanding for sample size. This study focused mainly on understanding trends in large groups of students, with focus group interviews used to corroborate and explain the large-n quantitative results. I did not collect any data examining how individual students think about solving problems in calculus. There may be important differences in the way students in different classes solve problems: a student in one class might deploy a set of memorized steps to solve a related rates problem, while a student in another class might use more conceptual strategies. In particular, the work of Zazkis (2013) implies that students in the LDT class may use patterns of thinking inspired by their use of GSP applets, even when the applets are not available. Future work in this vein could combine large-n quantitative data with small-n, fine-grained understanding of how representative individual students think about calculus problems.

Another limitation of this study is that while I was able to compare the local institution to the national sample on the survey instrument, I had no national-level performance data. A national database including both survey responses and CCR or CCI scores would have allowed me to compare the performance of students at the local institutions to other institutions nationwide.

7.4. Future directions

My study raised a number of questions that could inform future work. I summarize several directions that I find particularly interesting.

7.4.1. Online homework systems and mathematical beliefs

In section 6.5, I presented evidence that online homework systems may have promoted instrumental approaches to learning, characterized by memorization of disconnected facts and focused on grades rather than mastery (Skemp, 1978; Seymour, 2006). It would be interesting to examine the link between online homework systems and mathematical beliefs. Potential research designs to assess this link might include conducting interviews with individual students that probe both their mathematical beliefs and the sources for those beliefs. Another possibility might be to conduct a control-treatment study, in which some sections use online homework with simple binary feedback and others use homework, likely pencil and paper, with more detailed feedback.

Hanson, Nunez, and Ellis (2014) examined the homework assignments given at various doctoral institutions. They found significant differences between the content, frequency, and level of feedback given at more successful and less successful institutions. Future work examining the effect of online homework on student beliefs could build on this foundation. Perhaps Bloom's taxonomy or the taxonomy developed by White and Mesa (2014) could be used to classify the homework problems given on online homework sets, to test the hypothesis that problems on online homework sets skew more toward mechanical problems with simple answers than toward complex, open-ended problems that afford students the opportunity to engage more deeply with mathematics. In section 7.5, I reflect on personal teaching experiences that led me to this hypothesis.

Roth, Ivanchenko, and Record (2008) devised a scheme for coding log data to determine the types of problem-solving behaviors exhibited by students working and reworking WeBWorK problems. It might be possible to build on this approach to develop a scheme for determining the mathematical beliefs that are active in students' problem attempts.

7.4.2. Interventions targeting students' mathematical beliefs

As reported in section 5.2.2, there were no paired beliefs items that showed a significant positive shift in mathematical beliefs, either in the local or the national data. This raises several questions: Is it possible to come up with some kind of intervention that specifically targets mathematical beliefs? What might such an intervention look like? How successful could it be?

Boaler (2013) reported on several studies that have examined interventions targeting the development of a growth mindset. A student with a growth mindset believes that intelligence can be learned, and that the brain can grow and develop through exercise; in contrast, a student with a fixed mindset believes that intelligence is a trait that is possessed in fixed quantities and cannot be increased. Approximately 40% of U.S. students show a growth mindset, approximately 40% show a fixed mindset, and the other 20% show characteristics of both. By teaching students about neuroplasticity, or the ability of the brain to physically change and develop new connections when exercised, students can be moved from a fixed to a growth mindset; the shift in mindset carries both immediate and long-lasting effects on students' performance in school.

This work is thought-provoking, as it shows that shifting students' beliefs about the nature of intelligence is both possible and productive. It may be that such interventions can be adapted to target more discipline-specific beliefs and attitudes about mathematics.

7.4.3. Extending analyses presented here

I collected a great deal of demographic data, including SAT scores, mathematical preparation, gender, race/ethnicity, and parental education, which did not factor substantially into my analysis. There are many interesting questions that can be explored with this data: How do demographic factors impact beliefs? persistence? performance? Further work could employ multiple regression to build a model for various outcome variables using various demographic factors as input.

The differential impact techniques employed in section 4.7 could be extended to many other groups. For instance, how did mean final exam scores in the four classes compare across levels of socioeconomic status? across race/ethnicity? Further, these techniques could be applied to different outcome variables. For instance, how did beliefs in the four classes compare across various groups?

7.4.4. Psychometric properties of the CCI

The four classes were not distinguishable by their performance on the CCI, which was assumed to be a measure of conceptual understanding. So, either there were no significant differences in conceptual understanding, or the instrument is not a reliable measure of conceptual understanding. Against the first hypothesis, I have the analyses reported in section 4.6.2, which indicate that there were significant differences in performance on the conceptual portions of the final exam. Thus, might it be the case that the CCI is not a reliable instrument?

I have conducted some preliminary analyses of the CCI data I collected, using item response theory. The preliminary results appear to indicate problems with the psychometric properties of the CCI: the instrument does not appear to be unidimensional, as a concept inventory should normally be; several items do not adequately differentiate between students of low and high ability; and several items do not have plausible-enough distractors. Additionally, although the developer of the CCI has reported that detailed psychometric evaluations have been conducted (see Epstein, 2006), no validation studies have been published in peer-reviewed journals. A team of researchers from several universities, of which I am a part, are pooling CCI data collected in studies on students' understanding of calculus to investigate its psychometric properties.

Even if we find that the performance of the instrument is not good, it is a useful foundation on which better instruments can be built. In particular, to address the question of unidimensionality, the instrument could be fragmented into a number of more clearly unidimensional pieces, perhaps assessing student understanding of function, derivative, integral, and limit. Additionally, individual items with poor performance can be refined to have more plausible distractors and better discrimination between students of low and high ability.

7.5. Personal reflections on teaching calculus for life scientists

During the past academic year, I taught a year-long course entitled Calculus for the Life Sciences, a calculus course with a particular emphasis on mathematical modeling. The two-semester sequence included a review of functions, an introduction to modeling and measuring error, discrete dynamical systems, limits, derivatives (including applications such as optimization), differential equations (including numerical solutions via Euler's method), integration, and several integration techniques. It thus covered approximately the same set of material as one semester of mainstream calculus, with the addition of mathematical modeling topics. Each topic was linked to the solution of problems that arise in mathematical modeling. The course also includes a computer lab component, in which students used Excel and Maple to model sets of data. An effort is currently underway to replace the two-semester sequence with one semester covering essentially the same material.

Courses like this are another plausible way toward a lean and lively calculus. I am not arguing for fragmenting mainstream calculus into many different calculus courses tailored to the interests of each discipline; as discussed in section 7.2.2, in my mind, this is inadvisable and fraught with logistical problems. However, the practice of grounding each concept in realistic problems is a way to help students see the applicability of calculus. Never, during my year of teaching Calculus for the Life Sciences, did a student ask me when we were ever going to use this. I thus hypothesize that a fruitful area for future work in creating and assessing leaner and livelier calculus curricula is to build on modeling real-world data.

The Calculus for the Life Sciences classes I taught used WeBWorK as an online homework system. During the first semester, I had approximately 80 students in my class, and I was thus able to assign and grade occasional paper-and-pencil homework assignments. During the second semester, I had approximately 240 students, so grading that many paper-and-pencil assignments would have been unfeasible.

