Title
Explanation and Reflection in Calculus Learning

Author
Reinholz, Daniel Lee

Publication Date
2014

Peer reviewed|Thesis/dissertation

Explanation and Reflection in Calculus Learning

by

Daniel Reinholz

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy

in

Science & Mathematics Education

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Alan Schoenfeld, Chair
Michael Ranney
Maryl Gearhart

Spring 2014

Explanation and Reflection in Calculus Learning

Abstract

Explanation and Reflection in Calculus Learning

by

Daniel Reinholz

Doctor of Philosophy in Science & Mathematics Education

University of California, Berkeley

Professor Alan Schoenfeld, Chair

This dissertation focuses on explanation and reflection in mathematics and offers an instructional intervention for supporting student learning. Explanation and reflection are considered hallmarks of deep understanding; they are also tools for promoting learning. The primary context of study was calculus, because it is often considered a gatekeeper that prevents students from accessing higher-level mathematics and Science, Technology, Engineering, and Mathematics (STEM) careers. Simultaneously, there is concern that even students who are successful (in the conventional sense) lack deep mathematical understanding. I collected both conventional and deeper performance measures, because both have implications for learning and performance.

Over three semesters of instruction I used design-based research methods to iteratively develop and refine an intervention. The intervention, Peer-Assisted Reflection (PAR), used peer-review and self-reflection as means to promote explanations. The core PAR activities required students to: (1) engage in meaningful problems, (2) reflect on their own work, (3) analyze a peer's work and give and receive feedback, and finally (4) revise their own work based on insights gained throughout this cycle. PAR was supported by other aspects of the instructional environment, such as adequate training and opportunities for students to explain their ideas regularly during class sessions.

The first semester (pilot study) took place in an introductory algebra classroom in a community college; the second and third semesters (Phases I and II) took place in an introductory college calculus course in a research university. The purpose of the pilot study was to refine theoretical principles to create an effective instructional intervention. Phase I was the first full implementation of the intervention, and Phase II served to replicate and extend the findings from Phase I. During Phases I and II quasi-experimental methods were used to compare students in an experimental section to students in parallel sections of the same course. I collected the following data: students' common exams, pre- and post-surveys about beliefs, student interviews, video observations of class sessions, copies of students' PAR assignments, and audio records of student conversations.

Students in the experimental sections were more successful in calculus than their coun-

terparts in the comparison sections. During Phase I the experimental success rate (A's, B's, and C's in the course) was 13% higher than the comparison section. The difference in success was 23% during Phase II, due to iterative refinements of the intervention. These improvements were statistically significant on common department exams in Phase I (average exam score 73.03% vs. 66.84% and 67.32% in the two comparison conditions) and Phase II (average exam score 75.20% vs. 64.17%). Improvements were evident on both conceptual and procedural problems. Students in the experimental sections also improved their written explanations and demonstrated improved persistence in problem solving.

In sum, engaging in PAR helped students improve their: success rates, exam scores, written explanations, and persistence. Because PAR is not grounded in specific mathematical (or formal) content, it should be a broadly applicable intervention for improving student learning in a variety of contexts. The findings of this dissertation also contribute to a theoretical understanding of peer-analysis and self-reflection as tools to promote explanations. For example, explaining to an authentic peer audience helped students focus on their communication, rather than just finding the correct answer. The findings also provide further evidence of the importance of conceptual understanding; by engaging in PAR, an intervention specifically targeted at improving explanations, students also improved their success on measures of procedural understanding.

Contents

List of Figures

List of Tables

Acknowledgments

I am indebted to Alan Schoenfeld for supporting me to pursue this project. Years ago, Alan told me that if my dissertation project didn't represent something I was truly passionate about, then he would have failed me as an adviser. He taught me to "fail early, fail often," and helped me deal with the uncomfortable uncertainty of finding a research site. Alan's support and insightful feedback were invaluable throughout all phases of the study.

I would also like to thank Michael Ranney for his guidance and friendship. I fondly remember many cups of tea shared while discussing research ideas. Maryl Gearhart provided countless insights and encouragement throughout the dissertation process. Her attention to detail was invaluable. I would also like to thank Xiaoxia Newton for her feedback and support on my prospectus.

Ken Klopfenstein is the person responsible for pulling me into the wonderful world of mathematics. He also helped me find a research site when everything else fell through. For his support and friendship, I will always be grateful.

I would like to thank Gerhard Dangelmayr for providing the moral and financial support to conduct my study. Mary Pilgrim made countless contributions in implementation and analysis, and was a *fancy nice* person to work with. Thanks to Hilary Freeman, Luke Bayens, Steve Benoit, Tim Hodges, Bahaudin Hashmi, and countless others in the mathematics department for all of their contributions. A sincere thanks to Jeff Grabhorn for working as my research assistant and John Ramsey for providing PAR grading support.

Thank you to Malcolm Swan from the Shell Centre and Mara Landers and Julie Von Bergen from Los Medanos for their support on the initial versions of my design.

Functions and the Reasoning Group both played pivotal roles in helping me learn to do research, and were so kind and patient to pilot test countless instruments for me. I thank my other friends and family as well for playing a similar role.

Thank you to Dan Brake for his help generating beautiful visualizations of PAR conversations that will hopefully find their way into a paper after my dissertation.

Bill Zahner introduced me to *Calculus Problems for a New Century*, which was perhaps the most wonderful source of calculus problems I have ever seen. Dick Stanley helped me improve my initial tasks through many engaging conversations.

Finally, my sincerest gratitude to my partner Suparna for all of her support along the way. A wise person once told me that each year you spend in graduate school is like two years for your spouse. Thank you for being with me through a *decade* of PhD studies.

1 Introduction

Numerous mathematical domains are widely recognized as difficult for learners: e.g., ratio and proportion, negative numbers, algebra, calculus, and proof. The present study focuses on two of these: algebra (in the pilot work), and calculus (in the main experiments). This dissertation is premised on the notion that these difficulties cannot be avoided, but rather, that students must be supported to grapple with and overcome them; learning that sidesteps these fundamental difficulties is unlikely to be robust.

A considerable body of literature addresses the learning of specific mathematical concepts. One could imagine combining all of these results (e.g., on limits, derivatives) to create a course of instruction, but this approach has limitations and is unlikely to generalize to a larger population of instructors. Instead, I focus on methods that should be more likely to generalize across domains be accessible to many instructors.[1] Specifically, I focus on explanation and reflection, which are widely recognized as hallmarks of deep understanding, and are essential to promoting self-guided learning. The logic of the intervention is to: (1) provide access to meaningful problems, and (2) create a supportive learning environment in which students can engage them; both are essential.

This dissertation studies took place over three semesters, in which I iteratively designed and implemented an intervention for promoting deep understanding. After reviewing the relevant literature, I dedicate a chapter to each of these semesters. The first semester (pilot study) took place in an elementary algebra class in a community college, and the next two semesters (Phases I and II) took place in introductory college calculus. The focus of the pilot study was to translate theoretical principles into a working intervention for promoting explanation and reflection. Building on these results, the purpose of Phase I was to test a full implementation of the intervention and its impact on student learning. Phase II was intended to replicate the findings from Phase I, and extend them by examining the impact of iterative improvements to the design. Chapters 3, 4, and 5 focus primarily on the evolution of the design, with some supporting analysis of student exam scores. I withhold analyses of student explanations until Chapter 6 to facilitate comparisons of performance between Phases I and II, and to illustrate the impact of iterative improvements to the design. In Chapter 7, I outline the learning mechanisms that appeared to make PAR a meaningful activity for students. I close with a summary and discussion of implications.

[1] At time of writing, the methods have been used in two contexts beyond the dissertation: introductory ordinary differential equations, and introductory mechanics taught from a modeling approach.

1.1 History and Context

Each fall semester, nearly 85,000 students (28%) fail to successfully complete introductory college calculus, even though more than half of them have already studied calculus in high school (Bressoud, Carlson, Mesa, & Rasmussen, 2013). This lack of success persists despite a nationwide call over 2 decades ago to make calculus a "pump, not a filter" (L. A. Steen, 1988). This prompted numerous studies, most notably the group of studies known as the calculus reform movement (Hurley, Koehn, & Ganter, 1999), and honors-style calculus courses based on the Emerging Scholars Program (described below) aimed at improving success for minority students (Fullilove & Treisman, 1990). After these landmark efforts, calculus continues to be an area of research interest, but the number of studies has waned. Moreover, many of the studies still focus on calculus reform (e.g., Star & Smith, 2006; Roddick, 2003) and the Emerging Scholars Program (e.g., Burmeister, Kenney, & Nice, 1996; Moreno, Muller, Asera, Wyatt, & Epperson, 1999).

In general, interventions have focused on the tasks that students engage with and the conditions under which the students engage them. Many reform courses utilized "real-world problems" and introduced either technology or collaborative group work. The results of the studies have been mixed. Some of the interventions have been relatively successful, while others have had negative impacts on students. It is difficult to compare studies because they each used their own outcome measures. In their nationwide survey, Bressoud et al. (2013) found standard calculus exams were comprised of 71-88% procedural questions (while instructors self-reported a perception of 40-70%). Given that most calculus exams are very procedural, one might question the measures used by many of the studies. It is also difficult to generalize the results of the calculus reform movement, because most studies had no measure for the quality of the two variables that were changed: (a) the tasks, and (b) the structures used to promote collaborative learning. Because most of the studies did not focus explicitly on implementation itself, it is impossible to say much about these important dimensions. The need for more careful documentation of the intervention prompted the use of design-based research methodology for the present study (Brown, 1992; Cobb, Confrey, diSessa, Lehrer, & Schauble, 2003; Collins, Joseph, & Bielaczyc, 2004).

I now review some of the factors that make calculus so difficult and consider notable attempts to address them. Finally, I address the issue of choosing quality tasks; supports for student learning will be further discussed in the next chapter.

1.1.1 The Calculus Filter

Calculus is full of conceptual difficulties. Limits (e.g. Tall, 1992; Davis & Vinner, 1986) and functions[2] (e.g. Oehrtman, Carlson, & Thompson, 2008) are extremely difficult for students to understand. Students struggle with object-process duality of functions (Sfard, 1991), confusing functions and equations, thinking all functions can be algebraically-defined, and

[2]Although students have some exposure to functions before calculus, functions are usually peripheral to the curriculum, whereas in calculus they are a central object of study.

interpreting functions intuitively, as well as other issues (Oehrtman et al., 2008). Limits introduce linguistic and conceptual difficulties, such as understanding *arbitrarily large or small*, terms like *tends to* or *approaches*, and whether or not a limit can actually be reached (Tall, 1992). Calculus is especially difficult because intuitive understandings of concepts (e.g., of continuity) often contradict mathematical formalism (Tall & Vinner, 1981; Tall, 1990).

Entering college students have generally experienced superficial, rote learning in K-12 mathematics classrooms for years (e.g. Aud et al., 2011; Schoenfeld, 1991, 1992), which creates gaps in knowledge and reinforces unproductive learning dispositions (Schoenfeld, 1988). Some unproductive beliefs students develop are: (a) math is only for geniuses, (b) all problems have a single right or wrong answer, and (c) problems are generally solved in less than two minutes or not at all; these beliefs prevent students from engaging in mathematics in deep and meaningful ways. Despite the failures of traditional mathematics classrooms, there is growing evidence of the efficacy of reform curricula and teaching practices to improve student understanding (Boaler & Staples, 2005; Senk & Thompson, 2003).

1.1.2 The Calculus Reform Movement

The reform movement was fostered by a 1986 conference at Tulane University, focused on creating a "lean and lively calculus" (Douglas, 1986). A year later, at the *Calculus for a New Century* conference (Hurley et al., 1999), mathematics educators were given the famous charge to make calculus a "pump, not a filter." Calculus, a great achievement of human understanding, should impel students to want to learn *more* mathematics and science, rather than pushing them to pursue other majors, or so went the argument.

The calculus syllabus was reformed, due to widespread agreement that the current calculus syllabus was a mile wide but an inch deep (Schoenfeld, 1995). This new syllabus generally emphasized multiple representations of functions (e.g., algebraic, graphical, numeric). Reformed courses often utilized technology (e.g., hand-held calculators, computer labs, programming) and collaborative learning. As of 1994, reform calculus texts had sold over 80,000 copies, and in a spring 1994 survey, 68% of respondents indicated at least modest reform efforts (Schoenfeld, 1995). I now consider a few projects that are emblematic of the larger movement.

The University of Connecticut introduced a weekly computer lab session and converted its discussion session to group problem solving (Hurley et al., 1999). Students worked in collaborative groups of 3-5 members in both environments. Parallel sessions of reform and traditional courses were run with common exams for 6 years, and of the 7 semesters reported on during this period, the experimental section always outperformed the traditional section, ranging from 1%-10% differences (Hurley et al., 1999); this was a small positive effect on traditional outcomes. The authors note that a similar approach at Dartmouth University, resulted in no significant difference in the performance between traditional and reform classes.

As a part of Project CALC classes at Duke University were reformed to consist of three regular class periods with a weekly 2-hour lab session (Bookman & Blake, 1996).

Project CALC utilized real-world problems, activities, group work, writing, and technology. The findings of a multi-year evaluation were generally positive, with improved engagement, problem-solving skills, and increased enrollment in subsequent math-related courses (Bookman & Friedman, 1999). Nevertheless, students in traditional sections still had better computational skills (not statistically significant), and higher GPAs overall (statistically significant).

Calculus & Mathematica was an intervention that focused on teaching calculus through Mathematica, a Computer Algebra System (CAS). Depending on the institution, common final exams have favored both reform and traditional practices (Roddick, 2001). For example, at Ohio State University, traditional students were found to do much better in differential equations (mean GPA 2.7 vs 2.06), while reform students were found to have a higher GPA in introductory physics (mean GPA 2.83 vs 2.47) (Roddick, 2001). Accordingly, it was difficult to evaluate the impact of the reformed course.

In total, NSF funded 127 projects at 110 institutions as a part of its 1988-1994 calculus initiative (Ganter, 1999). Awards ranged from $1,500 to $570,283 per year, with a mean annual award for $186,458. Some common themes emerged from the projects (Ganter, 1999):

- student and faculty responses were almost uniformly negative at first, improving only after the project had been revised based on continuous feedback,
- students in reform courses generally had better conceptual understanding, higher retention, higher confidence, and higher continued mathematics enrollment, yet
- scores on common traditional exams were mixed, making it unclear whether or not reform students were falling behind on traditional skills.

Accordingly, the mathematics and mathematics-education communities remain divided on the virtues of reform calculus (Wilson, 1997). Keeping these lessons in mind, I turn to another set of prominent efforts in calculus education.

1.1.3 The Emerging Scholars Program

The Emerging Scholars Program was developed in response to a decade of data showing that 60% of African-American students at UC Berkeley failed introductory calculus. The program was based on a study of the working habits 20 African American students and 20 Chinese American students (Treisman, 1992; Fullilove & Treisman, 1990). The authors found that the groups approached mathematics learning in very different ways. The Black students worked diligently for about 6-8 hours per week, turned in all assignments on time, but primarily worked alone. In contrast, the Chinese students worked about 8-10 hours alone, but also spent about 4-6 hours working with other students. They checked one another's answers, worked on problems from old exams together, and were extremely aware of their standing in the course.

Thus began the Emerging Scholars Program (ESP), an honors program for minority students (Fullilove & Treisman, 1990). There are two key components to the program:

- convincing students that college success requires collaboration and the creation of an intellectual community with other students with shared interests and common professional aims, and
- providing students with a set of rich mathematical tasks to solve in a collaborative group environment.

Essentially, the ESP sought to reproduce the learning conditions under which Chinese students were found to be so successful. The ESP enrolls first-year students of all races, but typically 80% of participants are Black and Latino (Fullilove & Treisman, 1990). When students enroll in the program, they take part in special two-hour problem sessions, twice a week, in addition to their traditional calculus section. These sessions involve students organized in groups of 5-7 working on exceptionally difficult sets of problems.

From the years 1978-1984, 40% of non-ESP African American students failed calculus (115/284), while less than 4% of ESP African American students (8/231) failed the course (Fullilove & Treisman, 1990). Studies of implementations of the ESP at other institutions (e.g., UT Austin, CCNY) show similar improvements for minority students (Treisman, 1992). Both the collaborative group setting and meaningful mathematical tasks are considered to be crucial to the program's success (Treisman, 1992). However, the ESP requires four additional hours of weekly problem solving, which is nontrivial to import this structure into a regular calculus section. Nevertheless, the principles of collaborative problem solving and meaningful tasks underlie the intervention I describe in this dissertation.

1.2 Worthy Mathematical Problems

Calculus is difficult for students to learn and efforts to reform calculus are perhaps even more difficult. Notable calculus reform efforts moved away from procedural exercises towards more meaningful tasks and utilized collaborative work. As a result, reformed courses were able to improve student engagement and retention, but performance on traditional measures was mixed. In contrast, the ESP resulted in drastic student performance gains, but has its own barriers to widespread adoption.

As stated previously, most of the studies described above lacked any measures of the quality of the mathematical tasks used in the interventions. To avoid this difficulty, I have justified my choice of tasks drawing from Complex Instruction (Featherstone et al., 2011) and Alan Schoenfeld's (1991) work on problem solving.

Complex Instruction is a theory and set of techniques for implementing equity-oriented group work, most commonly used in K-12 mathematics learning (E. G. Cohen & Lotan, 1997). One of the principles of Complex Instruction is that group work will only be effective if you give a group something worthy of group engagement, a group worthy problem. Given the equity-oriented focus of Complex Instruction, multiple entry points and the need for multiple competencies are central, so that all students have an opportunity to meaningfully engage the given task. Based on this work, Featherstone et al. (2011, p 119) recommend three guiding questions for choosing and altering tasks to create group worthy problems:

1. How can I make sure this task will engage my students with big mathematical ideas?
2. How can I improve the chances that the context that the task proposed supports rather than distracts from mathematical investigation?
3. Does the task allow for multiple points of access and more than one solution path (promoting the visibility of mathematical connections)?

In a similar vein, Schoenfeld (1991) explains his *problem aesthetic* for potentially valuable problems. In general, good problems:

1. are (relatively) accessible;
2. have multiple solution paths, making mathematical connections visible;
3. serve as introductions to important mathematical ideas;
4. serve as "seeds" to mathematical exploration (i.e., are extensible and generalizable).

Both sets of criteria emphasize the importance of opportunities to make connections and generalize the mathematics to new situations. These principles provide a basis for choosing tasks, but as Triesman notes, it is often difficult to find worthwhile calculus tasks (Treisman, 1992). In this dissertation, I often had to choose tasks in which students generated examples or explained some sort of substantial mathematics (Reinholz, 2013b), because it was difficult to find problems with multiple solution paths. In the context of peer-assisted reflection, problems that allowed for a great deal of explanation or the generation of examples made it likely that students would have different solutions, and thus made mathematical connections more likely.

In summary, this chapter has highlighted some of difficulties of improving student learning in calculus. I have also provided some of the rationale for the tasks chosen in my study and the use of a design-based research methodology. Moving forward, I now turn my attention to the skills students need to think deeply about mathematics and how to help students develop them. These will be emphasized in the next chapter and in the design sections of all subsequent chapters.

2 Theoretical Framing

The previous chapter described some of the difficulties of calculus reform and the importance of meaningful tasks. I now focus on explanation and reflection, fundamental skills for deeply engaging such tasks. Explanation relates to students' abilities to express their mathematical understandings verbally and in writing. Crucially, explanation is a tool for developing understanding, not just communicating it. Reflection relates to the monitoring and self-regulation of learning and understanding. When students analyze others' work, they develop the skills to explain and reflect upon their own work. This is the crux of peer-assisted reflection, the intervention I later describe.

2.1 Explanation

Explanation is fundamental to mathematical problem solving and understanding. The Common Core State Standards for Mathematics consider explanation to be a "hallmark" of understanding (CCSSI, 2010). Of the five NCTM process standards, explanation is fundamental to three of them (reasoning and proof, communication, and connections), and important to the other two (problem solving and representation; NCTM, 2000). Explanations promote generalization through the recognition and extension of patterns (Lombrozo, 2006; M. Ranney & Schank, 1998), and mathematics is a "science of patterns" (L. Steen, 1988), making the centrality of explanation to mathematical practice hard to dispute. Major science education standards documents (AAAS, 1993; NRC, 1996) also emphasize explanation in scientific practice.

2.1.1 Explanation and Learning

Mathematical arguments (i.e., proofs) embody methods, tools, and strategies for solving other problems (Hanna & Barbeau, 2008). These explanatory aspects of proof (not just the proving aspects) provide the basis for generalizing between problems (Raman, 2003). In science education, explanation often relates to describing (usually causal) mechanisms behind natural phenomena (e.g., why a ball eventually falls when you throw it up). Related to explanation is argumentation, which refers to reasoning from evidence (e.g., why one species would be more likely to adapt to a new situation than another; Osborne & Patterson, 2011).[1] Other researchers have focused on the coherence of explanations, using computational methods to model individuals' reasoning (M. Ranney & Schank, 1998). In this dissertation, I group all of these processes together under the heading of explanation, because all of them support making connections and generalizations.

Chi, Bassok, Lewis, Reimann, and Glaser (1989) observed students learning physics from worked examples. The most successful students spent the most time explaining the reasoning behind the examples. The authors hypothesized that explanation supported generalizing principles from the worked examples to new problems. To follow up, Chi, De Leeuw, Chiu, and LaVancher (1994) experimentally manipulated whether or not students self-explained while reading a biology text. The students who self-explained learned more than the comparison students, who read the text twice rather than a single time, even though many of the self-explanations were incorrect. The authors hypothesized the key features of self-explanation were that: (1) it is constructive, (2) it promotes the integration of new and prior knowledge, and (3) it is carried out continuously, providing multiple opportunities to see conflict between one's evolving understanding and the text. Self-explanation has been used effectively in other learning contexts, such as in promoting understanding of global climate change (M. A. Ranney, Clark, Reinholz, & Cohen, 2012a, 2012b). Self-explanation pushes readers to make explicit unspecified connections (i.e., explanation supports reflection).

2.2 Reflection

Reflection relates to monitoring, regulating, and assessing one's own work, progress, or learning. Reflection features in formative assessment (particularly peer- and self-assessment), feedback, self-regulation, metacognition, and metacomprehension.

2.2.1 Formative Assessment

Constructivist theories emphasize that learning must be grounded in students' current understandings (Smith, diSessa, & Roschelle, 1993). Accordingly, formative assessment is concerned with *evoking* information about learning and using it to *modify* the teaching and learning activities in which students are engaged (Black, Harrison, & Lee, 2003); i.e. any instructional practice that adheres to basic constructivist principles must include some aspect of formative assessment. The positive impact of formative assessment on student learning is well-documented (e.g., Black & Wiliam, 1998; Black et al., 2003).

Peer- and self-assessment are core parts of formative assessment (Black & Wiliam, 2009), as students themselves have been argued as the definitive source of formative assessment (Andrade, 2010). Emphasizing peer- and self-assessment accounts for the role that students, not just the teacher, must play in guiding the learning process. Peer-assessment refers to a

[1] But these meanings for the terms are not universal.(For an example of a explanation in this sense being used to refer to argument, see Songer & Gotwals, 2012).

spectrum of activities in which students engage with the work of other students (cf. Topping, 1998). As a result, the mere assignment of grades is often conflated with activities that involve in-depth analysis and feedback. Because grades alone contain so little information about how to improve one's performance, peer-marking (the assignment of grades) is much less effective than in-depth analysis and feedback (cf. Hattie & Timperley, 2007). For conceptual clarity, I will use the term *peer-assisted reflection*, rather than peer-assessment, in this dissertation.

Reflection can be peer-assisted in two ways: (1) when students analyze a peer's work and provide feedback, they develop skills that they can later apply to their own work; and (2) when students receive feedback from a peer, they gain insight into their peer's perceptions of the solution, which helps them reflect on their work from the distanced perspective of another. During phases I and II of the experiment I define peer-assisted reflection (PAR) as a specific activity structure, but in theory, PAR could be implemented in other ways as well. PAR involves students analyzing one another's work and conferring about their analyses. As they confer, students explain both their own work and the work of their peers, which promotes deeper understanding.

Ultimately, peer-assisted reflection should help students develop self-reflection skills. Because self-reflection is difficult, PAR helps students develop the distanced objectivity required for self-reflection (Black et al., 2003). In order to effectively self-reflect, a learner must: (a) possess a concept of the standard or goal to be achieved, (b) be able to compare actual performance to this standard, and (c) engage in appropriate action to close the gap between (a) and (b) (Sadler, 1989). This requires practice assessing a large number of examples of various quality (ibid); PAR provides students with precisely this opportunity.

2.2.2 Self-regulation and Metacognition

Standard models of self-regulation involve planning, monitoring, regulation, and reflection (cf. D. Butler & Winne, 1995; Pintrich, 2004; Zimmerman, 2002). Self-reflection supports monitoring (judging the quality of products during their production) and reflection (judging the quality of products after their production). Individuals must make accurate judgments about their work in order to meaningfully guide their future actions, or they will act based on misunderstandings.

External feedback often plays an important role in self-regulation. In general, feedback that focuses on task processing (e.g., error detection strategies, use of cues) and self-regulation (e.g., generating internal feedback, assessing the correctness of one's response) is the most productive (Hattie & Timperley, 2007), especially when it is used to improve one's initial performance (Sadler, 1989). Feedback that focuses on the self (e.g., praise) is usually detrimental, and should be avoided (Hattie & Timperley, 2007; Shute, 2008). Ultimately, as individuals learn to accurately self-reflect, they move away from reliance on external feedback to self-regulation (Sadler, 1989).

Metacomprehension is the ability to judge one's own understanding. Although this ability is often assessed using limited measures of understanding (e.g., text comprehension and multiple choice questions; Dunlosky & Lipko, 2007; Dunlosky & Thiede, 2013), individuals' accuracy is still quite low. Evidently, many individuals self-assess by using a generalized notion of cognitive ease (Kahneman, 2011), rather than actually attempting the task and measuring their performance. Pushing individuals to engage more substantively with a task before making judgments has been shown to improve their accuracy (Dunlosky & Lipko, 2007). More broadly, metacognition refers to one's knowledge, beliefs, and awareness of one's thinking and learning processes (Brown, 1987; Young, 2010), which is an important part of self-regulation and learning more generally (Bransford, Brown, & Cocking, 2000). Just as peer-assessment supports self-assessment (Black et al., 2003), learners often develop metacognitive skills by internalizing external, social practices (cf. Palinscar & Brown, 1984; Schoenfeld, 1985; White, Frederiksen, & Collins, 2009).

2.3 Summary

Explanation helps students connect new and prior knowledge, reflect on gaps in understanding, and generalize to new situations. Reflection is a core part of self-regulation, the ability to control and direct one's own learning. Peer-assisted reflection leverages the learning potential of explanation to help students develop the metacognitive skills necessary to transition from external feedback to self-regulation. The next chapter concerns my preliminary efforts to enact PAR in the classroom.

2.4 Research Questions and Overview

I focused on two primary research questions, related to the efficacy of the PAR intervention:

1. How does student performance change on standard measures of understanding (compared to other sections)?; and,
2. How do student explanations develop over the course of the semester?

I attempted to address these questions through three semesters of study (see Figure 2.1). The next three chapters (Pilot Study, Phase I, and Phase II) focus primarily on the evolution of the experimental intervention and its implementation in the classroom. To support these discussions, I provide analyses of student outcomes during Phases I and II, which demonstrate significant learning gains in the experimental sections; student success rates were much higher in the experimental sections than the comparison sections, as were exam scores, on both conceptual and procedural problems. Readers who are interested in the final version of the design but not its evolution should skip Chapters 3 and 4, and begin with Chapter 5 (Phase II). Chapter 6 provides qualitative analyses of student explanations in the Phase I and Phase II experimental sections, and a comparison section sampled during Phase II. These analyses highlight the improvements in explanations in the experimental sections. Chapter 7 provides a discussion of the mechanisms that appeared to make the intervention a powerful learning activity. The dissertation ends with a summary of the three semesters of study.

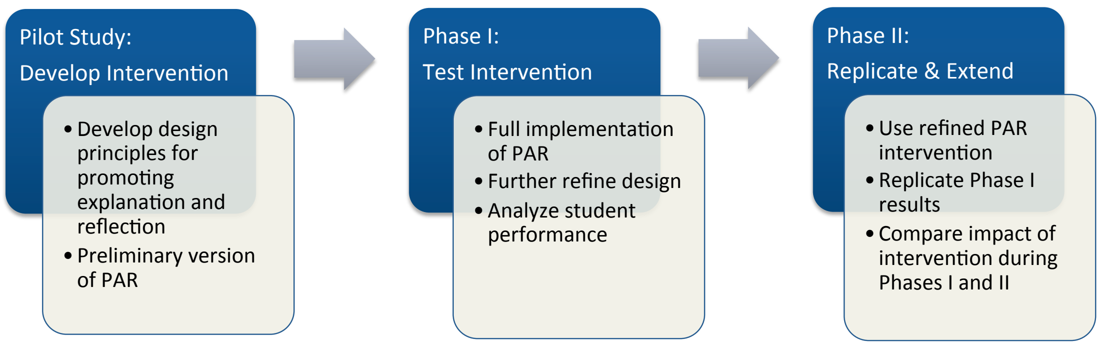


Figure 2.1: Overview of the 3-phase design-based research study.

3 Pilot Study

3.1 Chapter Overview

The purpose of the pilot study was to translate theoretical principles of explanation and reflection into an intervention to promote meaningful student learning. During the pilot study I acted as a teacher and researcher in a community college remedial algebra classroom. Because the students had already been unsuccessful in traditional algebra classes, this study would help me design activities to support deep learning for traditionally underserved students. I begin with an overview of the design as conceived, provide analyses of the design as implemented, and close with a discussion of student learning and design principles developed. The following two chapters (phase I and phase II) utilize the same structure.

The pilot study was exploratory, and I did not collect data to assess the impact of the experimental activities on student outcomes. Nevertheless, I uncovered important theoretical insights that served as the basis for future iterations of the intervention. Supporting explanation and reflection appeared to require: (1) helping students overcome the perceived need for expertise to construct and critique explanations (the "expert barrier"), (2) utilizing a small number of specific, focused prompts to facilitate reflection, and (3) leveraging the role of the audience of student explanations. Using these insights I constructed a standard form and activity structure for self-reflection and peer-analysis. I refrain from using the term peer-assisted reflection in this chapter to distinguish the pilot study activities from phases I and II of the study.

3.2 Design

The design had four main components: (1) a framework for analysis of explanations/solutions, (2) opening problems and self-assessments, (3) analysis tasks, and (4) opportunities for student explanations. Activities were chosen to provide as many opportunities as possible for students to explain and reflect upon mathematics. Choices relied heavily on the work of other researchers (e.g., Black et al., 2003; Sadler, 1989) and advice from experienced instructors in the developmental (remedial) mathematics program at Los Medanos College. I summarize the characteristics of the pilot study design in table 3.1.

Strand of Design	Key Features
Explanation Framework	• Focus on execution, explanation, justification • Rubric co-constructed by teacher and students • Used for reflecting on solutions
Opening Problems & Self-Assessments	• Written on board at beginning of class • Opening problem pushes students to think about lesson • Guiding question to help focus students • Five self-assessment questions
Analysis Tasks	• 3 types: explanation framework, unguided, standard form • Used primarily student work until end of semester • Variety of types used to see which would be most effective
Student Explanations	• Written explanations, class discussions • Student presentations to class (later in semester)

Table 3.1: Pilot Study Design Summary

3.2.1 Explanation Framework

To support students to engage in novel peer-analysis and self-reflection activities, instructors at Los Medanos College suggested that I should co-construct a rubric for analyzing the quality of mathematics solutions with my students. This rubric would help students meaningfully engage in analysis tasks throughout the course. The actual rubric that was co-constructed is presented in the analysis section.

3.2.2 Opening Problems and Self-Assessments

The basic lesson structure was inspired by the Learning Mathematics through Representations (LMR) project (Saxe, de Kirby, Le, Sitabkhan, & Kang, in press). Each LMR classroom lesson consisted of five phases; the first phase (opening problem) was designed to promote initial reflection by students, and the final phase (closing problem) was designed to provide insight into students' understandings at the end of the lesson, to help inform future lessons. I used these phases as a part of the daily lesson structure to promote reflection.