I regretted not being able to give pencil-and-paper assignments in the second semester, because I felt that the pencil-and-paper assignments I gave during the first semester exposed students to creative, open-ended parts of mathematics. For instance, one assignment asked them to choose a city, find average monthly temperature data for that city, and fit a trigonometric model to the data. They then compared this model to a model for another city that we created during class time, and I asked them to use the differences in the models to explain which city they would rather live in. I offered bonus points for doing something interesting outside the requirements of the assignment, and many students did interesting things, like graphing the two models together to show their differences graphically, evaluating the error of their model, or using Excel's Solver feature to increase the goodness of fit of their model.

In contrast to the types of problems I gave in paper-and-pencil assignments, I felt that the problems assigned via WeBWorK were largely mechanical, having one correct closed-ended answer; there was no way for me to program WeBWorK to grade student attempts to "do something interesting." It is thus my hypothesis that online homework systems lend themselves more to mechanical, closed-ended problems than to creative, open-ended problems, and that this might be the mechanism behind students developing instrumental approaches to their homework assignments.

7.6. Concluding remarks

This dissertation has contributed to the field's understanding of what works, and how, in teaching college calculus. By providing the beginnings of a list of best practices for the inverted model, and by discussing ways to strengthen the affordances of innovative pedagogical strategies, I have responded to PCAST's (2012) recommendation to "catalyze widespread adoption of empirically-validated teaching practices" (p. 16).

This study is fundamentally a mixed-methods study, rather than a concatenation of a quantitative study and a qualitative study. I view the qualitative and quantitative analyses as mutually informing and reflexively linked. The story of my analysis of the Inverted class illustrates this link. After reading the literature on the inverted model, I had expected that students in the Inverted class would outperform those in the other classes. However, the quantitative data revealed that this was not the case, and drove me to the qualitative data to find explanations. The student concerns voiced in the qualitative data drove me back to the literature, where I found that there were consistencies in the successful implementations that were not shared by this implementation.

By combining quantitative and qualitative analyses, I was able to ask and answer questions that I would not have known to ask or been able to answer without either of the sets of data. Thus, the quantitative and qualitative analyses worked together to yield something greater than the sum of their parts.

The practical benefit for me of conducting a mixed-methods study is that I was able to develop both quantitative and qualitative research skills. I hope to be able to use these skills in contributing to further research in the important field of student success in calculus.

Appendix A: Student Start-of-Term and End-of-Term Surveys

Characteristics of Successful Programs in College Calculus - Student Start

You have been selected to be part of a national survey of calculus instruction in colleges and universities across the United States. This research project is conducted by the Mathematical Association of America. Your answers are important to help us determine who is taking calculus, why they are taking it, and what mathematics background they bring to this course. At the end of this term, you will be asked to reflect on your experiences in this course.

Your participation is voluntary and does not affect your course grade. At the end of the survey you will have the option of providing us with your e-mail address. If necessary, your e-mail will allow us to arrange a telephone call with you to clarify any responses that are unclear. All information that you submit will be held in complete confidence. All personal information such as your e-mail address and telephone number will be disposed of prior to analysis. By continuing on to complete the survey you consent to participate in this study.

If you have any questions about this project, please call Olga Dixon at (202)-319-8498 or via e-mail odixon@maa.org.

I. High School Experience

1. My placement in calculus was determined by ***(Mark all that apply):***

- ☐ My ACT or SAT score
- ☐ My score on a placement exam
- ☐ My successful completion of prerequisite courses
- ☐ My AP exam score
- ☐ Don't know

2. Did you take the SAT exam?

- ○ Yes
- ○ No

3. My SAT scores were:

SAT critical reading	
SAT mathematics	

4. Did you take the SAT Subject Test in mathematics?

- ○ Yes
- ○ No

5. What was your score on the:

SAT Subject Test, Mathematics Level 1	
SAT Subject Test, Mathematics Level 2	

Characteristics of Successful Programs in College Calculus - Student Start

6. Did you take the ACT exam?

- ◯ Yes
- ◯ No

7. My ACT scores were:

Composite score

Mathematics score

8. My mathematics courses in high school have prepared me to:

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
complete calculations without a calculator	◯	◯	◯	◯	◯	◯
solve word problems	◯	◯	◯	◯	◯	◯
factor expressions	◯	◯	◯	◯	◯	◯
solve equations	◯	◯	◯	◯	◯	◯
solve inequalities	◯	◯	◯	◯	◯	◯

9. The teacher of my last mathematics course in high school:

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
lectured most of the time	◯	◯	◯	◯	◯	◯
primarily showed us how to get answers to specific questions	◯	◯	◯	◯	◯	◯
frequently had us work in groups	◯	◯	◯	◯	◯	◯
frequently had us solve challenging problems	◯	◯	◯	◯	◯	◯
cared that I was successful in the course	◯	◯	◯	◯	◯	◯

10. Please select the most appropriate answer:

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
I am comfortable in using a graphing calculator	◯	◯	◯	◯	◯	◯
I am comfortable in using a computer algebra system (e.g., Maple, MATLAB)	◯	◯	◯	◯	◯	◯

Characteristics of Successful Programs in College Calculus - Student Start

11. My high school mathematics teachers used an electronic response system (such as clickers) to poll students during class

- ◯ Yes
- ◯ No

12. In high school I was allowed to use graphing calculators on exams

- ◯ Always
- ◯ Sometimes
- ◯ Never

13. In high school I was allowed to use calculators that performed symbolic operations on exams (e.g., TI-89, TI-92)

- ◯ Yes
- ◯ No
- ◯ Sometimes

14. Which of the following calculus courses were offered at your high school? ***(Mark all that apply)***

- ☐ Calculus (Non AP)
- ☐ AP Calculus AB
- ☐ AP Calculus BC
- ☐ International Baccalaureate (IB) Calculus
- ☐ Online calculus course
- ☐ Calculus course taught at my high school for which students also received two-year college credit
- ☐ Calculus course taught at a two-year college for which students also received high school credit
- ☐ Calculus course taught at my high school for which students also received four-year college or university credit
- ☐ Calculus course taught at a four-year college or university for which students also received high school credit

Characteristics of Successful Programs in College Calculus - Student Start

15. For each (Non-AP) mathematics course listed below that you took in high school, please indicate the course level, your grade in school when taking the exam, and your final grade ***(If you did not take a particular course, leave that row blank)***

	Course Level	Grade Level in High School	Final Grade
Geometry			
Algebra II			
Integrated Math			
Pre-Calculus			
Trigonometry			
Statistics (Non AP)			
Calculus (Non AP)			
Other course taken senior year			

16. Did you take any AP Calculus or AP Statistics in high school?

- ◯ Yes
- ◯ No
- ◯ I don't know

17. For each of the following AP course that you took in high school, please indicate the score you earned on the exam, your grade in school when taking the exam, what grade you earned in the course, and the gender of the teacher

	AP Exam Score (if taken)	Grade Level in High School	Final Grade	Gender of Teacher
Calculus AB				
Calculus BC				
AP Statistics				

18. Did you take a calculus course in COLLEGE prior to this one?

- ◯ Yes
- ◯ No

Characteristics of Successful Programs in College Calculus - Student Start

19. Where was your previous college calculus course taken?

- ◯ At this college or university
- ◯ At another 4-year college or university
- ◯ At another 2-year college