Students were presented with a guiding question and an opening problem to begin each lesson (see appendix D for sample guiding questions). The guiding question was intended to help students stay focused throughout the lesson on a core mathematical idea. The opening problem either reviewed old ideas or pushed students to think about new mathematics in order to set up the lesson. As soon as students entered the classroom they were expected to take out a sheet of paper, write down the guiding question, and work on the opening problem. At the end of each lesson, students completed five prompts:

1. Answer the day's guiding question.
2. "Traffic light" how well you understood the lesson (traffic lights are described below).
3. Explain something you learned today as if you were explaining it to a classmate who missed the lesson.
4. Ask one question about mathematics.
5. Tell me something else you think I should know.

These questions were meant to focus students on their learning processes and make progress explicit, which is a key component of increasing self-efficacy (Bandura, 1997). Ongoing self-assessments were intended to improve the accuracy of students judgments (Dunlosky & Lipko, 2007). By explaining mathematics each day, students were meant to deepen their understanding. Finally, these self-assessments helped the instructor modify instruction to meet learners' needs (i.e., to practice formative assessment).

Students frequently used "traffic lights" to self-assess their understanding of classroom activities, homework problems, and entire lessons (Black et al., 2003). Green (G) represented a strong understanding, yellow (Y) represented a partial understanding, and red (R) represented a lack of understanding. These self-assessments informed future classroom practices (e.g., students often chose a "red" homework problem to discuss with a partner). The instructor was explicit about the purpose of traffic lights, their role in informing instruction, and how to use them to guide learning.

3.2.3 Analysis Tasks

Students need explicit guidance, access to a variety of examples, and a lot of practice to learn to analyze mathematical work (cf. Sadler, 1989). Malcolm Swan, a lead designer for the Shell Centre's formative assessment lessons, suggested that I utilize primarily hypothetical student work, so I could engineer learning opportunities around specific conceptual difficulties (Malcolm Swan, personal communication, November 14, 2011). Accordingly, most of the analysis tasks used during the beginning of the semester utilized hypothetical student work. Analyzing hypothetical student work would also provide a safer classroom environment.

Students engaged in three types of tasks: (1) general analysis tasks, (2) tasks utilizing the explanation framework, and (3) tasks using the standard form that was later developed. In (1), students had to explain other students' thinking or rate the quality of various solutions. In (2), students applied the explanation framework to a sample of work (hypothetical or their own). In (3), near the end of the semester, students analyzed real and hypothetical work (see appendix B). The peer-feedback and self-reflection components of the standard form were identical, which was intended to facilitate the transfer of peer-analysis to self-reflection.

The tasks used in the class were meant to be accessible but nontrivial, afford opportunities for explanation and multiple solution methods, and engage the students with deep mathematics, to promote making mathematical connections. The high-quality mathematical tasks developed by the Mathematics Assessment Resource Service (MARS) were a source of inspiration (see http://map.mathshell.org/materials/lessons.php). In fact, a number of the formative assessment lessons, such as gold rush and counting trees, were used during the semester.

3.2.4 Student Explanations

During the beginning of the semester, students primarily explained their reasoning in response to written prompts. This took place through homework and in-class activities. As the intervention evolved, students were afforded increased opportunities to explain their reasoning verbally. This occurred through group work and student presentations.

3.3 Method

Observations took place in two classrooms at a community college in the San Francisco Bay Area. Students in the experimental section ($N = 14$ after dropouts), were involved in a first year experience program, which grouped recent high-school graduates into a cohort that took the same three classes: elementary algebra (remedial algebra I), remedial English, and a study skills course. The comparison section ($N = 17$ after dropouts) was supposed to be another cohort of the first year experience program, but due to low enrollment the section was opened up to all students and was not tied to the English or study skills courses.

Both algebra sections met in parallel (i.e., all classes took place simultaneously in different classrooms). The class met five days a week at 8:00am for 50 minutes each session. The urban classroom consisted primarily of students of color (roughly 80%). As a teacher and researcher, I taught the experimental section and collected copies of all of the students' in-class work and homework assignments, and a stationary video camera in the back of the classroom captured lectures and in-class discussions. Additionally, I kept field notes of daily observations and reflections. In the comparison section, I video recorded five class sessions and collected copies of students' final exams.

3.4 Analysis

To begin, I provide a brief description of the comparison section. This is to provide a baseline comparison for standard instruction in the community college. This stands in contrast to the experimental activities that I later describe.

3.4.1 Comparison Section

The comparison section was taught by a cooperating instructor, Sarah. Sarah's class was typical in most respects. She began each day by writing assignments and other important reminders on the left side of the blackboard, and then began by titling the blackboard with the name of the day's lecture. In general, lectures flowed smoothly, and the instructor's organization of the board was excellent. All material that was written was very clear and organized, and if the students were simply to copy the board work down verbatim, they would have a complete set of notes. Class sessions proceeded with the instructor explaining the main ideas and working out a large number of examples. The class sessions were instructor-driven, giving students relatively few opportunities to ask questions or give input in problem solving. Of the 5 class sessions that were observed, the class remained in whole class lecture the entire time.

3.4.2 Implementation of Intervention

Implementation was hindered in a number of ways: modest attendance (mean 71.6%), students coming to class late (often 20 or more minutes late), and a large number of homework assignments not turned in. Basic literacy was an issue; most students struggled to write complete sentences. This prompted the use of more verbal, rather than written, activities. Other members of the mathematics department informed me that these issues were typical for this particular mathematics course and early meeting time (8:00am).

3.4.2.1 Explanation Framework

During the second class session, the class co-constructed a set of criteria for high-quality mathematics solutions. After the students came up with the criteria for high-quality solutions, the instructor grouped them into three categories: execution, explanation, and justification. A few of the items given in table 3.2 that were not suggested by the students were added by the instructor (e.g., check units), but the majority of the suggestions came from the students.

Execution **(what did you do?)**	**Explanation** **(why did you do it?)**	**Justification)** **(did you do it correctly?)**
Show all solution steps in order	State any assumptions made	Check units
Define all variables in the problem	Explain why you chose to solve the problem a certain way	Estimate or solve a simpler problem
Write down important information from the problem statement	Draw a picture or diagram	Interpret answer using the problem context
Include units on all quantities	Use words to explain meaning of arithmetic operations in problem context	Solve the problem using another method
Answer all questions asked in the problem	Explain choice of representations	

Table 3.2: Explanation Framework for High Quality Solutions

The various components of the rubric - execution, explanation, and justification - were designed to provide a sense of what constitutes an explanation (more than recounting steps). Use of the framework was often modeled during class and students were asked to apply it to sample work. Nevertheless, students did not find the framework very useful. Consider two questions from homework 6:

- Choose one of the problems you solved on either a previous homework, exam, or in-class assignment. Write your original solution. Then analyze the solution in terms of execution, explanation, and justification. Write down at least two ways in which you could improve the original solution.
- Based on your answer to the previous question, rework the problem that you just analyzed your solution for (don't just copy from the solutions!).

Two samples of typical student responses are given, one from Teresa (see figure 3.1), the other from Josh (see figure 3.2).

Teresa chose a problem from a recent in-class assignment of matching equations and graphs with each other. However, she only focused on the correctness of the solution, without applying the framework for analysis.

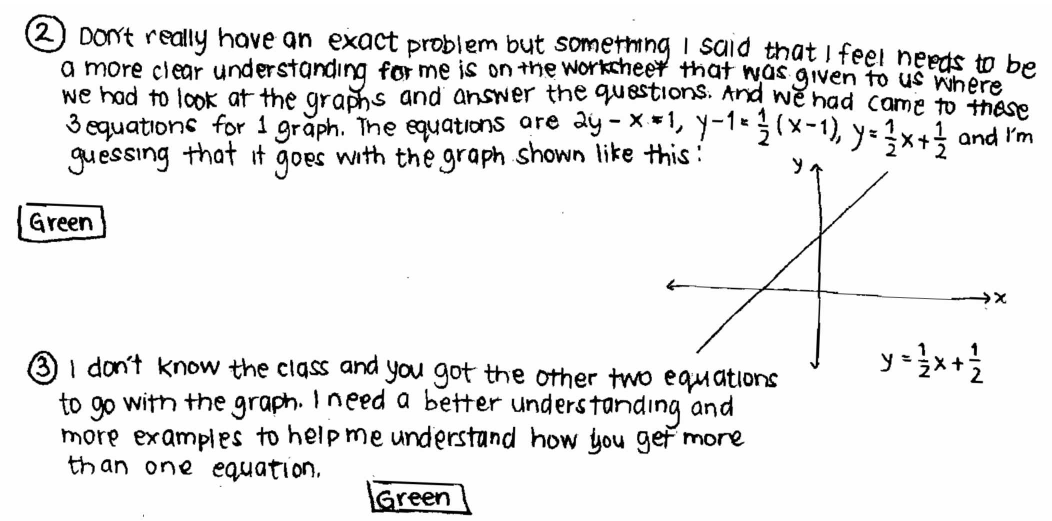


Figure 3.1: Teresa's Solution for the Homework 6 Revision Problem

Josh also picked out a problem from a previous class session. Students had to create a graphical representation of a standard story situation: riding on the bus, walking, sitting at class, walking back to the bus, and riding home. Josh reworked his initial solution, fixing one of the errors (the distance traveled did not change while the student was in class), even though other errors remained (and were introduced). Just as for Teresa, the framework did not support Josh to reflect on his work. Due to its inefficacy, the explanation framework was phased out near the middle of the semester.

3.4.2.2 Opening Problems and Reflections

When students arrived on time, they generally began working on the opening problem quickly. However, students who came to class late were slow to work on the problem, spending a long time and disrupting the flow of the class session, and often did not work

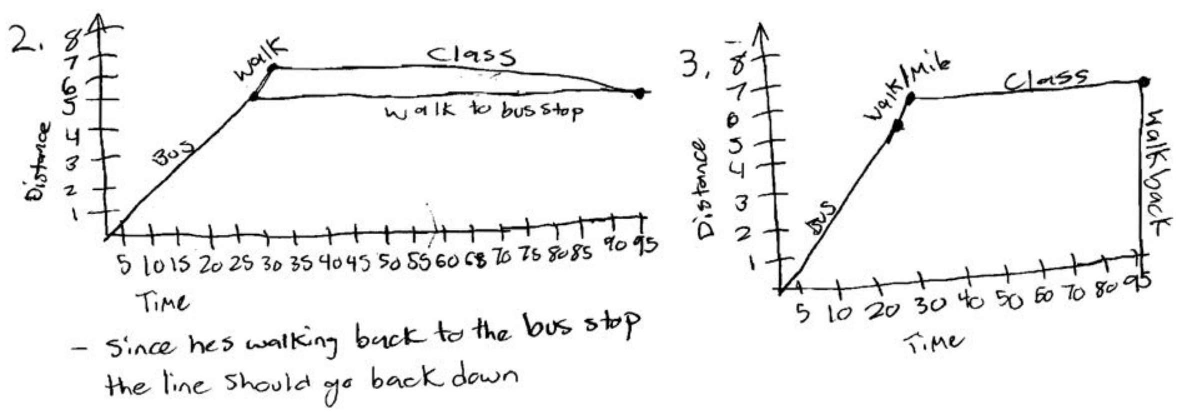


Figure 3.2: Josh's Solution for the Homework 6 Revision Problem

on the problems at all (because they came halfway through the session). Students were not graded for on-time attendance. Opening problems were usually simple. For instance, on the second day of class students were asked to multiply various combinations of positive and negative numbers together. The prompts were: $5 \cdot 5$, $(-1) \cdot 6$, $(-2) \cdot (-2)$, and $6 \cdot (-1)$.

Guiding questions either summarized a big idea from the lesson or were used to interrogate student thinking about a specific topic. For instance, the guiding question corresponding to constructing the explanation framework was *what does a "good" math solution look like?* Student responses are given:

- I have no clue.
- A good math solution has multiple elements to show you did it right. One, what did you do, two why did you do it, 3 did you do it correctly: execution, justification, and explanation.
- 25, -6, +4, -6
- $+ \cdot + = +$, $- \cdot + = -$, $- \cdot - = +$, $+ \cdot - = -$,
- A good one likes one were you can see how you got from point A to point B.
- I would like to better understand the in-class activities because I recognize the equation but can't explain.
- A good math solution either has two positives or two negatives.
- A good solution would be a complete answer that is correct with a box around it.
- A good math solution looks like when you're able to solve everything in order and it's easy to understand the process. The shorter the work the better the solution would look.

These responses provided a number of insights: (1) only 9 of 20 students enrolled (at that time) attended the second day of class, (2) most students misunderstood the self-assessment task, and (3) student responses were mostly unrelated to the class discussion. Future class sessions were modified to address these issues. By the third week of the class, nearly all students understood how to answer the guiding question. Nevertheless, guiding questions had a number of limitations: (1) students seemed to only think about them at the end of the lesson, not as a tool to stay focused during class, (2) students often rushed to finish their reflections, and (3) it was often difficult to develop a meaningful guiding question before each lesson (e.g., sometimes lesson goals were emergent).

In contrast, students found using traffic lights (cf. Black et al., 2003) to track their progress useful, but also felt limited by only having three choices. While students could easily identify "red lights," students frequently marked work that was mostly incorrect as green, greatly overestimating their abilities (cf. Dunlosky & Lipko, 2007). Student reflections about traffic lights illuminated the source of some of their difficulties. For example, Vanessa stated:

> "Red flag" doesn't help me - it only makes me frustrated I don't have an understanding of solving or graphing. That is my frustration for the entire homework.

Vanessa notes that a "red flag" doesn't help, because it provides insufficient grist for working through partial understandings. A general, non-specific method of self-assessment (such as the traffic lights) rests on the tacit assumption that students have the skills to self-assess and reflect but not the predilection to do so. However, most students made vague, general comments such as "I'm confused" or "I don't get it," which are not very effective for self-assessment because they do not specify "what" exactly needs to be resolved.

This problem was ameliorated to some extent through the design of the standard form for peer-analysis and self-reflection (see appendix B.1). Withholding further discussion for the next section, I provide a direct comparison between traffic lights and the standard form made by Teresa after class one day:

> This new assessment form is a lot better than the traffic lights. It actually gives you questions to answer, which makes it a lot easier to use.

The third self-assessment prompt asked students to explain something they had learned today as though they were explaining it to a classmate who missed the lesson. In practice, students almost never provided an explanation for a classmate. Instead, most students interpreted the question as *what did you learn today?* Here is a representative sample of student responses (from week 5):

- *(Blank)*
- I remembered x and y axis. We learned about input and output.
- *(Blank)*
- $d(t) = st$, $d =$ distance $=$ traveled like likes, time means hours
- I learn x, y axes.
- We used an input/output box meaning whatever goes into a function gives us an output.
- To find what the output is you have to know what the input is first and put that into the function to get the output.

- Today we learned about graphing.
- Input goes into a function and then it gives us a output
- Today I relearned about graphs and input and output.

The vast majority of these explanations would not support a classmate's learning; they are simply statements of what students felt they learned.

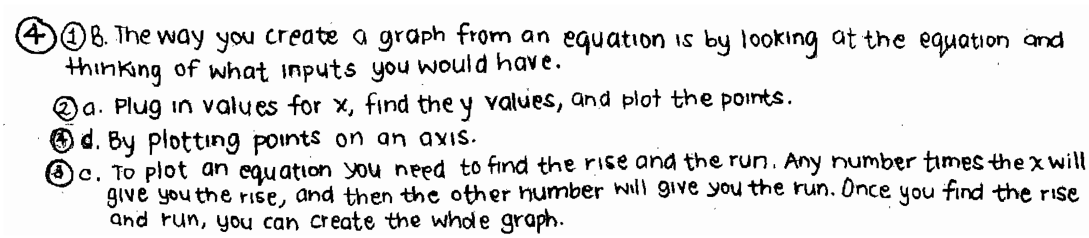
The final two self-assessment prompts required students to ask a question and say something else they wanted the instructor to know. Both of these prompts provided useful information for the instructor. Students asked questions that they may have felt uncomfortable asking in class, and talked about how personal struggles could be affecting their attendance and performance.