20. What was the delivery mode of the calculus course you completed in college prior to this one? *(Mark all that apply)*

- ☐ Online
- ☐ Through correspondence
- ☐ Face-to-face with an instructor

21. Why are you taking this course again? *(Mark all that apply)*

- ☐ It did not count toward the credits I need
- ☐ I passed, but I need/want a higher grade (e.g., for my major)
- ☐ I did not pass the course
- ☐ I dropped the class
- ☐ I wanted to get a better grade
- ☐ I wanted to improve my understanding of calculus
- ☐ My college advisor told me to

22. Did you take a pre-calculus course in COLLEGE prior to this course?

- ◯ Yes
- ◯ No

23. Where was your previous college precalculus course taken?

- ◯ At this college or university
- ◯ At another 4-year college or university
- ◯ At another 2-year college

Characteristics of Successful Programs in College Calculus - Student Start

24. What was the delivery mode of the precalculus course you completed in college prior to this one? *(Mark all that apply)*

- ☐ Online
- ☐ Through correspondence
- ☐ Face-to-face with an instructor

25. What grade do you expect in this calculus course?

○ A ○ B ○ C ○ D ○ F

26. Do you intend to take Calculus II?

- ○ Yes
- ○ No
- ○ I don't know yet

27. How important is a good grade in this course in influencing your decision whether or not to take Calculus II?

- ○ Very important
- ○ Important
- ○ Slightly Important
- ○ Slightly Unimportant
- ○ Unimportant
- ○ Not important at all

28. Is Calculus II required for your major?

- ○ Yes
- ○ No
- ○ I don't know

II. Attitude Toward Mathematics

Characteristics of Successful Programs in College Calculus - Student Start

29. Please select the most appropriate answer.

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
I believe I have the knowledge and abilities to succeed in this course.	◯	◯	◯	◯	◯	◯
I am confident in my mathematical abilities.	◯	◯	◯	◯	◯	◯
I understand the mathematics that I have studied.	◯	◯	◯	◯	◯	◯
I enjoy doing mathematics.	◯	◯	◯	◯	◯	◯

30. If I take another calculus course after this one, it will be because *(Mark all that apply)*

- ☐ it is required
- ☐ I want to

In the following questions, choose number:
1 if you completely agree with option **A**
2 if you do not completely agree with option **A**, but agree with option **A** more than option **B**
3 if you do not completely agree with option **B**, but agree with option **B** more than option **A**
4 if you completely agree with option **B**

31. Please select the most appropriate response below:

	A Not at all certain 1	2	3	B Very certain 4
How certain are you in what you intend to do after college?	◯	◯	◯	◯

32. Please select the most appropriate answer below:

	A I try hard to figure it out on my own 1	2	3	B I quickly seek help or give up trying 4
When experiencing a difficulty in my math class:	◯	◯	◯	◯

Characteristics of Successful Programs in College Calculus - Student Start

33. Please select the most appropriate answer below:

	A a natural part of solving the problem 1	2	3	B an indication of my weakness in mathematics 4
For me, making unsuccessful attempts when solving a mathematics problem is:	◯	◯	◯	◯

34. Please select the most appropriate answer below:

	A solve specific kinds of problems 1	2	3	B make connections and form logical arguments 4
My success in mathematics PRIMARILY relies on my ability to:	◯	◯	◯	◯

35. Please select the most appropriate answer below:

	A I understand the covered material 1	2	3	B I can do things the way the teacher wants 4
My score on my mathematics exam is a measure of how well	◯	◯	◯	◯

36. Please select the most appropriate answer below:

	A I would never take another mathematics course 1	2	3	B I would continue to take mathematics 4
If I had a choice,	◯	◯	◯	◯

37. Please select the most appropriate answer below:

	A memorize it the way it is presented 1	2	3	B make sense of the material, so that I understand it 4
When studying Calculus I in a textbook or in course materials, I tend to:	◯	◯	◯	◯

Characteristics of Successful Programs in College Calculus - Student Start

38. Please select the most appropriate answer below:

	A understand underlying mathematical ideas 1	2	3	B find answers to problems 4
When solving mathematics problems, graphing calculators or computers help me to:	◯	◯	◯	◯

39. Please select the most appropriate answer below:

	A work problems so students know how to do them 1	2	3	B help students learn to reason through problems on their own 4
The primary role of a mathematics instructor is to:	◯	◯	◯	◯

40. Please select the most appropriate answer:

	Strongly Disagree	Disagree	Mildly Disagree	Mildly Agree	Agree	Strongly Agree
Mathematics instructors should show students how mathematics is relevant	◯	◯	◯	◯	◯	◯

41. Please select the most appropriate answer for each of the following.

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
If I am unable to solve a problem within a few minutes, it is an indication of my weakness in mathematics.	◯	◯	◯	◯	◯	◯
Mathematics is about getting exact answers to specific problems.	◯	◯	◯	◯	◯	◯
The process of solving a problem that involves mathematical reasoning is a satisfying experience.	◯	◯	◯	◯	◯	◯

III. Demographic Information

Characteristics of Successful Programs in College Calculus - Student Start

42. Please indicate whether each of the following was born in the U.S.:

	Yes	No
You	◯	◯
Male Parent or Guardian	◯	◯
Female Parent or Guardian	◯	◯

43. What was the highest level of education for your male parent or guardian?

- ◯ Did not finish high school
- ◯ High school
- ◯ Some college
- ◯ Four years of college
- ◯ Graduate school

44. What was the highest level of education for your female parent or guardian?

- ◯ Did not finish high school
- ◯ High school
- ◯ Some college
- ◯ Four years of college
- ◯ Graduate school

***45. What was your home zip code at the time you graduated from high school?** *If you graduated from a high school outside the U.S., enter 00000*

***46. What is your birth date?**

Enter numbers as indicated > MM / DD / YYYY

***47. What is your Red ID?**

Characteristics of Successful Programs in College Calculus - Student Start

48. Do the following people see you as a person who is good at mathematics?

	Not at all 1	2	3	4	5	Very much 6
Yourself	○	○	○	○	○	○
Parents/Relatives	○	○	○	○	○	○
High school mathematics teacher	○	○	○	○	○	○
Friends	○	○	○	○	○	○

49. What is your gender?

○ Male ○ Female

50. What is your race? ***(Mark all that apply)***

☐ White
☐ Black
☐ Asian
☐ Pacific Islander
☐ American Indian or Alaska Native
☐ Other (please specify)

51. Are you of Hispanic origin?

○ Yes
○ No

52. Was English the primary language spoken in your household?

○ Yes
○ No

53. What year are you in college?

○ Freshmen
○ Sophomore
○ Junior
○ Senior
○ Graduate Student
○ Other

Characteristics of Successful Programs in College Calculus - Student Start

54. What is your current type of college enrollment?

- ◯ Full-time
- ◯ Part-time

55. To what degree was your home environment supportive of your studying math?

- ◯ Not at all
- ◯ Somewhat
- ◯ Strongly
- ◯ Very Strongly

56. Who encouraged you to take mathematics classes? ***(Mark all that apply.)***

- ☐ No One
- ☐ Mother/Female Guardian
- ☐ Father/Male Guardian
- ☐ Siblings
- ☐ Other Relative
- ☐ School Counselor
- ☐ Math Teacher
- ☐ Other Teacher
- ☐ Coach

57. Where did you attend high school?

- ◯ American school in the United States
- ◯ American school abroad
- ◯ Other (please specify)

58. Approximately how many hours per week do you expect to work at a job this semester/term?

- ◯ 0
- ◯ 1-5
- ◯ 6-10
- ◯ 11-15
- ◯ 16-20
- ◯ 21-30
- ◯ More than 30

59. Approximately how many hours per week do you expect to participate in organized extracurricular activities such as sports, college paper, or clubs this semester/term?