3.4.2.3 Analysis Tasks

The explanation framework was discussed above, so I now consider general analysis tasks and the standard form that was later developed. General analysis tasks focused on analyzing sample student work (primarily at the beginning of the semester). Students developed their analytic skills slowly, and in the absence of a guiding framework, each student applied idiosyncratic methods. Consider another problem from homework 6:

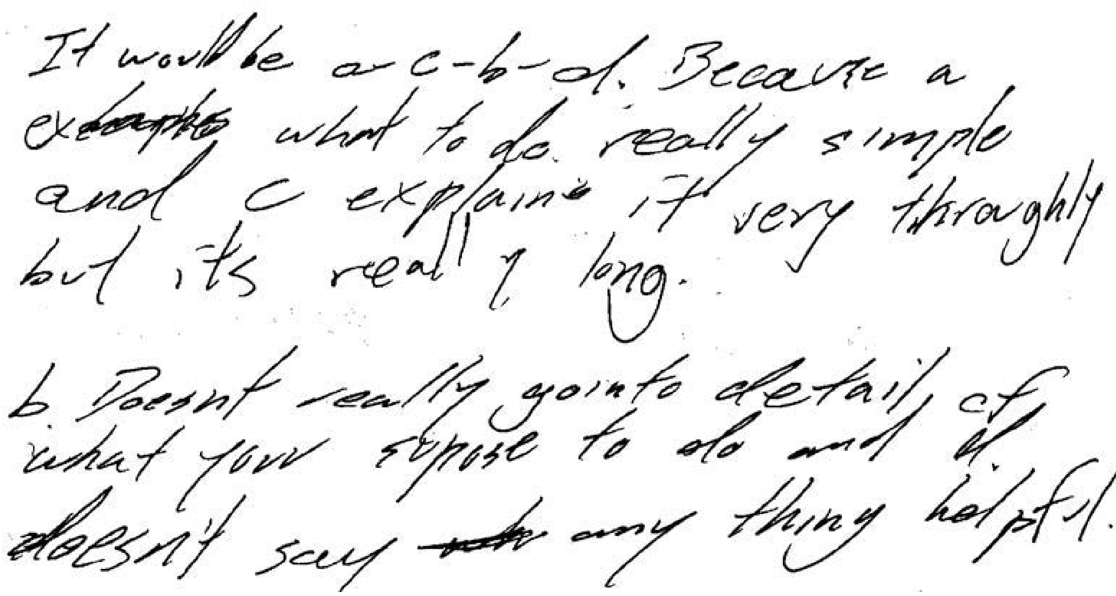
> Four answers to the question "how do you create a graph from an equation?" are given below.
>
> (a) Plug in values for x, find the y values, and plot the points.
> (b) The way you create a graph from an equation is by looking at the equation and thinking of what inputs you would have.
> (c) To plot an equation you need to find the rise and the run. Any number times the x will give you the rise, and then the other number will give you the run. Once you find the rise and run, you can create the whole graph.
> (d) By plotting points on an axis.
>
> Order these explanations from best (1) to worst (4). Explain why you ordered the explanations how you did.

The four choices were modified samples of student work. Example (c) was designed as a decoy to see if students would consider it to be the best because of its length. Indeed, not a single student ranked (c) as the worst explanation, and some ranked it as the best. I now consider three samples of student work, Teresa (see figure 3.3), Joseph (see figure 3.4), and Vanessa (see figure 3.5). These samples represent higher quality responses from students; many students skipped these problem, wrote down that they didn't know what to do, or attempted to solve the problem while ignoring the sample work.



④ ① B. The way you create a graph from an equation is by looking at the equation and thinking of what inputs you would have.
② a. Plug in values for x, find the y values, and plot the points.
④ d. By plotting points on an axis.
③ c. To plot an equation you need to find the rise and the run. Any number times the x will give you the rise, and then the other number will give you the run. Once you find the rise and run, you can create the whole graph.

Figure 3.3: Teresa's Solution for the Homework 6 Analysis Problem



It would be a-c-b-d. Because a explains what to do really simple and c explains it very throughly but it's really long.

b Doesnt really go into detail of what your suppose to do and it doesn't say any thing helpful.

Figure 3.4: Joseph's Solution for the Homework 6 Analysis Problem

Teresa ranked the four explanations in the order of (b)-(a)-(c)-(d). Although she rated (c) near the bottom, she ranked (b) near the top, which is also a very flawed explanation. Teresa provided no justifications for her answers.

Joseph chose the order (a)-(c)-(b)-(d). He ranked (c) as second because it was "very thorough," even though it appears he viewed the length as a drawback. Joseph evidently did not recognize the numerous flaws in the explanation.

Finally, Vanessa ranked the solutions as (c)-(a)-(d)-(b). Vanessa ranked the decoy (c) as the best solution, but her reasoning was difficult to follow.

In summary, Joseph gave partial reasoning, Vanessa offered less complete reasoning, and Teresa offered no reasoning. Of all of the students, not a single one lucidly articulated the specific limitations of the explanations (e.g., (b) gives you no idea how to get outputs, (c) gives incorrect ways for finding rise and run and doesn't specify how to create the graph once they are found, and (d) gives you no idea how to get the points to plot). Even after six weeks, the students had relatively little ability to analyze mathematical explanations. Yet, they knew it was an expectation of the course. Jacki and Tia frequently referred to the

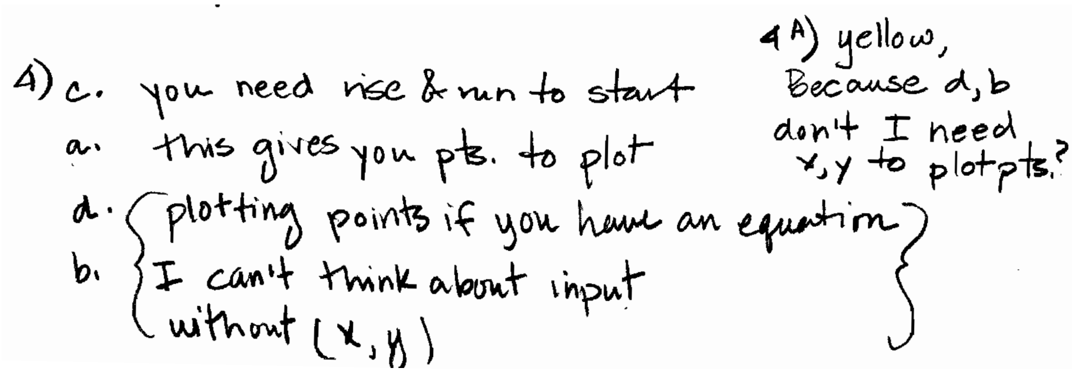


Figure 3.5: Vanessa's Solution for the Homework 6 Analysis Problem

class as "the explain class," and during an in-class discussion when a number of students asked me to "just tell them" how to solve the problem, Sigfried exclaimed "you know he's not going to tell us the answer!"

The student work and reflections shown previously provided insight into the first major theoretical breakthrough, the *expert barrier*. Most students in the class believed that only "experts" are capable of explaining mathematics, which was a barrier to their engagement with explanation activities. Yet, as a constructive activity, explanation is one of the very practices students must engage to develop expertise. Explanation provides students with an opportunity to surface, examine, and reflect upon their current understandings, and refine them toward more normative understandings (Chi et al., 1994). This is an "expert barrier" – the perceived need for expertise prevents students from developing expertise. This perception is related both to issues of self-efficacy (Bandura, 1997) and perceived locus of mathematical authority (Boaler & Greeno, 2000).

This theoretical insight prompted design modifications. Because students didn't feel they had the expertise to assess "right or wrong" or "good or bad" I reframed the tasks so that students wouldn't feel as though they had to make such judgments. This resulted in the development of the standard form for peer-analysis and self-reflection (see appendix B.1). I now provide the three prompts used in the form:

- **Big Picture:** What was the approach to solving the problem?
- **Accuracy:** Do you think the problem was solved correctly? How do you know? (Note any errors you found.)
- **Questions:** What are you unsure about in the problem or response? What do you want to know more about?

The design goals for the standard form were as follows: (1) provide students with a standard tool for both peer-analysis and self-reflection, (2) focus students on what their peers were doing, and (3) move students away from acting as judges (e.g., by leaving in room for uncertainty in the third prompt). This resulted in the third type of analysis tasks that students engaged in, primarily during the final 4 weeks of the semester. Students typically used the standard form to analyze one another's homework solutions, but they also used it to analyze sample solutions in class. Use of the form seemed to relate to shifts in engagement. As Teresa noted after analyzing a peer's work:

> I didn't get it before, why you were always asking us to explain, but now it makes sense. When you don't explain things people can't tell what you're doing.

When the framing for the activity was shifted to focus on *understanding* the explanations and providing feedback rather than *assessing* them, students were better able to engage the activity. Consider an excerpt from a lesson in which students engaged sample student work with the standard form[1]:

> Instructor: That's good, so what questions might you ask him?
> ...
> Dahlia: I don't get it, because he's not explaining anything.
> Julius: I want him to say more about two people, I didn't really get that part.
> Instructor: Okay, so you want him to clarify what he meant about two people.
> Julius: Yeah, I don't really understand that part. "When you have two people it's easier to do it by yourself;" what do you mean by that?

Even though the task was framed as understanding the given explanations, it ultimately led to the assessment of the explanations, as evident in Dahlia's comment that he "wasn't explaining anything." In his self-assessment that day, Julius noted:

> I learned that explaining a problem *wrong* even with the right answer can be confusing.

This shift in framing was effective because the perceived need for expertise was reduced. Further, I recognized that students needed an authentic (peer) audience for their explanations. Awareness of their audience made students more likely to see communication as important. I elaborate more below.

3.4.2.4 Student Explanations

During the first half of the semester, students primarily explained through written prompts or during whole-class discussion. As a result, when students engaged with the self-assessment prompts (as shown above) they simply stated what they learned, rather than actually trying to explain the ideas. It appears that the instructor was the students' primary (perhaps only) audience, and they tailored their responses accordingly. Students explained their homework problems in the same way.

[1]The work was a part of the Shell Centre's formative assessment lesson, "Golden Rectangles," which is freely available online at http://map.mathshell.org/materials/download.php?fileid=1226

I conjectured that a peer audience would highlight the communicative aspects of explanation, because students would receive visceral feedback when their explanations were not understood. Accordingly, student presentations of mathematics became a crucial component of the course. Students primarily presented homework problems and opening problems, but occasionally novel problems as well. Thus, students replaced their instructor as discussion leaders for these problems. Student presentations were framed as opportunities to work out partial understandings, not to teach classmates "the answer."

Just two weeks after students began regularly presenting their work, Teresa and Tia were vigorously discussing who would present the opening problem, in contrast to two weeks earlier when they both deliberately avoided being in front of the class. In reviewing homework 13, students decided amongst themselves who would work out problems, and didn't even ask the instructor to present. On another occasion Julius stated that he would prefer to work out the problem in front of the class and get feedback rather than having someone else just show the solution to him.

Analysis activities with the standard form appeared to supported student presentations. To illustrate, I contrast a student presentation from before the use of the form and after. On 2012-04-03, before the use of the form, Sigfried presented about a situation in which the students were given a right triangle with two sides equal to 3, 4, and the other side was an unknown.

> Sigfried: So class, today we learned $a^2 + b^2 = c^2$. As we see these two right here, we can plug in either or for a or b. [writes $4^2 + 3^2 = c^2$ on board]
>
> Sigfried: Okay so we plugged in 4 as a, and 3 as b.
>
> *(side conversation)*
>
> Sigfried: so 4^2 simplifies as 16. 3^2 okay... [writes $16 + 9 = c^2$]
>
> Sigfried: so 25. Then you gotta do the square root of 25, cuz it's c^2 and that would give you 5.

After the presentation a few students asked him to clarify his steps, but the conversation remained focused on answer-getting.

In contrast, during Sigfried's (2012-04-25) presentation (after the introduction of the standard form), students had a lively discussion (see figure 3.6 for the task). At first the presentation looked as before, but it took an unexpected turn.

Sigfried's presentation is given:

> Sigfried: You see right here it's like they are points on a single line. If you see on this side, it goes up by 2 (it goes 0,2,4,6). And on this side it goes up by 5 - it goes from -3, to 2, to 7, to 12 - it goes up by 5. [writes +5 to the right of table, +2 to the left of it]
>
> Sigfried: So if it was $y = 22$, you'd have to go up...
>
> *(Sigfried gets confused for a minute if x or y has to be 22, but it is resolved)*
>
> Sigfried: So this goes up by 5, so the next would be 17, and then 22.
>
> Sigfried: So if you go 2 more down on this one, it would be 8, then 10

Linear Relationship

The table below gives samples of points on a single line.

x	y
0	−3
2	2
4	7
6	12

Based on this table, what is the value of x when y = 22?

Figure 3.6: Homework 13, problem 1

At this point Teresa asked "is there an equation you could do?," launching a deeper mathematical discussion:

Sigfried: Honestly, I didn't get an equation. But if we had an equation then we could check it...do you know an equation we could do?

Dan (Instructor): So this is good. You solved the problem, but now this relates to the whole questions thing. Even though you solved the problem, there's still more math. We're pushing one level deeper here.

. . .

Dan: So how would we find a line for this?

Sigfried: We need an equation, or a graph.

Dan: So what do you want to do? Do you want to try to graph it, or look for an equation?

Sigfried: I think we should try to graph it

Tia: I wanna graph it.

This was the first time in the class that students pushed deeper with the mathematics rather than stopping as soon as the problem was "solved." Continuing on, the instructor worked with the students to construct a graph from the table. The instructor was at the board, but the students told him the process for graphing and how to proceed. For the next 4 minutes the students engaged with the difficulties of constructing a graph, and then came up with an equation based on the graph. Then Sigfried had his major insight.

Sigfried: Oh!! Oh I see!!

Sigfried: You can just find it from the thing. The rise over the run is the pattern, the y over the x, it's like the differences that they move.

Sigfried: All right, so now I see a better way to get it. Fo sho! [gives a high-five to another student]
Teresa: Good job!

The class engaged mathematics with an unprecedented interest. On the self-reflection Sigfried had turned in a day earlier, two questions were written: "What is the line?" and, "how do I know it's a line?" which may have supported his deeper inquiry. In sum, the standard form seemed to help students see the importance of communication and opened up space for deeper mathematical conversations.

3.4.3 Results

Students provided evidence that they recognized the expectation to explain and understand mathematics deeply, and began to recognize the importance of communication. Nevertheless, the standard form was only used during the final 4 weeks of class, so I could not document student growth over time. I was also unable to collect evidence supporting the specific reflection prompts used, and recognized that they would likely be revised later. Also, due to the nature of the pilot study, it was difficult to compare the written work of students between sections in any statistically or qualitatively meaningful way.

Nevertheless, on the final exam, I included the classic NAEP buses problem (Carpenter, Lindquist, Matthews, & Silver, 1983) (as a free-response item rather than multiple choice). Student responses were as follows: 8/10 correctly answered 32 buses or clearly explained there would be 31 full buses and part of the 32^{nd} bus, 1/10 students performed the correct calculation but answered with 'remainder buses,' and one student could not perform the correct calculation (contrast this 80% correct to 23% who answered correctly in the original study). This gives some indication that most students had understood the point that mathematics *should* make sense.

3.5 Discussion

The results of the pilot study highlighted the difficulties of translating broader theories of learning into a specific design. I learned as much from the repeated failures to promote meaningful student engagement at the beginning of the semester as I learned from the eventual successes. The results of the pilot study provided the basis for an intervention that would be fully implemented during Phase I, in order to assess its impact on student learning.

3.5.1 Revisions for Theory and Intervention

I now take stock of the key lessons learned during the pilot study.

1. **The importance of framing.** Effective analysis or assessment tasks need to address the expert barrier. Tasks should be framed to consider some aspect of communication or describing thinking and should leave room for uncertainty.
2. **The importance of a few, simple prompts.** Students will best be able to engage in an analysis task if they are given a simple framework to guide their learning.
3. **The importance of audience.** The perceived audience influences the ways in which students engaged in the analysis tasks. In order for students to engage in meaningful explanations, they should perceive an authentic audience that requires a more in-depth discussion than that which is standardly communicated to an instructor.

The first of these insights built on notions of self-efficacy (Bandura, 1997) and authority (Boaler & Greeno, 2000). When students were asked to assess the correctness of solutions, they felt as though they did not have the expertise to do so. Yet, when the framing was shifted to focus on providing useful feedback, students were better able to engage, because they did not perceive themselves as lacking adequate expertise.