- ◯ 0
- ◯ 1-5
- ◯ 6-10
- ◯ 11-15
- ◯ 16-20
- ◯ 21-30
- ◯ More than 30

Characteristics of Successful Programs in College Calculus - Student Start

60. Approximately how many hours per week do you expect to spend preparing for your classes this semester/term (studying, reading, writing, doing homework or lab work, analyzing data, rehearsing, or other academic activities)?

- ◯ 0
- ◯ 1-5
- ◯ 6-10
- ◯ 11-15
- ◯ 16-20
- ◯ 21-30
- ◯ More than 30

Concerning Your Career Plans:

61. Which of the following BEST describes your current career goal?

62. Please select the appropriate response for each of the following:

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
I am anticipating difficulty paying for college.	◯	◯	◯	◯	◯	◯
In order to succeed in calculus at a college or university, I must have taken it before.	◯	◯	◯	◯	◯	◯

63. Optional: We may want to contact you to ask follow-up questions about your math-related experiences. All communications will be kept in the strictest confidence and your email will NOT be disclosed to any third party.

Your email address:

Characteristics of Successful Programs in College Calculus - Student End

This is a continuation of the national survey of calculus instruction in colleges and universities across the United States. This research project is conducted by the Mathematical Association of America. Whether or not you completed the survey at the start of the term, please complete this survey now. We want to know your thoughts about the calculus class you took this term. Your answers are important to help us determine what works and how calculus instruction can be improved for all students.

Your participation is voluntary and does not affect your course grade. At the end of the survey you will have the option of providing us with your e-mail address and your student ID. If necessary, your e-mail will allow us to arrange a telephone call with you to clarify any responses that are unclear. Your student ID will allow us to match your survey to your final course grade. All information that you submit will be held in complete confidence. All personal information such as your e-mail address, telephone number, and student ID will be disposed of prior to analysis. By continuing on to complete the survey you consent to participate in this study.

If you have any questions about this project, please contact Olga Dixon at (202) 319-8498 or via e-mail odixon@maa.org.

1. What grade do you expect (or did you receive) in this course?

○ A ○ B ○ C ○ D ○ F

2. Is Calculus II required for your intended major?

○ Yes ○ No ○ I'm not sure

3. Do you intend to take Calculus II?

○ Yes
○ No
○ I'm not sure

4. If you do not intend to take Calculus II, check all reasons that apply.

- ☐ I never intended to take Calculus II.
- ☐ I changed my major and now do not need to take Calculus II.
- ☐ My experience in Calculus I made me decide not to take calculus II.
- ☐ I have too many other courses I need to complete.
- ☐ To do well in Calculus II, I would need to spend more time and effort than I can afford.
- ☐ My grade in Calculus I was not good enough for me to continue to Calculus II.
- ☐ I do not believe I understand the ideas of Calculus I well enough to take Calculus II.

Characteristics of Successful Programs in College Calculus - Student End

5. When you started this class, did you intend to take Calculus II?

○ Yes ○ No ○ I wasn't sure

6. Select the appropriate response to each of the following:

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
This course has increased my interest in taking more mathematics.	○	○	○	○	○	○
I am confident in my mathematical abilities.	○	○	○	○	○	○
I am good at computing derivatives and integrals.	○	○	○	○	○	○
I am able to use ideas of calculus (e.g., differentiation, integration) to solve word problems that I have not seen before.	○	○	○	○	○	○
My previous math courses prepared me to succeed in this course.	○	○	○	○	○	○
I enjoy doing mathematics.	○	○	○	○	○	○
Mathematics is about getting exact answers to specific problems.	○	○	○	○	○	○

In the following questions, choose number:
1 if you completely agree with option **A**
2 if you do not completely agree with option **A**, but agree with option **A** more than option **B**
3 if you do not completely agree with option **B**, but agree with option **B** more than option **A**
4 if you completely agree with option **B**

7. Please select the most appropriate response below.

	A Not at all certain 1	2	3	B Very certain 4
How certain are you in what you intend to do after college?	○	○	○	○

8. Please select the most appropriate response below.

	A I try hard to figure it out on my own 1	2	3	B I quickly seek help or give up trying 4
When experiencing a difficulty in my math class:	○	○	○	○

Characteristics of Successful Programs in College Calculus - Student End

9. Please select the most appropriate response below.

	A a natural part of solving the problem 1	2	3	B an indication of my weakness in mathematics 4
For me, making unsuccessful attempts when solving a mathematics problem is:	○	○	○	○

10. Please select the most appropriate response below.

	A solve specific kinds of problems 1	2	3	B make connections and form logical arguments 4
My success in mathematics PRIMARILY relies on my ability to:	○	○	○	○

11. Please select the most appropriate response below.

	A I understand the covered material 1	2	3	B I can do things the way the teacher wants 4
My score on my mathematics exam is a measure of how well:	○	○	○	○

12. Please select the most appropriate response below.

	A I would never take another mathematics course. 1	2	3	B I would continue to take mathematics. 4
If I had a choice:	○	○	○	○

13. Please select the most appropriate response below.

	A memorize it the way it is presented. 1	2	3	B make sense of the material so I understand it. 4
When studying Calculus I in a textbook or course materials, I tend to:	○	○	○	○

Characteristics of Successful Programs in College Calculus - Student End

14. Please select the most appropriate answer below:

	A understand underlying mathematical ideas 1	2	3	B find answers to problems 4
When solving mathematics problems, graphing calculators or computers help me to:	○	○	○	○

15. Please select the most appropriate response below.

	A work problems so students know how to do them 1	2	3	B help students learn to reason through problems on their own 4
The primary role of a mathematics instructor is to:	○	○	○	○

16. Please select the most appropriate response below.

	A waited for a student to answer 1	2	3	B answered the question if no one responded quickly 4
When my Calculus instructor asked a question addressed to the whole class, s/he:	○	○	○	○

17. Please select the most appropriate response below.

	A solved the problem for me 1	2	3	B helped me figure out how to solve the problem 4
When I asked a question about a problem I was having difficulty solving, my instructor:	○	○	○	○

Characteristics of Successful Programs in College Calculus - Student End

18. My Calculus instructor:

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
asked questions to determine if I understood what was being discussed.	◯	◯	◯	◯	◯	◯
listened carefully to my questions and comments.	◯	◯	◯	◯	◯	◯
discussed applications of calculus.	◯	◯	◯	◯	◯	◯
allowed time for me to understand difficult ideas.	◯	◯	◯	◯	◯	◯
helped me become a better problem solver.	◯	◯	◯	◯	◯	◯
provided explanations that were understandable.	◯	◯	◯	◯	◯	◯
was available to make appointments outside of office hours, if needed.	◯	◯	◯	◯	◯	◯
discouraged me from wanting to continue taking Calculus.	◯	◯	◯	◯	◯	◯

19. During class time, how frequently did your instructor:

	Not at all 1	2	3	4	5	Very often 6
show how to work specific problems?	◯	◯	◯	◯	◯	◯
have students work with one another?	◯	◯	◯	◯	◯	◯
hold a whole-class discussion?	◯	◯	◯	◯	◯	◯
have students give presentations?	◯	◯	◯	◯	◯	◯
have students work individually on problems or tasks?	◯	◯	◯	◯	◯	◯
lecture?	◯	◯	◯	◯	◯	◯
ask questions?	◯	◯	◯	◯	◯	◯
ask students to explain their thinking?	◯	◯	◯	◯	◯	◯

20. How frequently did your instructor?

	Not at all 1	2	3	4	5	Very often 6
prepare extra material to help students understand calculus concepts or procedures?	◯	◯	◯	◯	◯	◯
require you to explain your thinking on your homework?	◯	◯	◯	◯	◯	◯
require you to explain your thinking on exams?	◯	◯	◯	◯	◯	◯
assign sections in your textbook for you to read before coming to class?	◯	◯	◯	◯	◯	◯

Characteristics of Successful Programs in College Calculus - Student End

21. My Calculus instructor:

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
made students feel nervous during class.	◯	◯	◯	◯	◯	◯
encouraged students to enroll in Calculus II.	◯	◯	◯	◯	◯	◯
acted as if I was capable of understanding the key ideas of calculus.	◯	◯	◯	◯	◯	◯
made me feel comfortable in asking questions during class.	◯	◯	◯	◯	◯	◯
encouraged students to seek help during office hours.	◯	◯	◯	◯	◯	◯
presented more than one method for solving problems.	◯	◯	◯	◯	◯	◯
did not speak English very well.	◯	◯	◯	◯	◯	◯
made class interesting.	◯	◯	◯	◯	◯	◯

22. Indicate how often the following occurred.

	Never	Some class sessions	About half the class sessions	Most class sessions	Every class session
My instructor assigned homework.	◯	◯	◯	◯	◯
Homework was collected (either hard copy or online).	◯	◯	◯	◯	◯
My instructor gave a short quiz.	◯	◯	◯	◯	◯
My instructor used technology.	◯	◯	◯	◯	◯

23. Which of the following technologies did you use during your calculus class? (*Check all that apply*)

- ☐ None
- ☐ Graphing Calculator
- ☐ Computers
- ☐ Clickers or some other electronic response system

Characteristics of Successful Programs in College Calculus - Student End

24. How frequently were the following technologies used *during* class?

	Never	Some class sessions	About half the class sessions	Most class sessions	Every class session
My instructor demonstrated mathematics with a graphing calculator.	◯	◯	◯	◯	◯
I used a graphing calculator.	◯	◯	◯	◯	◯
My instructor demonstrated mathematics with computer algebra system (e.g., Maple, Mathematica, MATLAB).	◯	◯	◯	◯	◯
I used a computer algebra system (e.g., Maple, Mathematica, MATLAB).	◯	◯	◯	◯	◯

25. How did you use technology during your class? (*Check only those that apply*)

- ☐ To find answers to problems.
- ☐ To understand underlying mathematical ideas.
- ☐ To check written answers after I worked them out by hand

26. How did your instructor use technology during your class? (*Check only those that apply*)

- ☐ To illustrate ideas.
- ☐ To find answers to problems.
- ☐ To check answers after we worked them out by hand.
- ☐ To illustrate motion/dynamic animations.

27. Does your calculator find the symbolic derivative of a function?

◯ Yes ◯ No ◯ N/A (I don't use a calculator)

28. Were you allowed to use a graphing calculator during your exams?

◯ Yes ◯ No

29. Assignments completed *outside of class* time were:

	Not at all 1	2	3	4	5	Very often 6
completed and graded online.	◯	◯	◯	◯	◯	◯
graded and returned to me.	◯	◯	◯	◯	◯	◯
returned with helpful feedback/comments.	◯	◯	◯	◯	◯	◯
submitted as a group project.	◯	◯	◯	◯	◯	◯
challenging but doable.	◯	◯	◯	◯	◯	◯

Characteristics of Successful Programs in College Calculus - Student End

30. The assignments completed *outside of class time* required that I:

	Not at all 1	2	3	4	5	Very often 6
solve word problems.	◯	◯	◯	◯	◯	◯
solve problems unlike those done in class or in the book.	◯	◯	◯	◯	◯	◯
use technology to understand ideas.	◯	◯	◯	◯	◯	◯

31. The *exam questions* required that I solve:

	Not at all 1	2	3	4	5	Very often 6
word problems.	◯	◯	◯	◯	◯	◯
problems unlike those done in class or in the book.	◯	◯	◯	◯	◯	◯

32. Please select the most appropriate response for each of the following.

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
My Calculus exams were a good assessment of what I learned.	◯	◯	◯	◯	◯	◯
My exams were graded fairly.	◯	◯	◯	◯	◯	◯
My homework was graded fairly.	◯	◯	◯	◯	◯	◯

33. During class:

	Never	Some class sessions	About half the class sessions	Most class sessions	Every class session
I contributed to class discussions.	◯	◯	◯	◯	◯
I was lost and unable to follow the lecture or discussion.	◯	◯	◯	◯	◯
I asked questions.	◯	◯	◯	◯	◯
I simply copied whatever was written on the board.	◯	◯	◯	◯	◯

34. How often did you do the following?

	Never	Once a month	A few times a month	Once a week	More than once a week
Read the textbook prior to coming to class.	◯	◯	◯	◯	◯
Visit your instructor's office hours.	◯	◯	◯	◯	◯
Use online tutoring.	◯	◯	◯	◯	◯
Visit a tutor to assist with this course.	◯	◯	◯	◯	◯

Characteristics of Successful Programs in College Calculus - Student End

35. Check the box that describes your level of agreement with the following statements.

	Strongly Disagree	Disagree	Slightly Disagree	Slightly Agree	Agree	Strongly Agree
The homework for the course helped me learn the material.	◯	◯	◯	◯	◯	◯
The textbook and/or class materials helped me learn the material.	◯	◯	◯	◯	◯	◯
The textbook or reading materials for the course were readable.	◯	◯	◯	◯	◯	◯
I completed all my assigned homework.	◯	◯	◯	◯	◯	◯

36. Did you meet with other students to study or complete homework outside of class?

◯ Yes ◯ No

37. Did you belong to a calculus study group organized by your instructor or department?

◯ Yes ◯ No

38. Does your math department or university provide a walk-in tutor center for mathematics?

◯ Yes ◯ No

39. Approximately how many hours per week did you work at a job this semester/term?

◯ 0
◯ 1-5
◯ 6-10
◯ 11-15
◯ 16-20
◯ 21-30
◯ More than 30

Characteristics of Successful Programs in College Calculus - Student End

40. Approximately how many hours per week did you participate in organized extracurricular activities such as sports, college paper, or clubs?

- ◯ 0
- ◯ 1-5
- ◯ 6-10
- ◯ 11-15
- ◯ 16-20
- ◯ 21-30
- ◯ More than 30

41. Approximately how many hours per week did you spend preparing for all classes (studying, reading, writing, doing homework or lab work, analyzing data, or other academic activities) this semester/term?