The next major insight was the importance of having a limited number of effective prompts. When students used the explanation framework, they had too much information that they didn't know how to apply. On the other hand, without a guiding framework to analyze solutions, students did so in idiosyncratic and largely ineffective ways. Once the standard form was used, focusing on a limited number of specific prompts, it appeared to be much more effective for guiding student activity.

The final theoretical insight was the importance of audience for engaging in peer-analysis. Depending on who students perceive as their audience, they are likely to engage in explanations in very different ways. By actually having students explain to their peers, issues of audience became immediately apparent as real obstacles that needed to be overcome, rather than just as the pickiness of their instructor. Furthermore, by having students to explain to one another, students are afforded powerful opportunities to learn from one another's explanations.

In addition to the breakthroughs above, a number of smaller changes were made to the intervention as well.

1. Opening problems would henceforth be given to students on a piece of paper in order to facilitate a quick start in class. Giving students a physical piece of paper would also make it easier to change the self-assessment questions for the end of the class more easily to survey students about various topics.
2. The self-assessment prompts at the end of each class period would be reduced from 5 to 3. The guiding question would be dropped altogether, as would the explanation prompt (explain something you learned as though you were explaining it to a classmate). This would reduce the amount of time spent on the daily self-assessments.
3. Rather than using traffic lights, students would instead use a 0% to 100% scale at the end of each class period. This would correspond with the idea of a grading scale which is familiar, so it should be easy to adopt, and it would address the issue of lack of

choices with 3 colors. Students would no longer need to rate their understanding on the homework assignments, only each class period.

4. Students would begin engaging with the analysis tasks using the standard tool starting from the beginning of the semester. This would allow for framing and training related to these specific prompts. The explanation framework was dropped altogether.
5. The process of analyzing solutions and reflecting upon them needed to become streamlined into a regular, predictable activity. Also, this activity should involve a real audience rather than mostly hypothetical work. As a result, a regularized cycle of analysis, peer-assisted reflection (PAR), was introduced from the start of the semester.
6. Students would be given regular opportunities for in-class presentations and to present to one another in smaller groups or pairs. This discussion aspect of mathematical ideas would be made a regular part of PAR.

All of these changes will be considered in greater detail in the next chapter, as I discuss the next iteration of the design in detail.

4 Phase I

4.1 Chapter Overview

The purpose of Phase I was to fully implement the intervention developed during the pilot study. This would provide opportunities to assess the impact of the intervention on student learning, and also to further develop theoretical principles to refine the intervention. Phase I took place in an introductory college calculus course, and the intervention was taught by a cooperating instructor. As before, I begin with the design as conceived. Next, I analyze student surveys and exam scores to describe the impact of the experimental intervention. I follow up with an analysis of the design and implementation, and close with a discussion of results.

Students in the experimental section performed significantly better than their comparison counterparts on the common department exams. Students in the experimental section performed better on conceptual problems as well as purely procedural problems. These improvements were also evidenced by a (marginally significant) 13% improvement in success rates (A, B, C's in the course).

Phase I also provided theoretical insights. In terms of design, the peer-analysis form was modified to focus on communication rather than directly asking about a peer's reasoning. The explicit focus on communication facilitated meaningful feedback and analysis. The importance of adequate training also became evident during Phase I. These insights were used to improve the final version of the intervention, which was implemented in Phase II.

4.2 Design

The four components of the design were all revisions from the pilot study: (1) reflective questions, (2) opening problems and self-assessments of learning, (3) peer-assisted reflection, and (4) opportunities for student explanations. Reflective questioning replaced the ineffective explanation framework from the pilot study to promote reflection related to the process of problem solving, not just the outcome. Opening problems and reflections looked similar to those in the pilot study, but the number of prompts was reduced and the logistics of the process were streamlined. Peer-assisted reflection was an outgrowth of the peer-analysis and self-reflection activities from the pilot study. Finally, verbal opportunities for student presentations were recognized as crucial to ensure that students had an authentic audience for their explanations. I summarize the characteristics of the phase I design in table 4.1.

Strand of Design	**Key Features**
Reflective Questioning	• Questions about reasoning, justification, conceptualization • Poster at front of class reminds students of questions • Questions used for reflection and problem solving
Opening Problems & Self-Assessments	• Handed out on a single piece of paper • Opening problem pushes students to think about lesson • Rate understanding, ask two questions, tell something else
Peer-Assisted Reflection	• Cycle: attempt problem, reflect, peer conference, revise • Some training opportunities through sample work • Parallel self-reflection and peer-feedback forms
Student Explanations	• Written explanations, class discussions, PAR conferences • Regular problem solving in small groups • Regular student presentations (especially opening problems)

Table 4.1: Phase I Design Summary

4.2.1 Reflective Questioning

Reflective questions were intended to move students away from simply *doing* mathematics (e.g., plug and chug) to reflect on the meaning of what they are doing (cf. Schoenfeld, 1985). Students can use reflective questions both as a tool to facilitate problem solving and to probe their own understanding. The underlying message was that if you can't answer these questions, your understanding is not as deep as you might think. There were three basic questions:

- Why would you. . . ?
- Why can you. . . ?
- What does it mean that. . . ?

These questions are exemplars of categories of questions (see appendix A for the full list of questions). Through metacognitive modeling, the instructor demonstrates how to use these questions to the students. By repeatedly asking these questions of the class, students were intended to integrate them into their regular practice, just as other interventions that promote metacognitive development through social practice (cf. Palinscar & Brown, 1984; Schoenfeld, 1985; White et al., 2009). The use of a small number of questions was intentional, to make such reflective practice easier for students to internalize.

4.2.2 Opening Problems and Self-Assessments of Learning

Students spent five minutes at the beginning of each class working independently on the opening problem. The opening problem was provided on a double-sided sheet of paper with the self-assessment questions on the back, so that students could begin to work immediately upon entering class. The three self-assessment questions asked students to rate their understanding of the lesson from 0% to 100%, ask two questions, and tell something else of importance (see appendix D).

Self-assessments allowed students to ask questions that they would otherwise not have time to ask or not feel comfortable asking. This helped the instructor practice formative assessment by regularly responding to student difficulties. Also, practice articulating one's uncertainties was meant to help students learn to better self-reflect and become independent learners who could recognize and remedy their own struggles. By rating a 0 to 100% value for each session, students would have a better idea of what to study for exams.

4.2.3 Peer-Assisted Reflection

Peer-assisted reflection (PAR) took place on a weekly basis, providing opportunities to apply the standard form developed in the pilot study on a special PAR homework problem. Students: (1) completed the PAR problem (individually), (2) self-reflected on their understanding, (3) traded with a peer and gave peer-feedback during class, and (4) revised and turned in a final solution. When students traded work with a peer, they spent about 5 minutes filling out peer-feedback forms silently, and then about 5 more minutes to discuss their feedback and the problem more generally. Students were forced to first analyze their partner's work before their discussion to make sure that students focused on one another's reasoning, not just the problem more generally. The activity was designed to help students internalize insights from peer-analysis for improved self-reflection (cf. Black et al., 2003). The self-reflection form, peer-feedback form, and tasks can be found in the appendices (B, C, and G).

Students practiced explanation; gave, received, and utilized feedback; and practiced analyzing others' work. PAR feedback was timely (before an assignment was due; cf. Shute, 2008) and the activity structure (submission of both a initial and final solution) supported the closure of the feedback cycle (Sadler, 1989). Repeated practice analyzing a variety of examples is crucial to transition from external feedback to self monitoring (ibid). Through the use of hypothetical work, students were trained to meaningfully engage in PAR. PAR positioned students as authors of mathematics (in solving nontrivial problems) and as contributors to the work of others through peer-feedback (Engle, 2011), to promote positive mathematics identities (Boaler & Greeno, 2000). Because students were not asked to judge one another's work, the PAR design addressed the expert barrier discovered in the pilot study (Reinholz, 2013a).

PAR tasks were inspired or modified from a number of sources: the Shell Centre, the Mathematics Workshop Problem Database, *Calculus Problems for a New Century*, and existing homework problems from the course. Modifications provided more opportunities for explanations and multiple approaches (e.g., solution paths, generation of examples). The PAR tasks can be found in appendix G.[1] PAR tasks were numbered based on when they were assigned (i.e., PAR1 was assigned during the first week of class (due the second), PAR2 was assigned the second week of class (due the third), and so on).

4.2.4 Student Explanations

Students frequently explained their ideas both verbally and in written work. In order to provide an authentic audience (evidenced as important in the pilot study), students regularly explained their work at the board and through the PAR activity structure. Finally, students were given regular opportunities to collaborate with one another on problems during class. Even office hours (outside of class) focused on students working out the problems and explaining them to one another rather than teacher exposition.

4.3 Method

4.3.1 Participants

Phase I (fall 2012) took place in a university-level introductory calculus course targeted at students majoring in engineering and the physical sciences. The course met four days a week for 50 minutes at a time. About 300-500 students register each semester, participating in about 10 parallel sections. Most sections have a class size of 30-40, but there are also a few large sections with 50-100 students. The intervention was implemented in a single experimental section, and three comparison sections were observed to provide a baseline for standard instructional practices. I will use pseudonyms to refer to the instructors of the various sections.

For phase I, the experimental instructor, Michelle, was a PhD in mathematics education, with nearly 10 years of teaching experience. Michelle also taught one of the comparison sections. The other two comparison sections were taught by a full-time instructor with over a decade of experience teaching mathematics (Heather) and an advanced PhD student in mathematics who was graduating that semester (Logan). All instructors had taught the course a number of times before. As a researcher I conducted all observations, interviews, and collected samples of student work. Homework assignments were graded by undergraduate graders, except for the PAR assignments, which I graded personally with the assistance of a trained grader. Exam grading was performed by the group of instructors for the course, as well as a handful of undergraduate graders (for a total of approximately 10-12 graders for each exam).

[1] I provide the prompts for the PAR tasks during Phase II; changes in tasks between phases I and II were minor, except for the second and eighth PAR problems, which were redesigned for phase II. Also, phase II introduced extension questions as part of the tasks.

4.3.2 Data Collection

I collected data in one experimental section and three comparison sections (the four sections in the study). The exam scores from the remaining sections were used to compare exam scores to the department average. In the four observed sections I scanned students' exams, performed video observations, and offered a mathematics belief survey for extra credit. In the experimental section, I also scanned students' PAR assignments and daily self-assessments of learning, and made audio records of students' conversations during peer-conferences. Because instructors in each section were free to choose their own homework problems (from a provided pool of problems), homework was not included as a part of the data collection. A summary of the data collected is given in table 4.2. These are the numbers for students who remained in the course after the W-drop date,[2] so the original enrollments were higher.

	Experimental	Comp. 1	Comp. 2	Comp. 3	Comp. (Other)	Total
Instructor	Michelle	Michelle	Heather	Logan	-	-
Students	56	18	38	67	222	401
Participants	53	17	29	54	-	163
Pre-survey	48	14	29	41	-	132
Post-survey	44	12	29	48	-	133
Video Obs.	45	6	6	6	-	63
Interviews	14	-	-	-	-	14
PAR Conf.	66	-	-	-	-	66

Table 4.2: Phase I Data Collection Summary

4.3.2.1 Exams

In this coordinated course, all students completed three common midterm exams and one common final. Therefore, I calculated the department average scores in addition to the observed sections.

4.3.2.2 Surveys

I administered identical mathematics beliefs pre- and post-surveys using Qualtrics. If students completed the survey or an optional homework assignment they would receive extra credit. The surveys consisted of seven open response items and 18 quantitative items from the Modified Indiana Mathematics Beliefs System (MIMBS; see appendix E),[3] which had

[2] If students withdrew from the class before this date, they would receive a W on their transcript, which does not influence their GPA.

[3] To ensure consistency with the other belief scales, which focus on positive mathematics beliefs, I reversed the coding of the nature of mathematics scales. In my scale, items that focused on mathematics as conceptual were rated positively, and items that focused on mathematics as procedural were reverse coded.

been used in prior studies of the same student population (Pilgrim, 2010). I used three constructs from MIMBS: (1) self-efficacy ($\alpha = 0.76$), (2) nature of mathematics ($\alpha = 0.55$), and (3) concepts ($\alpha = 0.81$). I chose self-efficacy ($r = 0.201, p < 0.001$) and nature of mathematics ($r = 0.155, p = 0.001$) due to their correlation with the first midterm, a significant predictor of performance.[4] Although the concepts construct was not a significant predictor, I included it due to theoretical interest.

4.3.2.3 Video Observations

I performed all video observations with two stationary video cameras, one for the teacher, one for the class. Non-consenting students were assisted to sit out of the range of the cameras. As a researcher, I attended all class sessions, taking field notes of student behaviors and class discussions.

4.3.2.4 Interviews

After the second midterm, I interviewed students in the experimental section. The entire class was solicited for interviews, and all students who agreed were interviewed. Students were offered no incentives to complete the interview. Interviews took place in a one-on-one interview setting, with a semi-structured protocol. In appendix F I provide the interview protocol.[5] All students were asked at least the set of questions listed in the protocol; other follow-up questions were asked as appropriate.

4.3.2.5 PAR and Self-Assessments of Learning

I scanned students' PAR packets (initial and final solutions, self-reflections, and peer-feedback), opening problems, and self-assessments (approximately 10,000 pages total). During the first two weeks I captured peer-conferences with video cameras, but the audio quality was poor, students were hesitant to be recorded, and it was logistically difficult in a small classroom with 60 students. Starting with week three I instead captured peer-conferences using less intrusive handheld audio recorders. I could record up to seven conversations in a full room of 60 students, with high-quality audio (although gestures were lost).

4.4 Analysis

4.4.1 Survey Results

Student responses to the seven open-ended questions were very short, generally consisting of less than a single complete sentence. As a result, these responses were not analyzed,

[4] Self-efficacy was also a significant predictor ($r = 0.400, p < 0.001$) of final exam score.

[5] The interview protocol given is from Phase II; it was similar to protocol used in Phase I, but it also contained some additional questions.

and I instead focused on the quantitative scales. The scales focused on: (1) students' self-confidence, (2) beliefs that mathematics is non-procedural, and (3) beliefs about the value of concepts in mathematics. I now provide sample questions from each of these scales (the entire survey can be found in Appendix E):

1. **Self-confidence:** I feel I can do math problems that take a long time to complete.
2. **Mathematics is non-procedural:** There is no procedure to solve many math problems.
3. **Concepts:** In addition to getting a right answer in mathematics, it is important to understand why the answer is correct.

All of these beliefs were measured on five-point Likert scales, with a five representing a high endorsement of the belief, and a one representing low endorsement. Table 4.3 shows the changes in beliefs in the experimental section. As the table shows, all changes in belief were not statistically significant in the experimental section.

Belief	Pre-survey	Post-survey	Change	Paired t-test ($df = 47$)
Self-Efficacy	3.75	3.83	0.08	$t = 0.8998$, $p = 0.3728$
Nature of Mathematics	3.37	3.43	0.06	$t = 0.8804$, $p = 0.3831$
Concepts	4.09	3.98	−0.11	$t = -1.4868$, $p = 0.1437$

Table 4.3: Changes in Beliefs in the Experimental Section (On a Five-Point Scale)

Table 4.4 shows the changes in beliefs in the comparison sections. The table shows that students in the comparison sections decreased in their beliefs about the nature of mathematics; after completing a semester of calculus, they believed that mathematics was *more* procedural in nature.

Belief	Pre-survey	Post-survey	Change	Paired t-test ($df = 107$)
Self-Efficacy	3.60	3.70	0.10	$t = 1.8311$, $p = 0.0699$
Nature of Mathematics	3.32	3.14	−0.18	$t = -2.9870$, $p = 0.0490$*
Concepts	3.95	3.96	0.01	$t = 0.1596$, $p = 0.8735$

Table 4.4: Changes in Beliefs in the Observed Comparison Sections (On a Five-Point Scale)

To compare changes across sections, I computed change scores (post-test minus pre-test) and conducted independent samples t-tests for each belief scale.[6] As one would expect

[6] If either the pre-test or post-test score was missing for a respondent I imputed the mean score for their section.

based on Tables 5.3 and 5.4, the differences in mean changes in beliefs (+0.06 vs -0.18) that mathematics was nonprocedural were significant ($t = 2.5045, p = 0.0137$). There were no significant differences for the other scales (self-efficacy: $t = 0.1988, p = 0.8429$ and concepts: $t = 1.3255, p = 0.1883$).