- ◯ 0
- ◯ 1-5
- ◯ 6-10
- ◯ 11-15
- ◯ 16-20
- ◯ 21-30
- ◯ More than 30

42. Approximately how many hours per week did you spend preparing for calculus (studying, reading, doing homework or lab work) this semester/term?

- ◯ 0
- ◯ 1-5
- ◯ 6-10
- ◯ 11-15
- ◯ 16-20
- ◯ 21-30
- ◯ More than 30

Characteristics of Successful Programs in College Calculus - Student End

***43. What is your home zip code at the time you graduated from high school?** *If you graduated from a high school outside the U.S., enter 00000*

***44. What is your birth date?**

MM DD YYYY

Enter numbers as indicated -> / /

Optional: We may want to contact you to ask follow-up questions about your math-related experiences. All communications will be kept in the strictest of confidence and your email will NOT be disclosed to any third party.

45. Your email address:

46. Is there anything else you want to tell us about your experience in Calculus I?

Appendix B: Student Focus Group Protocol

Thank you again for meeting with me to talk about your calculus class. As you may know, I am conducting a study comparing the four different calculus classes at your university. My goal is to better understand how students felt about the way their classes were taught, which can then lead to recommendations for similar institutions. We have a number of questions that we want to ask, but we want to keep the discussion somewhat informal and allow for you guys to let us know what is important for you. The interview should not take more than an hour. So let's get started.

Question
Q1. I wanted to start by asking everyone if they intend to take Calc 2 and why? Also, if your major does not require calc 2 would you take it anyway. So maybe we could just go around. *(Probe for major if not offered)*
Q3. What kinds of questions do you answer on exams/homework? o Applications (substantial or just basic word problems?) o Computational problems? ("Symbol pushing?") o Questions that ask for explanations or justifications? o Mostly problems that are just like examples explained in class? A. How are these questions similar or different than in other math courses you have taken? B. How important is it that you memorize what your teacher did in class on exams and homework? C. How do you study to prepare for the exams?
Q4. How often are you asked to solve a problem that you have never seen before? *(Probe for example, and when these happen)*
Now we want to talk about your actual Calculus class.
Q5. Please describe for me what a typical class looks like. o How often and how much does your instructor lecture?

- How interactive are the instructor's lectures?
- How often do you work in groups?
- What is group work like in your class?
- Do you work problems in class?
- Do you give presentations in class?
- What is your favorite part of the class?
- What is your least favorite part of the class?

A. Describe a typical week in class – what happens on different days?

- Class vs. section

Q6. What do you need to do to be successful in this class?

Q7. [Inverted classes] How did your experience in this class compare to your experiences in traditional classrooms?

- **Did your instructor explain why online video was being used?**
- **Do you think it was good for your learning?**
- **What did you like about it? What did you dislike about it?**
- **If other classes were offered using this model, would you sign up for them?**

Q8. How does your instructor use technology?

- Web to post text, problems and solutions?
- Online textbooks resources?
- Graphing calculators (with or without a computer algebra system)?

○ Computer algebra systems such as Maple, Wolfram alpha or Mathematica? ○ Interactive visualizations A. Did your instructor explain why they used technology? B. Do you think the way your instructor used technology was good for your learning? C. How do you use technology ? (inside and outside of class)
Q9. What type of things happen in class that help you learn calculus content? A. What types of things do you think could be done differently to improve your learning of the content?
Q10. Another area that is important for a successful calculus program are the resources that are available to students. If you need help with the class, where do you go? If that doesn't work, what then? What then?.. etc. ○ Tutoring center? ○ Faculty/TA Office hours? ○ Computer labs? ○ Study groups? ○ Review sessions? ○ Supplemental instruction? ○ Problem solving sessions? ○ Prep or bridge programs? ○ Wolphram Alpha? ○ Khan Academy?

A. How often do you use these resources? B. Which of these resources actually help you with your learning of calculus? How? C. How did you find out about these resources? D. Are there resources you wish were available to you that you that are not? Or resources you wish functioned differently?
Q13. How did you end up in Calculus I vs. taking pre-calculus or Calculus II? ○ Did you have an advisor? ○ Did you take a placement exam?
Q14. Are there other types of assignments that we haven't already talked about *(Probe if something important not mentioned yet)*? ○ Homeworks ○ Labs ○ Projects ○ Group assignments ○ Writing projects ○ Tests
Q15. Do you work with other students outside of class? A. Where do you go? B. *(if applicable)* How did you decide who to study with? C. *(if applicable)* How often do you work with other students? D. Is this typical of students in your major? E. Does your instructor encourage this?
Q16. Can you describe specific ways this class has helped you understand why Calculus is important to your field of study?

Q17. Has your experience in Calculus I increased your confidence in your ability to do mathematics (*Probe for elements from discussion that connect to this)*?   **Do you enjoy doing mathematics? What effect has your experience in Calculus I had on your enjoyment of mathematics?**
Q18. What would you say is your instructor's attitude towards students? A. Does your teacher seem to care about your learning? B. Does your teacher think that students are capable of understanding calculus? C. Do you think this is typical of the teachers in this math department?
Q20. We talked about a number of features your Calculus class. Are there things that strike you as important that we missed or didn't talk about?
END Thank you again for talking with us about Calculus. This has been very informative and interesting.

References

Ambrose, R., Clement, L., Philipp, R., & Chauvot, J. (2004). Assessing prospective elementary school teachers' beliefs about mathematics and mathematics learning: Rationale and development of a constructed-response-format beliefs survey. *School Science and Mathematics*, *104*(2), 56-69.

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of educational research*, *70*(2), 181-214.

Bergmann, J., & Sams, A. (2008). Remixing chemistry class. *Learning and Leading with Technology*, *1*, 22–27.

Boaler, J. (1998). Open and closed mathematics: Student experiences and understandings. *Journal for Research in Mathematics Education*, *29*(1), 41. doi:10.2307/749717

Bowers, J. S., & Nickerson, S. (2001). Identifying cyclic patterns of interaction to study individual and collective learning. *Mathematical Thinking and Learning*, *3*(1), 1–28. doi:10.1207/S15327833MTL0301_01

Bowers, J., & Zazkis, D. (2012). Do students flip over the "flipped classroom" model for learning college calculus? In L. R. Van Zoest, J.-J. Lo, & Kratky, J. L. (Eds.), *Proceedings of the 34th Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (849-852). Kalamazoo, MI: Western Michigan University.

Bressoud, D., Carlson, M., 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. doi:10.1080/0020739X.2013.798874

Bressoud, D., Carlson, M., Pearson, J. M., & Rasmussen, C. (2012). Preliminary results of the study on characteristics of successful programs in college calculus. *12th International Congress on Mathematics Education*. Seoul, Korea, July 8-15, 2012. Retrieved from http://www.maa.org/cspcc/TSG13_Bressoud.pdf

Brousseau, G. (1997). *Theory of didactical situations in mathematics.* Edited and translated by N. Balacheff, M. Cooper, R. Sutherland, & V. Warfield. Dordrecht: Kluwer.

Brown, A. L., & Campione, J. C. (1996). Psychological theory and the design of innovative learning environments: On procedures, principles, and systems. In L.