4.4.2 Exam Performance

The guiding principle for exam grading (from the course coordinator) was to score "what students know" rather than "how well students communicated what they know." Because poor communication necessarily obscures students' ideas, poorly written responses tended to score lower, even though the quality of the communication was not a part of the scoring. Nevertheless, responses with highly elaborated reasoning tended to score the same as those with less elaborated reasoning. Considering the following prompt and two sample responses:

> If $\int_0^1 f(x)dx = 4$, then $\int_0^{1/2} f(x)dx = 2$. If the statement is true, explain how you know it's true. If it is false, give a counterexample.

Now consider the following solutions to the problem (see figure 4.1 and figure 4.2).

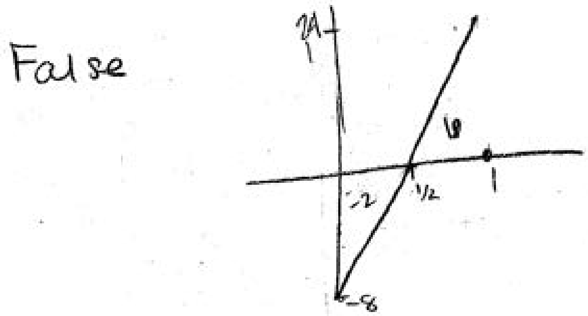


Figure 4.1: Exam 3, sample solution 1

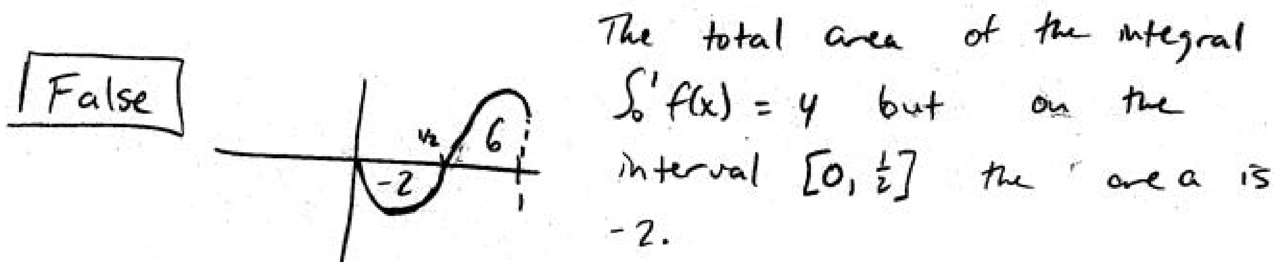


Figure 4.2: Exam 3, sample solution 2

Both solutions provide a similar counterexample, yet the second also provides an explanation why it is a counterexample. Both of these solutions received 4/4 points on the exam, even though the second was more elaborated. Other problems on the exams were graded similarly. Keeping this scoring philosophy in mind, I will compare student scores on exams for the experimental and comparison sections. Before doing so, I explicate the exam design and scoring in greater depth.

4.4.2.1 Exam Design and Logistics

Common exams were administered in the evenings, approximately every four weeks. The first three exams covered material from the preceding 3-4 weeks, while the final exam was comprehensive, with a slight emphasis (approximately 1/3) on topics after the third midterm. The final exam was weighted as 28.5% of students' final grades.

Exam writing typically began $1\frac{1}{2}$ weeks before the actual date of the exam. Each exam was written by a five-member team, consisting of the course coordinator, Michelle, myself, and two other instructors for the course (chosen by the course coordinator).[7] After a few members created an initial draft, the course coordinator flagged problems for modification or removal, suggested new problems to be added, and revised wording for consistency with prior semesters, the study guides, and textbook. This draft was sent to two other team members who engaged in individual review. The course coordinator reviewed the comments, and once again a larger meeting commenced to finalize the exam. In general, the exams were designed to consist of about 2/3 conceptual problem-solving and explanation problems, and about 1/3 rote use of skills and procedures.

Elaborated study guides (3-4 pages per exam) guided exam review and design. These study guides, given to students 2-3 weeks in advance, consisted of a detailed list of the expectations for conceptual understanding and some examples of potential problems for the exams. Each exam was reviewed to ensure it covered a breadth of topics on the study guide, and didn't cover anything not on the study guide. In addition, students had access to prior exams from 2-3 semesters prior, which gave them a reasonable idea of the types of problems they could expect and the format of the exams (even though all exams were created from scratch).

Exams were graded by the group of all course instructors, the course coordinator, myself, and a number of undergraduate graders (for an approximate total of 10-12 graders for each exam). Grading was blind (with exams from all sections mixed together), with each problem delegated to a single team of graders (typically 2-3 members), to ensure objectivity. Partial credit was assigned for all problems other than pure computations of limits, differentiation, and integration. To receive full credit on other problems students needed to show their work and explain their reasoning.

[7]The two additional members rotated on each exam, as chosen by the course coordinator. Thus, the 5-member team was different for each exam.

4.4.2.2 Exam Content

All exams had a similar format (see appendix H for a sample exam). As a result, the exam problems fit into a simple typology, given in table 4.5.

Problem Type	Description
Problem Solving and Explanations	Non-rote mathematical problems. Over 80% had multiple parts and required written explanations.
True/False	Students must explain why it is true, or provide a counterexample and explain why it is false.
Pure Computation	Procedural practice of limits, derivatives, and integrals. Graded all or nothing.
Miscellaneous	Multiple choice, fill-in-the-blank, and curve sketching problems. No written explanations required.

Table 4.5: Midterm Exam Problem Types

Problems in the first two categories (problem solving and true/false) all required written explanations, and comprised about 2/3 of each exam. The other types of prompts (pure computation and miscellaneous) did not require written explanations.[8] Thus, even though communication was not an explicit focus of grading, there were considerable amounts of explanation included on each exam. On average the midterm exams had nine questions, mostly multi-part, and the final exams were slightly longer. By percentage of total possible points, the breakdown of points for each exam is given in table 4.6.

Type	Exam 1	Exam 2	Exam 3	Final Exam
Problem Solving	45.9%	49.1%	46.2%	53.3%
True/False	18.3%	15.4%	15.4%	20.0%
Pure Computation	22.9%	19.2%	19.2%	26.7%
Miscellaneous	12.9%	16.3%	19.2%	-

Table 4.6: Percentage of exam points by problem type

[8]The only exception to this occurred on exam 2 during spring 2013, in which a 4-part computation problem had a single prompt of explanation, which was worth four points. Due to the way test scores were gathered, it was not possible to disaggregate these points from the remaining computation portion of the question.

4.4.2.3 Results

I first analyzed student success rates (number of A,B,C's in the course vs D,F,W's). The experimental section had a success rate of 82% while the comparison sections had a success rate of 69% (difference of 13%). I ran chi-squared to test for the significance of this effect ($\chi^2 = 3.4247, df = 1, p = 0.0642$), which was marginally significant.

Table 4.7 shows the average exam scores by condition. These scores provide the average for all of the exams simultaneously, rather than considering each exam separately (i.e., the experimental section was considered as having 218 exams total: 56 for Exam 1, 56 for Exam 2, 54 for Exam 4, and 52 for the Final Exam). I tested for significant differences in exam scores across conditions. The results can be found in Table 4.8. The results show that the experimental section performed significantly better than students in the comparison section taught by the same instructor. Students in the experimental section also performed significantly better than students in the comparison sections with different instructors. The effect sizes are small, according to Cohen (J. Cohen, 1988). Taken together, these results demonstrate the impact of the PAR intervention.

	Mean Exam Score	N
Experimental	73.03	218
Comparison (same instructor)	66.84	71
Comparison (all other sections)	67.32	1261

Table 4.7: Average Exam Scores, By Condition (Percentage Out of 100)

Exp. vs. Comp.	Independent samples $t-$test	Effect size	Sign.
Same Instructor	$t = 2.0291, df = 99.456, p = 0.04512$	$d = 0.28$	*
All Other Sections	$t = 4.2018, df = 307.005, p = 3.465 \cdot 10^{-5}$	$d = 0.31$	**

Table 4.8: Statistical Comparisons of Average Exam Scores

Table 4.9 shows statistically significant differences in the exam scores, looking at each exam individually. The table gives the average exam scores for the experimental section (row 1, $N = 56$), the comparison section taught by the experimental instructor (row 2, $N = 18$), and the average exam scores for all other sections not included above (row 4, $N = 335$). Numerically, the experimental section performed consistently better throughout the semester, with the gap between experimental and comparison sections increasing throughout the semester. Recall the small sample size of the comparison section with the experimental instructor, which accounts for the fact that those t-tests were not significant.

	Exam 1	Exam 2	Exam 3	Final Exam
Experimental	70.2%	79.8%	74.4%	67.3%
	$N = 56$	$N = 56$	$N = 54$	$N = 52$
Comparison (same instructor)	66.7%	75.3%	67.9%	56.9%
	$N = 18$	$N = 18$	$N = 18$	$N = 17$
Difference (same instructor)	**3.5**	**4.5**	**6.6**	**10.4**
t/p-values	0.69 / 0.50	1.03 / 0.31	1.06 / 0.30	1.32 / 0.20
Comparison (all other sections)	67.2%	75.1%	66.3%	60.2%
	$N = 327$	$N = 321$	$N = 311$	$N = 302$
Difference (all other sections)	**3.0**	**4.8***	**8.1****	**7.1***
t/p-values	1.18 / 0.24	2.20 / 0.03	3.08 / 0.003	2.20 / 0.03

Table 4.9: Phase I, Mean Exam Scores (Significant for $p < 0.05$*, $p < 0.01$**)

To test for significant changes in exam scores from the first exam to final exam in each section, I conducted a repeated measures ANOVA with between-subjects factors. Each of the four exam scores was treated as a repeated measure of the same outcome variable, and I compared the outcomes across the three groups (Michelle experimental, Michelle comparison, and all other comparison sections). The assumption of sphericity was violated (Mauchly's $W = 0.890, \chi^2 = 42.574, df = 5, p < 0.001$), so the Greenhouse-Geisser correction was used. There was a significant main effect for time ($df = 2.791, F = 43.073, p < 0.001$) and Group ($df = 2, F = 3.879, p = 0.022$), but interactions between group and time were not significant ($df = 5.581, F = 0.902, p = 0.487$). Following up with the post-hoc test Tukey's HSD (for Group), Michelle's experimental section vs. all other comparison sections was the only significant result (Difference of means $= 6.419, SE = 2.32, p = 0.016$). Michelle's experimental section vs. comparison section was not significant (Difference of means $= 6.8, SE = 4.42, p = 0.274$). These results are consistent with the t-tests shown in Table 4.9.

Comparing the experimental section to all other sections with a different instructor, the effect sizes were as follows: $d = 0.32$ (Exam 2), $d = 0.45$ (Exam 3), $d = 0.32$ (Final Exam). These indicate small to medium effect sizes, according to Cohen (J. Cohen, 1988). Using the above typology of problem types (see Table 4.5), I disentangle student performance on different problem types on each exam. Given the focus on explanation and deep conceptual understanding, one would expect that students in the experimental section would perform better on these specific problem types. However, given the link between conceptual understanding and more robust procedural knowledge, one might also suspect that they would do better on more rote types of problems as well. See table 4.10 for the results.

	Exam 1	Exam 2	Exam 3	Final Exam
Problem Solving	1.63	2.27	4.08	6.06
t/p-values	0.77 / 0.44	1.09 / 0.28	1.41 / 0.16	1.73 / 0.09
True/False	11.38**	8.28**	7.27*	7.45*
t/p-values	2.94 / 0.004	2.68 / 0.009	2.25 / 0.03	2.33 / 0.02
Pure Computation	1.48	10.47**	19.52**	5.42
t/p-values	0.38 / 0.70	3.11 / 0.003	4.9 / $< 10^{-5}$	1.57 / 0.12
Miscellaneous	10.31**	-0.27	4.21	-
t/p-values	3.53 / 0.0006	-0.08 / 0.94	1.70 / 0.09	

Table 4.10: Mean (Percentage) Differences Between Experimental and Comparison (including Michelle) by Problem Type (Significant for $p < 0.05$*, $p < 0.01$**)

As shown in Table 4.6, different problem types were weighed more heavily than others in the exam (e.g., many more points were assigned for problem solving as compared to multiple choice questions). As a result, Table 4.10 shows scores scaled by the total number of possible points in that problem type category, in order to make comparisons more meaningful. Thus, a 100 in this table would mean that the experimental section received all possible points on average, and that the comparison section received 0 points on average. Positive numbers mean that the experimental section scored higher, while negative numbers mean the experimental section scored lower. Note that for these analyses I have collapsed all comparison sections together; due to the small sample size of Michelle's comparison section, differences between her two sections were not statistically significant.

Numerically, students in the experimental section virtually scored higher across the board (including procedural problems). The only exception was miscellaneous problems on exam 2, for which there was a negligible difference in performance. The experimental section performed statistically significantly better on all exams on the true/false problems. One would expect that extra practice explaining one's mathematical ideas and larger exposure to classmates' thinking would definitely be supportive of increased understanding in this area. In particular, the ability to assess the quality of mathematical work might also be supportive of the ability to assess the quality of (hypothetical) mathematical claims (i.e., the hypothetical claim that the conjecture is true or false).

Somewhat surprisingly, students in the experimental section performed significantly better on the purely computational problems of the exams, especially on midterms 2 and 3. Recall that the problems in this category were purely computational, such as: compute this limit, derivative, or integral, and required no problem-solving. This gives evidence of a conceptually-oriented intervention providing significant gains for basic skills as well.

While the experimental section was considerably higher for these computational problems on midterms 2 and 3, the gains dropped down for the final exam. There were three purely computation problems on the final exam: one associated with limits, one for derivatives, and one for integrals. The scores for experimental and comparison sections (as percentages) are given in table 4.11.

	Limits	Derivatives	Integrals
Experimental	45.13	89.04	63.85
Comparison	49.68	78.34	51.97
Difference	**−4.55**	**10.69****	**11.86***
t/p-values	-0.95 / 0.34	2.84 / 0.005	2.32 / 0.02

Table 4.11: Phase I final exam pure computation problems (percentage out of 100)

The table shows that the experimental section performed quite a bit better on both the derivative and integral problems (both statistically significant), but actually performed worse than the comparison section for the limit computations (but the difference was not significant). There are a number of possible explanations. Looking at the PAR problems that were assigned, only two of the 14 were related to limits, and neither of the PAR problems involved limit computation. In that sense, the students didn't really experience the intervention in the context of limits. Also, limits took place at the beginning of the semester before students had much experience engaging with the experimental techniques. It is also possible that there were systematic differences in the types of review different sections did at the end of the semester, or perhaps other sections were simply more fluent with limit computations than the experimental section.

4.4.3 Comparison Sections

To contextualize the activities that took place in the experimental section, I provide a brief description of the standard practices that were observed in the comparison sections.

4.4.3.1 Michelle

Michelle taught the experimental section ($N = 56$) and one comparison section. Her comparison section was small ($N = 18$), and met early in the morning. Despite the early time of day, students generally showed up on time and participated in class activities. At least half of the students would participate on a given day, and students asked questions and were willing to go to the board to present material.

The instructor's teaching style was similar in the experimental and comparison sections she taught. All activities and lesson plans other than the experimental aspects of the intervention (PAR, opening problems, and self-assessments) were used in both sections. Michelle's teaching style was fast-paced and consisted of a large number of Initiation-Response-Evaluation (IRE) style interactions. In the experimental section the use of reflective questions sometimes slowed the class down and helped Michelle probe deeper into students' thinking.

In addition to IRE interactions, Michelle frequently utilized group work, typically one or more times a week. In the comparison section students were often slow to get started with group work, and many times students would work individually and only consult their group members regarding their answers. As a result, these students may not have benefitted fully during their group work sessions. In contrast, students in the experimental section seemed more willing to work together in their groups, perhaps due to the other features of the experimental classroom. In contrast to the other sections, Michelle presented no formal proofs to the class, only conceptual explanations; proofs had not appeared on exams for many years.

4.4.3.2 Heather

Heather's section ($N = 38$) met in the early afternoon. In contrast to Michelle's fast-paced style, Heather moved more slowly and probed the topics in greater depth. As a result, Heather was likely to cover fewer examples than Michelle, but more thoroughly. The level at which the class explored topics generally exceeded the depth required by exams and homework assignments. Students often asked difficult questions and did so spontaneously without prompting.