Schauble & R. Glaser (Eds.), *Contributions of instructional innovation to understanding learning* (pp. 289-325). Hillsdale, NJ: Erlbaum.

Carlson, M. (1999). The mathematical behavior of six successful mathematics graduate students: Influences leading to mathematical success. *Educational Studies in Mathematics*, *40*(3), 237–258.

Carlson, M., Bloom, I., & Glick, P. (2008). Promoting effective mathematical practices in students: insights from problem solving research. In M. P. Carlson, C. Rasmussen, M. P. Carlson, & C. Rasmussen (Eds.), *Making the connection* (pp. 275–288). Washington DC: Mathematical Association of America.

Carlson, M., Buskirk, T., & Halloun, I. (1999). Assessing college students' views about mathematics with the views about mathematics survey. Unpublished manuscript.

Carlson, M., Madison, B., & West, R. (2010). The Calculus Concept Readiness (CCR) instrument: Assessing student readiness for calculus. *arXiv preprint arXiv:1010.2719*. Retrieved from http://arxiv.org/abs/1010.2719

Carlson, M. P., & Bloom, I. (2005). The cyclic nature of problem solving: An emergent multidimensional problem-solving framework. *Educational Studies in Mathematics*, *58*(1), 45–75. doi:10.1007/s10649-005-0808-x

Characteristics of Successful Programs in College Calculus. (2013). Characteristics of successful programs in college calculus. Retrieved from http://www.maa.org/cspcc/

Cobb, P., & Yackel, E. (1996). Constructivist, emergent, and sociocultural perspectives in the context of developmental research. *Educational Psychologist*, *31*(3-4), 175–190. doi:10.1080/00461520.1996.9653265

Deslauriers, L., Schelew, E., & Wieman, C. (2011). Improved learning in a large-enrollment physics class. *Science*, *332*(6031), 862–864. doi:10.1126/science.1201783

Douglas, R. G. (ed.). (1986). *Toward a lean and lively calculus, MAA Notes 6*. Washington, DC: MAA.

Ellis, J., Kelton, M. L., & Rasmussen, C. (2014). Student perceptions of pedagogy and associated persistence in calculus. *ZDM,* 1-13. doi:10.1007/s11858-014-0577-z

Epstein, J. (2006). The Calculus Concept Inventory. In D. Deeds & B. Callen (Eds.), *Proceedings of the National STEM Assessment Conference,* Washington, D.C. (60-67).

Fennema, E., & Sherman, J. A. (1976). Fennema-Sherman mathematics attitudes scales: Instruments designed to measure attitudes toward the learning of mathematics by females and males. *Journal for Research in Mathematics Education*, *7*(5), 324–326.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *PNAS, 111.* doi:10.1073/pnas.1319030111

Gannod, G. C. (2007). WIP: Using podcasting in an inverted classroom. *Proceedings of the 37th IEEE Frontiers in Education Conference. IEEE*. Retrieved from http://www.colorado.edu/MCDB/MCDB5650/InvertedClassroom07.pdf

Gannod, G. C., Burge, J. E., & Helmick, M. T. (2008). Using the inverted classroom to teach software engineering. *Proceedings of the 30th international conference on Software engineering* (pp. 777–786). Retrieved from http://dl.acm.org/citation.cfm?id=1368198

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Goldin, G. A. (2004). Problem solving heuristics, affect, and discrete mathematics. *ZDM*, *36*(2), 56–60. doi:10.1007/BF02655759

Gresalfi, M. S. (2009). Taking up opportunities to learn: Constructing dispositions in mathematics classrooms. *The Journal of the Learning Sciences*, *18*(3), 327-369.

Gresalfi, M. S., Barnes, J., & Cross, D. (2012). When does an opportunity become an opportunity? Unpacking classroom practice through the lens of ecological psychology. *Educational Studies in Mathematics*, *80*(1-2), 249-267.

Hanson, K, Nunez, G., & Ellis, J. (2014, February). Beyond plug and chug: The nature of calculus homework at doctoral institutions. Poster presented at the Conference for Research on Mathematics Education, Denver, CO.

Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. *Research in Collegiate Mathematics Education, 3*, 234-283.

Hauk, S., & Segalla, A. (2005). Student perceptions of the web-based homework program WeBWorK in moderate enrollment college algebra classes. *Journal of computers in mathematics and science teaching*, *24*(3), 229-253.

Henningsen, M., & Stein, M. K. (1997). Mathematical tasks and student cognition: Classroom-based factors that support and inhibit high-level mathematical thinking and reasoning. *Journal for Research in Mathematics Education*, *28*, 524-549.

Herbst, P., & Kilpatrick, J. (1999). "Pour Lire" Brousseau. *For the learning of mathematics, (19)*1, 3-10.

Huntley, M. A., Rasmussen, C. L., Villarubi, R. S., Sangtong, J., & Fey, J. T. (2000). Effects of Standards-based mathematics education: A study of the Core-Plus Mathematics Project algebra and functions strand. *Journal for Research in Mathematics Education*, 328-361.

Jackiw , N. (1995). *The Geometer's Sketchpad.* [Computer software]. Emeryville, CA: Key Curriculum Press.

Jiang, Z., & McClintock, E. (1997). Using the geometer's sketchpad with preservice teachers. *MAA NOTES*, 129–136.

July, R. A. (2001). Thinking in three dimensions: Exploring students' geometric thinking and spatial ability with the Geometer's Sketchpad. *ProQuest ETD Collection for FIU*, 1–269.

Kazemi, E., & Stipek, D. (2001). Promoting conceptual thinking in four upper-elementary mathematics classrooms. *The Elementary School Journal*, *102*(1), 59. doi:10.1086/499693

Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *The Journal of Economic Education*, *31*(1), 30. doi:10.2307/1183338

Martin, J. R., & White, P. R. R. (2007). *The Language of Evaluation*. Basingstoke: Palgrave Macmillan. Retrieved from http://www.palgraveconnect.com/doifinder/10.1057/9780230511910

McClintock, E., Jiang, Z., & July, R. (2002). Students' Development of Three-Dimensional Visualization in the Geometer's Sketchpad Environment. ERIC/CSMEE Publications, 1929 Kenny Road, Columbus, OH 43210-1080. Tel: 800-276-0462 (Toll Free). Retrieved from http://www.eric.ed.gov/ERICWebPortal/detail?accno=ED471759

Mehan, H. (1979). *Learning lessons: social organization in the classroom*. Cambridge, Mass: Harvard University Press.

Mehan, H. (1979). 'What time is it, Denise?': Asking known information questions in classroom discourse. *Theory into Practice, 28*(4), 285-294.

Mesa, V., & Chang, P. (2010). The language of engagement in two highly interactive undergraduate mathematics classrooms. *Linguistics and Education*, *21*(2), 83–100. doi:10.1016/j.linged.2010.01.002

Moravec, M., Williams, A., Aguilar-Roca, N., & O'Dowd, D. K. (2010). Learn before Lecture: A Strategy That Improves Learning Outcomes in a Large Introductory Biology Class. *Cell Biology Education*, *9*(4), 473–481. doi:10.1187/cbe.10-04-0063

Musallam, R. (2010). *The effects of using screencasting as a multimedia pre-training tool to manage the intrinsic cognitive load of chemical equilibrium instruction for advanced high school chemistry students*. The University of San Francisco. Retrieved from http://www.flipteaching.com/assets/Dissertation_Musallam.pdf

National Council of Teachers of Mathematics. (1989). *Curriculum and evaluation standards for school mathematics.* Reston, VA: Author.