Heather frequently used humor in her class sessions, which sometimes detracted from the focus of the lessons. A few weeks into the semester I learned that Michelle and Heather shared many of the same in-class activities, but only those which were not a part of the experimental intervention. Many of these activities supported the same types of group work that took place in Michelle's class. In contrast to Michelle, Heather presented some formal proofs, but she did not stress their importance to her students.

4.4.3.3 Logan

Logan's section ($N = 67$) met in the afternoon after Heather's section. In contrast to the other two sections, Logan's class was almost entirely lecture-driven; there was no use of partner work or group work. Logan presented with a document camera and projector, which gave students no opportunities to present to the rest of the class. Lectures were very focused on mathematical formalisms and nuances of language (e.g., the difference between a theorem and a law). Most lectures consisted of the instructor introducing the topic and presenting a number of examples to the class.

Logan did afford some opportunities for student participation, typically in short IRE sequences. When more open-ended questions were asked, if students did not answer them quickly, the questions were simplified to make them more direct. Students frequently spent time on their phones, and one student in the class even brought a laptop to class to play a flight simulator.

4.4.4 Implementation of Intervention

4.4.4.1 Reflective Questioning

The researcher and Michelle co-constructed a poster to highlight the three major reflective questions (see figure 4.3). This physical artifact was intended to remind students to think of the questions as they solved mathematical problems.

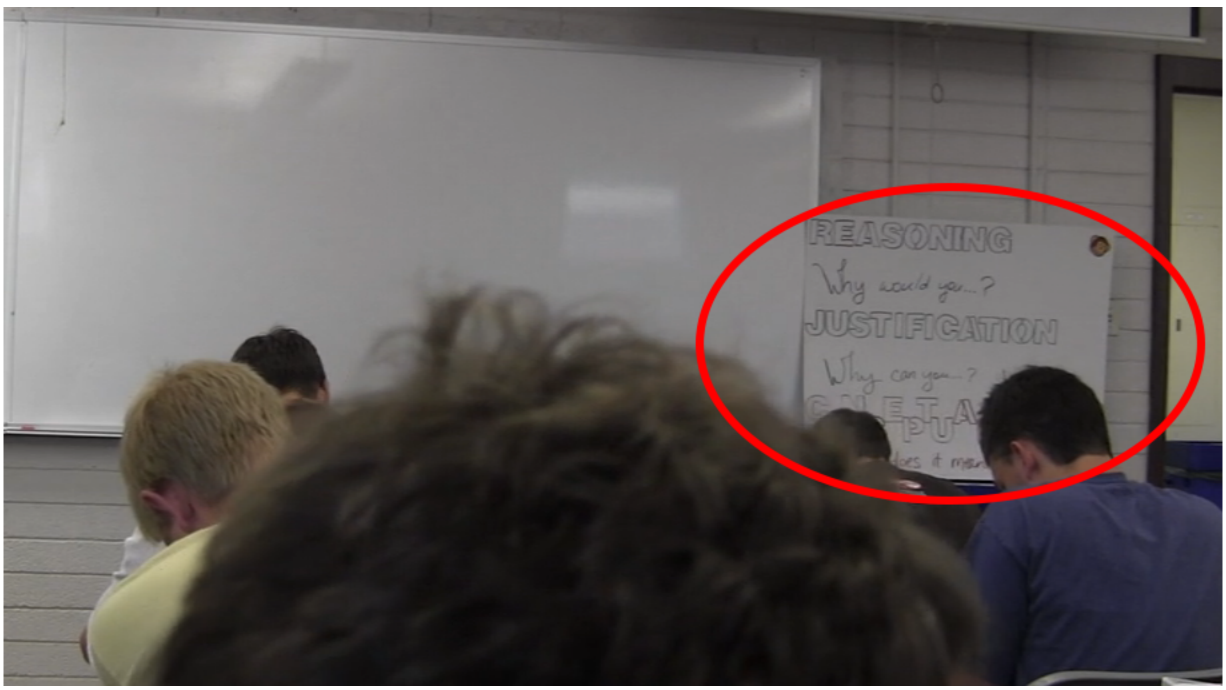


Figure 4.3: Poster with Reflective Questions

Michelle used the poster to reinforce her requests for explanation and justification, but didn't use the specific questions. I illustrate this with two examples from a class session on related rates problems. In the first episode, Michelle used the poster as a means to ask for further reasoning in a discussion about the relationship between the distance from a wall and the height up the wall of a moving ladder:

> Michelle: All right, so at 12 seconds, and this is going to be useful when we start building everything. We definitely know that the ladder is 15. What is x gonna be?
> Nathan: 10.
> Michelle: After 12 seconds?
> Nathan: Oh no, 7

Michelle: How did you get 7, please explain to me your reasoning (Michelle points at the poster).

Nathan: The ladder is moving in at a rate of 0.25 ft/s, and it's been moving for 12 seconds, so we'll move by 12 times 0.25.

Michelle points to the poster to ask the student to explain where they got their answer from, connecting the need for explanation and the physical presence of the poster. In episode two, the class was trying to find the rate of change of the volume of a cube with changing sides:

Michelle: So we've got all kinds of information here, let's talk about what we're after. What quantity are we after?

Laura: volume

Michelle: So volume. Just volume?

Nicki: rate of change of volume

Michelle: Now that's different, so which one is it? Do we want the volume, or rate of change of volume?

Ed: we want dV/dt.

Michelle: Ah, what indicates, give me reasoning and justification (points to poster), what indicates in that phrase that we're after dV/dt?

Ed: how is the volume changing.

Michelle: How is the volume changing. Right there, rate of change. That phrase tells us that we ultimately want dV/dt.

Once again, Michelle asks for reasoning and justification. Even though she doesn't use the specific questions, this type of probing into student thinking was not observed in Michelle's comparison section.

The above class session was an exception in terms of frequency of use of the poster; Michelle rarely used the poster, and by week 8 she had stopped referring to it altogether. During week 13 (of 15 weeks) Michelle stopped bringing the poster to class, after acknowledging that she wasn't really using it anymore. She noted even though she wanted to use it, her teaching style felt ingrained over many years of practice, so she forgot to use it. It is possible that the mere presence of the poster during most class sessions still had an impact, but only 2 of the 14 students interviewed referred to the poster this way. An excerpt from Robert's interview is given:

> Just seeing it every day, it makes you think about why would you do this, and all of those questions. Like it makes you think through that and for me it really helps with the explanations of everything. Why, how, all of those questions. It really helps me come up with accurate explanations and what graders are actually looking for.

Some students (6 of 14) noted that the poster was useful when the teacher pointed it out to them or that the instructor didn't use the poster often enough. When students did talk about the poster, many talked about using it mostly when stuck (6 of 14). For instance, when asked if he used the questions in his own learning, Matt responded:

> Big picture, I do think about why am I doing what I'm doing. Especially if I start to struggle now, I don't start to scramble and try to find an example or something someone told me in class. It's like, okay, why am I doing what I'm doing.

Jacob had a similar response:

> I feel like it's something where if you don't know where to start, you should start with that. When you're not completely clueless but stumped on something those are three guidelines you can take to get you in the right direction.

Although student use of the poster wasn't that high, there weren't any negative responses to the poster. It appears that, even weakly integrated, the poster and questions may have had some positive impact.

4.4.4.2 Opening Problems and Self-Assessments of Learning

Without the need to write down the opening problem, students began working on the opening problems immediately. The problems reviewed relevant concepts and introduced new ideas to help students get started thinking about the day's lesson. Opening problems provided many opportunities for student presentations, but student explanations were generally limited. I consider an episode from week 9 to illustrate a typical set of interactions. Students were given three conceptual true/false questions (explain if true, provide a counterexample if false) related to the first derivative test and critical points:

1. T/F: If $f'(5) = 0$, then there is either a local maximum or local minimum at $x = 5$.
2. T/F: If $f(2)$ is a local maximum, then $f'(2) = 0$.
3. T/F: If $f(7)$ is a local minimum at an interior point of D and $f'(7)$ is defined, then $f'(7) = 0$.

After a few minutes of silent work, Michelle requested that students come up to the board to show their work. Although students willingly presented their results to the task, Michelle limited their opportunities to fully explain their reasoning.

> Michelle: What about two? True or false?
> Mike: False
> Michelle: You got a counterexample or reasoning? Someone come and draw a counterexample.

Michelle: I can draw examples all day but can you guys?
Mike: Can a local max be an absolute max?
Michelle: Sure it can.
(Mike comes to board to draw)

Mike's work can be found in figure 4.4. After Mike drew his (correct) graph, Michelle elaborated his reasoning for him, rather than providing Mike with the opportunity to do it himself.

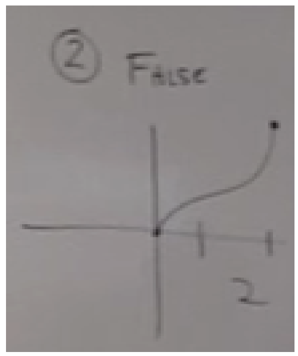


Figure 4.4: Mike's work on the opening problem

Michelle: So you got an endpoint.
Mike: Yes
Michelle: Right! Does it say an interior point? Not in the true/false statement. So we could be considering 2 could be an endpoint, and still be a local max. What else?
Lucy: It could be a minimum.
Michelle: Well we're talking about a maximum.
Evan: Since that graph has an endpoint, wouldn't it be absolute maximum instead of local?
Michelle: Can't it be both?
Evan: I dunno, can it?
Michelle: It can, think about it. In a neighborhood around this point (gestures to the top of the graph) is it the highest point?
Evan: Yes.

This episode was typical of how the opening problems were used. Students frequently presented to the class, but their responses were often limited to few words.

Students often referenced the opening problem in their self-assessment. Related to the above class session, a number of students asked what a critical point was. Ideally the instructor would be able to address this in real time during the class, but it was still important to know that students were still confused at the end of class. Other students provided feedback on the lesson, such as saying "true/false questions help a lot." A small subset of students asked irrelevant questions or left the prompt blank.

In interviews students noted that it was difficult to come up with questions, especially due to time constraints. To address student feedback, Michelle reduced the requirement of asking two questions about the lesson to only asking one question. Students were also provided with an alternative to asking questions: they could also generate a conceptual exam-style question if they felt like they had no questions to ask. Although it wasn't analyzed in depth, students generally rated their understanding levels as very high (80% or better), even when they asked questions on their self-assessments that indicated major conceptual misunderstandings. Nevertheless, students still may have found their relative ratings of understanding to be helpful.

Throughout the semester, Michelle frequently spoke about the usefulness of daily self-assessments outside of class, such as how they let her "know what's going on with her students on a day-to-day basis." Almost every few days Michelle would make notes about what she was noticing in the self-assessments, whether it was that "students were leading into the second derivative," that they have been asking "less-content related questions," or that they were struggling with a particular concept. Perhaps most telling, Michelle continued to integrate opening problems and daily assessments into her teaching practice in subsequent semesters.

4.4.4.3 Peer-Assisted Reflection

PAR Conversations

I provide an example from PAR1 (found in appendix G.1) to illustrate the opportunities PAR discussions afforded for students to engage in mathematical thinking. PAR1 was concerned with the length and radius of tape on the spools of a cassette tape as it plays. Aaron and Nancy's conversation illustrates how students worked out their ideas in discussion. The students were trying to determine how the radius of the spool affected the overall radius of the tape on the spool:

> Nancy: I'm not sure, because as I asked my adviser, err, tutor, he says it's 2.5 centimeters, that's for the center, including the spool itself, it's not just the tape. And the spool itself is never gonna decrease.
>
> Nancy: So if you plot points like this, you're never gonna be at 0, when your tape is going to, it's not going to be anymore. So you almost have to change it to 1/2 centimeters, not 2.5 centimeters.
>
> Aaron: Yeah, that's what I was thinking about, but, I was working with some of the kids I was living with, and um, they said they would have done it the exact same way with the 2.5 and the 1. Saying, once it gets to 1cm then the tape is over. It's like, so you don't want it to be exactly to 0.
>
> Nancy: So it's acceptable, but not necessarily specific, in a sense.
>
> Aaron: When it's going to zero, it's really going to 1.

Both students used their individual experiences (working with a tutor or roommates) to support their ideas in discussion; PAR promoted deeper conversations because students had already spent time working on the problems before discussing them together. During the first few assignments students mostly discussed the mathematics, but over time students became more direct in their feedback. Consider an example from PAR5:

> Sam: On number four, you said that this, like $f(a + h) - f(a)$ over h is the secant line. It would be the slope of the secant line. Cause that would be your average rate of change. And then...on...this one. You were pretty close except the slope of a secant line is through two points and you're trying to evaluate it at just that single point, so it can't really have a secant line at that one point.

Sam gives very direct feedback to her partner about specifics of mathematical language, which demonstrates her increasing proficiency in giving feedback.

Students also used PAR to compare their solutions, as in this excerpt from PAR5:

> Amanda: OK. What did you think about 'c'? Cause you had something different for 'c'.
>
> Joe: Let's see. What was the question? Explain why a limit is needed to find the slope of a tangent line? Um...
>
> Amanda: Cause you wrote the limit needed to be this. And that makes sense, too. So,
>
> Joe: Well both are kind of the same thing because you're just saying if you plug in zero for the h then it would be like anything, like, I dunno. I can't really remember my thought process I just know that the limit from the example we used in class, where it was absolute value of x over x, that the limit didn't exist.

This is a small excerpt from a larger conversation. Because the students noticed discrepancies between their two solutions they were clued into an important area for them both to focus on. Students frequently made such comparisons.

Most PAR tasks worked well, but some had significant flaws. For instance, PAR8 was a curve sketching problem, which resulted in very shallow conversations (so the task was replaced for phase II). PAR2, related to various hypothetical definitions of a limit (with situations such as $\lim_{x \to a} f(x) = L$, $f(a) = L$, or $\lim_{x \to a} f(x) \neq L$), was so confusing that most students could not figure out how to even answer the question. Consider the following excerpt:

> Steve: Okay, so what's the question?
>
> Patrick: I don't understand the first question, because is seems like you come up with some random L, literally, and you hope to God your limit on the function you came up with isn't the L you came up with.
>
> Steve: Right, because that's what I felt it was like. Just pick something random and not close.
>
> ...

Patrick: I'm not sure how I got number 2. It was an explanation of how L and a are not related, but...

Steve: $f(x)$ approaches something, but it doesn't approach L. I don't know, I was confused.

...

Patrick: but if for this one, if it's the exact same problem, would it still work for question 2?

Steve: pretty much.

Both of these students were confused about the problem, but the PAR conversation did not help them resolve their confusion. This task was significantly reworked during Phase II, which resulted in less student confusion.

Other tasks received only minor revisions, such as PAR7. PAR7 was a related rates problem about a cone with increasing radius and decreasing height. Students were tasked to find the rate of change of the volume, and this task focused students solely on procedural conversations (see appendix G[9]). Nevertheless, a few students made interesting insights into the problem situation, such as the fact that with a decreasing height the volume of the cone would eventually have to go 0, no matter what the starting conditions were. These insights were used to improve the task for phase II.

A small subset of students had very short and superficial conversations. Consider the following discussion between Alex and Nicole, related to PAR10 (approximating the area underneath one's hand using simpler shapes).

Alex: Ridiculous.

Nicole: I think you did it right except for the last 3 parts.

Alex: Yeah.

Nicole: Do you know how to do it, just using triangles?

Alex: Yeah, I got that.

Nicole: You gotta add the ones underneath, and subtract the other ones.

Alex: Yep.

Nicole: It looks pretty good, and then for more accuracy, you could do some more triangles.

Alex: Even more triangles.

Nicole: And more triangles!

Alex: I said yours is awesome, and, yeah.

Nicole: Thanks.

Alex: Uh huh.

This was the extent of the students' conversation, which lasted less than a minute. Only a few students had such conversations, which occurred especially when these students repeatedly worked with their friends in the class. Michelle had no formalized method for making

[9]The version of PAR7 using during phase I only consisted of parts (a)-(c), which meant that there was essentially no explanation required of students.

students work with other students, so this small subset of students continually worked together and had superficial conversations. This observation prompted a modification to the intervention; during Phase II students were told to sit randomly as they entered the class on PAR day. This eliminated all superficial conversations of this type. I also added extension questions, which could be used to push students to continue thinking deeper when they felt like they were done.