National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

National Research Council. (1996). *National science education standards.* Washington DC: National Academy Press.

Nickerson, S., & Bowers, J. (2008). Examining interaction patterns in college-level mathematics classes: A case study. In M. Carlson & C. Rasmussen (Eds.), *Making the connection: Research and teaching in undergraduate mathematics* (pp. 179–190). Washington, DC: MAA.

Olive, J. (1998). Opportunities to explore and integrate mathematics with the Geometer's Sketchpad. *Designing learning environments for developing understanding of geometry and space*, 395–417.

Overmyer, J. (2013, February). *The flipped classroom model for college algebra: Effects on student achievement.* Poster presented at the Conference on Research in Undergraduate Mathematics Education, Denver, CO.

Pajares, F., & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of counseling psychology*, *42*(2), 190.

Pascarella, E. T., & Terenzini, P. T. (1980). Predicting Freshman Persistence and Voluntary Dropout Decisions from a Theoretical Model. *The Journal of Higher Education*, *51*(1), 60. doi:10.2307/1981125

Pascarella, E. T., & Terenzini, P. T. (1991). *How college affects students*. K. A. Feldman (Ed.). San Francisco: Jossey-Bass.

Plake, B. S., & Parker, C. S. (1982). The development and validation of a revised version of the mathematics anxiety rating scale. *Educational and Psychological Measurement*, *42*(2), 551–557.

President's Council of Advisors on Science and Technology (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics.* Washington, D.C.: Office of Science and Technology Policy.

Purdy, D. C. (2000). Using The Geometer's Sketchpad to Visualize Maximum-Volume Problems. *The Mathematics Teacher*, *93*(3), 224–228. doi:10.2307/27971341

Rasmussen, C., & Ellis, J. (2013). Students who switch out of calculus and the reasons why they leave. In Martinez, M. & Castro Superfine, A (Eds.). *Proceedings of the 35th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education* (pp. 457-464). Chicago, IL: University of Illinois at Chicago.

Rasmussen, C., Yackel, E., & King, K. (2003). Social and sociomathematical norms in the mathematics classroom. *Teaching mathematics through problem solving: Grades*, 6–12.

Rocca, K. A. (2010). Student participation in the college classroom: An extended multidisciplinary literature review. *Communication Education*, *59*(2), 185-213.

Roth, V., Ivanchenko, V., & Record, N. (2008). Evaluating student response to WeBWorK, a web-based homework delivery and grading system. *Computers & Education*, *50*(4), 1462-1482.

Sonnert, G., Sadler, P., Sadler, S., & Bressoud, D. (2014). The impact of instructor pedagogy on college calculus students' attitude toward mathematics. Submitted to *International Journal of Mathematics Education for Science and Technology.*

Scher, D. (2000). Lifting the curtain: The evolution of the Geometer's Sketchpad. *The Mathematics Educator*, *10*(1), 42–48.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense-making in mathematics. In D. Grouws (Ed.), *Handbook for Research on Mathematics Teaching and Learning* (pp. 334–370). New York: MacMillan.

Schommer, M., Calvert, C., Gariglietti, G., & Bajaj, A. (1997). The development of epistemological beliefs among secondary students: A longitudinal study. *Journal of Educational Psychology*, *89*(1), 37–40. doi:10.1037/0022-0663.89.1.37

Schommer-Aikins, M., Duell, O. K., & Hutter, R. (2005). Epistemological Beliefs, Mathematical Problem-Solving Beliefs, and Academic Performance of Middle School Students. *The Elementary School Journal*, *105*(3), 289–304. doi:10.1086/428745

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

Seymour, E. (2006). *Testimony offered to the Research Subcommittee of the Committee on Science of the U.S. House of Representatives hearing on Undergraduate*

Science, Mathematics, and Engineering Education: What's Working? Available from http://commdocs.house.gov/committees/science/hsy26481.000/hsy26481_0f.htm

Sinclair, N., Zazkis, R., & Liljedahl, P. (2003). Number Worlds: Visual and Experimental Access to Elementary Number Theory Concepts. *International Journal of Computers for Mathematical Learning*, *8*(3), 235–263. doi:10.1023/B:IJCO.0000021780.01416.61

Skemp, R. R. (1976). Instrumental understanding and relational understanding. *Mathematics Teaching*, *77*, 20-26.

Steen, L. A. (1987). *Calculus for a New Century: A Pump, Not a Filter. Papers Presented at a Colloquium (Washington, DC, October 28-29, 1987). MAA Notes Number 8*. Washington, DC: Mathematical Association of America.

Stipek, D., Salmon, J. M., Givvin, K. B., Kazemi, E., Saxe, G., & MacGyvers, V. L. (1998). The value (and convergence) of practices suggested by motivation research and promoted by mathematics education reformers. *Journal for Research in Mathematics Education*, *29*(4), 465. doi:10.2307/749862

Strang, G. (1991). *Calculus.* Wellesley, MA: Wellesley-Cambridge Press.

Strauss, A. L. (1987). *Qualitative Analysis for Social Scientists*. Cambridge: Cambridge University Press. Retrieved from http://ebooks.cambridge.org/ref/id/CBO9780511557842

Talbert, R. (2014). Inverting the linear algebra classroom. *PRIMUS*, *24*(5), 361-374. doi:10.1080/10511970.2014.883457

Tallman, M., & Carlson, M. P. (2012). A characterization of calculus I final exams in U.S. colleges and universities. In *Proceedings of the 15th annual conference on research in undergraduate mathematics education* (pp. 217–226). Portland, OR: Portland State University.

Tinto, V. (1975). Dropout from higher education: a theoretical synthesis of recent research. *Review of Educational Research*, *45*(1), 89. doi:10.2307/1170024

Tinto, V. (1997). Classrooms as communities: exploring the educational character of student persistence. *The Journal of Higher Education*, *68*(6), 599. doi:10.2307/2959965

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Wasserman, N., Norris, S., & Carr, T. (2013, February.) *Comparing a "flipped" instructional model in an undergraduate Calculus III course.* Paper presented at the Conference on Research in Undergraduate Mathematics Education, Denver, CO.

Watson, A. (2003). Opportunities to learn mathematics. *Mathematics education research: Innovation, networking, opportunity*, 29–38.

Wood, T., Williams, G., & McNeal, B. (2006). Children's mathematical thinking in different classroom cultures. *Journal for Research in Mathematics Education*, *37*(3), 222–255.

Yackel, E., & Rasmussen, C. (2003). Beliefs and norms in the mathematics classroom. *Beliefs: A Hidden Variable in Mathematics Education?*, 313–330.

Zazkis, D. (2013). *Calculus students' representation use in group-work and individual settings* (Unpublished doctoral dissertation). San Diego State University, San Diego, CA.

Zhang, G., Carter, R., Thorndyke, B., Anderson, T. J., and Ohland, M. W. (2003). Are engineering students different from others? In *Proceedings of the American Society for Engineering Education, Southeast.*