Self-Reflection and Peer-Feedback Forms

Over time, flaws in the initial design became evident. In particular, the big picture prompt (what was the approach to solving the problem?) didn't support either self-reflection or peer-feedback. Consider Evan's response on PAR1,

> Use prior knowledge and common sense to determine the nature of the tape recorder in order to sketch graphs loosely depicting this.

or Sam's,

> Understanding the situation and visualizing the graphs.

These responses were typical, containing little to no information about the student's thinking. Such reflections are neither helpful for understanding the student's reasoning (from the perspective of a teacher) nor for generalizing from it (from the perspective of the student). To address this design flaw, the form was modified to ask students about communication rather than the approach itself. When students focused on communication, they ultimately analyzed one another's reasoning anyway. Students also struggled to determine whether or not they thought a problem was solved correctly. Accordingly, the prompts were changed to instead focus more explicitly on justification, to help students think about how to determine if a problem was solved correctly or not (see appendix C, phase I). Parallel changes were made to the self-reflection forms, giving students a number of check boxes to self-reflect (see appendix B, phase I).

In the days following these changes, a number of students came up to me and said that the new version of the form made much more sense and was easier to use. Although there was a definite improvement with the focus on communication, students still struggled to justify their answers. Typical responses to what evidence was provided were as follows (Cynthia):

> rational logic used when analyzing the graph that correspond to key calculus definitions

or (Mike):

> Most evidence does make sense but should probably be elaborated further.

These responses highlight that students did not a strong sense of how to justify their solutions.

Training Students

On the second day of class, the instructor walked through a sample analysis exercise with the students. Students were given the prompt:

> Train A leaves Fort Collins station traveling at 50 miles an hour on Track X. Three hours later, Train B leaves the station traveling 60 miles an hour on Track Y, which is parallel to Track X. How long does it take Train B to catch up with Train A?

This classic algebra problem affords many solution paths, so it is likely that students would solve it in a number of ways. After students worked through the problem individually, they were given a few minutes to complete a self-reflection, just as they would with PAR. Then, Michelle gave the students two sample solutions, and the students worked in pairs to fill out the peer-feedback forms as if they were responding to the given student. After working in pairs the class came back together to discuss what they had come up with, and Michelle offered her own advice as well.

Throughout the semester students had other opportunities to discuss the analysis of sample work, such as in the opening problems. Other times special activities were developed in response to student difficulties on PAR problems. For instance, after PAR5 students were given 3 sample responses to different prompts and asked specific questions to help them analyze the work and areas that might be lacking (for instance, poor use of pronouns). These activities were used sporadically. Although sample work could be collected from the class, it was difficult to collect samples and develop an activity that could be used in a timely manner. As a result, a large library of samples of common misconceptions, and work that highlighted other common mistakes, was gathered from phase I for use during phase II.

4.4.4.4 Student Explanations

Students had opportunities to present work during class, but their opportunities for explanation were often limited. The repeated use of quick IRE sequences during discussions usually limited student responses to fragments of sentences. Nevertheless, students had frequent opportunities to explain mathematics through group work and during PAR conversations.

4.5 Discussion

Quantitative results supported the experimental intervention. Students in the experimental section performed significantly better on the exams than the comparison section with the same instructor (average scores of 73.03% vs. 66.84%) and the comparison of all sections with different instructors (average scores of 73.03% vs. 67.32%). The experimental section also performed numerically better on nearly all aspects of the common exams. For the final three exams, the differences in overall exam scores were statistically significant. Most impressively, students made large gains on purely procedural skills, even though they were not an explicit target of the intervention. In contrast to students in the comparison section who developed unproductive beliefs that mathematics is procedural, the surveys showed no belief changes in the experimental section.

Michelle was positive about the intervention and continued to incorporate it into her teaching in subsequent semesters. The iterative nature of the PAR cycle supported students to go to office hours, to revise their work based on feedback, and to revise their work based on thinking about the problem more than once before turning in a final solution. Opening problems and self-assessments were generally successful. As Brad noted, the opening problems helped him get into "math mode" at the beginning of the lesson. Other students made similar remarks. Opening problems also afforded opportunities for student presentations. Daily self-assessments provided a wealth of information for the instructor, even though students often wrote overly favorable estimates of understanding and struggled to construct questions.

4.5.1 Revisions for Theory and Implementation

The design evolved considerably over the semester. The key lesson learned seemed to be as follows: focusing on communication helped students focus on one another's reasoning. Students struggled to articulate one another's reasoning based on the initial analysis forms, but this was remedied with a focus on communication. This forced students to focus on *what* it is that makes thinking visible, and ultimately it helped them better think about thinking. Simultaneously, it improved their communication skills. In addition to this major insight, a number of smaller changes were made to the intervention for Phase II.

1. To address the difficulties of training students to analyze solutions, a large corpus of sample student work was collected. This would serve as the basis for training exercises for phase II, which would occur on a weekly basis (or more frequently).
2. To address student difficulties with self-assessment (coming up with an accurate percentage), I introduced a two-column format into the opening problems. This would allow students to compare their initial and final solutions. Students were encouraged to do this multiple times during each lesson, so that they could assess how well they understood the material.
3. Given that a subset of students did not fully engage the PAR activity, I designed extension questions for each PAR assignment. The intent was that these would give students something to do if they finished early, so that all students would be engaged the entire time. Also, all students were sat randomly during PAR days so they worked with a large variety of partners.
4. The reflective questions and the categories of reflective questioning were modified slightly in ways that seemed more natural and well-suited for the problems.
5. PAR2 and PAR8 were replaced, and minor adjustments to the wording or the addition of explanations prompts were added to some of the problems.

6. Other minor changes were made to the self-reflection and peer-feedback forms, intended to simplify reflection and analysis.

Even though experimental instructor was able to effectively implement the peer-assisted reflection and daily self-assessment of learning activities, she struggled to engage in the use of reflective questioning during class sessions. Accordingly, I decided to teach an experimental section for phase II, in order to research the impact of the full implementation of the experimental intervention. Phase II was intended to replicate the findings of Phase I, and also determine the impact of revisions and full implementation.

5 Phase II

5.1 Chapter Overview

The purpose of Phase II was to extend and replicate the findings from Phase I. My goals were to assess if the experimental students still performed significantly better than their comparison student counterparts, and also what impact revisions to the design had on improving student performance. Phase II once again took place in an introductory college calculus course. As before, I begin with the design as conceived. Next, I analyze student surveys and exam scores to describe the impact of the experimental intervention. I follow up with an analysis of the design and implementation, and close with a discussion of results.

Students in the phase II experimental section performed significantly better than comparison sections on the common exams, with differences even larger than in phase I. Once again, these improvements were evident on both conceptual and procedural exam problems. These improvements were also evidenced by a significant 23% improvement in success rates (A, B, C's in the course). These improvements in performance help highlight the importance of design revisions, particularly regular training in analyzing solutions.

5.2 Design

The design consisted of revised versions of the same four strands as in phase I: (1) reflective questions, (2) opening problems and self-assessments of learning, (3) peer-assisted reflection, and (4) opportunities for student explanations. I summarize the characteristics of the phase II design in table 5.1.

5.2.1 Reflective Questioning

As in phase I, reflective questions were used to help students think more deeply about mathematics. In particular, students were intended to use reflective questions both in problem solving and in analyzing and reflecting on their own work. Through metacognitive modeling, the instructor was to demonstrate practices that the students would ultimately incorporate into their own thinking. Once again a poster was used as a classroom aid to make students more aware of these questions and categories. Given that Michelle struggled to utilize reflective questioning in phase I, I adjusted the three categories of questions to try to make

Strand of Design	**Key Features**
Reflective Questioning	• Questions about approach, justification, and meaning • Poster at front of class reminds students of questions • Questions used for reflection and problem solving
Opening Problems & Self-Assessments	• Handed out on a single piece of paper • Opening problem pushes students to think about lesson • Rate % understanding, ask one question, tell something else
Peer-Assisted Reflection	• Cycle: attempt problem, reflect, peer conference, revise • Regular training through darts activity • Random seating to ensure different partners • Four categories for reflection and feedback
Student Explanations	• Written explanations, class discussions, PAR conferences • Regular group problem solving at chalkboards • Regular student presentations (especially opening problems)

Table 5.1: Phase II Design Summary

them more natural to include in classroom discussions. The new categories were: approach, justification, and meaning. Approach extends beyond *what* one does to *why* one does it. Justification relates to how one knows a solution is right. Meaning relates to how one makes sense of math. The questions associated with these categories can be found in appendix A.3.

5.2.2 Opening Problems and Self-Assessments of Learning

As in phase I, students were handed a piece of paper containing the opening problem and self-assessment questions for the day. Students were given 5 minutes at the beginning of class to work on the opening problem, and a few minutes at the end of class to answer the three questions: rate your understanding of the lesson from 0% to 100%, ask a question about mathematics, and (optional) tell me something else I should know. Unlike during phase I, students were only required to ask a single question.

To help students self-assess, I introduced a two-column format for opening problems. The left column was for an initial solution (before class discussion), and the right column provided room for corrections after discussion, allowing for a direct comparison of the solutions. Students were strongly encouraged to use this format during class. The sum of one's performance on these checkpoints would provide a grounded rationale for self-assessment. For instance, a student who completed one of four problems without assistance would mark a low understanding (less than 50%).

5.2.3 Peer-Assisted Reflection

Peer-Assisted Reflection (PAR) followed a similar format to phase I. Once again, students: (1) completed the PAR problem, (2) self-reflected, (3) engaged in peer-conferences, and (4) revised and turned in a final solution. Students had about 10 minutes for peer-conferences: 5 for silent analysis, and 5 for discussion. Once again, PAR problems were completely weekly.

In phase II, self-reflection forms consisted of 6 prompts, grouped into four categories: (1) completeness, organization, and labeling, (2) explanations, (3) use of language, and (4) justification. These four categories were intended to guide self-reflection and peer-analysis, by creating important categories to focus on. The peer-feedback form was also streamlined. Prompts were simplified (and examples were provided) so that students would only have a single prompt to answer. The phase I prompt:

> Was their solution communicate clearly? (Give at least one suggestion to improve the presentation of the solution.)

was changed to

> Give at least one suggestion to improve the presentation of the solution,

This eliminated meaningless responses such as "yes" in the peer-feedback form.

I also added extension questions to each PAR, so that interested students could dig deeper into the mathematics. Pedagogically, extension questions made it easier to hold students accountable during PAR. Students who finished early or didn't appear to be working could be pressed to answer the extension questions. Additionally, two of the PAR problems (2 and 8) were replaced and minor changes to the wording and additional explanation prompts were added to other problems. These were added in places that they could be used to support students to make mathematical connections.

PAR training was made systematic in phase II, through a weekly activity called darts. Each Wednesday, the opening problem presented students with three sample explanations to the PAR problem they just turned in: the "bullseye" (correct explanation), "on the board" (a mostly correct idea that is communicated poorly or has a minor error), and "off the mark" (incorrect) solutions. Students had to classify the explanations, stating how to improve the incorrect solutions, which led into a class discussion. Through darts, students could calibrate their analysis to an expert perspective and see different ways of looking at a large corpus of exemplars.

5.2.4 Student Explanations

As in phase I, students regularly presented mathematics at the board. Additionally, the classroom during phase II had a dozen large chalk boards on the wall that were used for collaborative group work. This allowed for about 8-10 groups of 4 students each to work collaboratively on problems. Work was visible to all students but only one student could write at a time, which supported meaningful group engagement. This was in contrast to phase I in which students were limited to group work using pencil and paper.

5.3 Method

5.3.1 Participants

Phase II once again had a single experimental section with observations in 3 comparison sections (in the same course as phase I). I taught the experimental section, to better ensure full implementation of the design during this phase. Two of three comparison sections were taught by (unobserved) graduate student instructors from phase I. The third instructor was a post-doctoral researcher working on mathematics education projects. To maintain objectivity, a research assistant performed all observations and interviews for phase II. The grading procedure was the same as in phase I, with the research assistant providing assistance as a grader for weekly assignments.

5.3.2 Data Collection

There were 9 parallel sections of calculus. Most data collection procedures were the same, except for a few minor changes: (1) one camera was used rather than two to reduce logistical difficulties, (2) a research assistant conducted interviews and performed video observations, and (3) students were offered one extra credit homework assignment as incentive to give an interview. Table 5.2 summarizes the data collected (reported after the W-drop date).

	Experimental	Comp. 1	Comp. 2	Comp. 3	Comp. (Other)	Total
Instructor	Dan	Sam	Tom	Bashir	-	-
Students	38	37	31	28	205	339
Participants	38	34	27	24	-	119
Pre-survey	32	23	18	18	-	91
Post-survey	32	19	20	14	-	85
Video Obs.	54	6	5	5	-	75
Interviews	22	-	-	-	-	22
PAR Conf.	89	-	-	-	-	89

Table 5.2: Phase II Data Collection Summary

Additionally, Tom included the PAR problems as a part of his regular homework assignments, to provide a point of comparison for student growth.

5.4 Analysis

5.4.1 Academic Background

I collected data on students' academic backgrounds to verify that the comparison sections were similar to the experimental section on relevant background characteristics. There were no significant differences in comprehensive ACT scores ($t = 1.2986, df = 64.6, p = 0.1987$), but the average scores were numerically slightly lower in the experimental section (mean of 25.45 vs 26.3). Similarly, there no significant differences in high school GPA ($t = 1.2685, df = 54.773, p = 0.21$), but the average scores were numerically slightly lower in the experimental section (mean of 3.43 vs 3.56). There were no significant differences in gender ($\chi^2 = 0.0019, df = 1, p = 0.9649$) or race (comparing White students and students of color; $\chi^2 = 1.256, df = 1, p = 0.26$) between the experimental and comparison sections. Of the students who responded to these surveys, there were 87 males and 21 females; in terms of race, there were 18 students of color, and 90 White students.

5.4.2 Survey Results

As in Phase I, student responses to the seven open-ended questions were very short, generally consisting of less than a single complete sentence. As a result, these responses were not analyzed, and I instead focused on the quantitative scales. The scales focused on: (1) students' self-confidence, (2) beliefs that mathematics is non-procedural, and (3) beliefs about the value of concepts in mathematics. I now provide sample questions from each of these scales (the entire survey can be found in Appendix E):

1. **Self-confidence:** I feel I can do math problems that take a long time to complete.
2. **Mathematics is non-procedural:** There is no procedure to solve many math problems.
3. **Concepts:** In addition to getting a right answer in mathematics, it is important to understand why the answer is correct.

All of these beliefs were measured on five-point Likert scales, with a five representing a high endorsement of the belief, and a one representing low endorsement. Table 5.3 shows the changes in beliefs in the experimental section. The table shows a significant improvement in beliefs related to self-efficacy and the non-procedural nature of mathematics.

Table 5.4 shows the changes in beliefs in the comparison sections. Like the experimental section, there was a significant improvement in beliefs regarding the procedural nature of mathematics.

To compare changes across conditions, I computed change scores (post-test minus pre-test) and conducted independent samples t-tests for each belief scale. There were no significant differences (self-efficacy: $t = 1.292, p = 0.2$, nature of mathematics: $t = -0.0006, p = 0.9953$, and concepts: $t = 1.3094, p = 0.194$).

Belief	Pre-survey	Post-survey	Change	Paired t-test ($df = 33$)
Self-Efficacy	3.52	3.74	0.22	$t = 2.4731$, $p = 0.0187$*
Nature of Mathematics	3.02	3.25	0.23	$t = 2.7167$, $p = 0.0104$**
Concepts	4.04	4.10	0.06	$t = 0.8118$, $p = 0.4227$

Table 5.3: Changes in Beliefs in the Experimental Section (On a Five-Point Scale)

Belief	Pre-survey	Post-survey	Change	Paired t-test ($df = 73$)
Self-Efficacy	3.67	3.74	0.07	$t = 1.0441$, $p = 0.2999$
Nature of Mathematics	3.00	3.24	0.24	$t = 4.2712$, $p < 0.0001$**
Concepts	4.04	3.96	−0.08	$t = 1.0614$, $p = 0.2920$

Table 5.4: Changes in Beliefs in the Comparison Sections (On a Five-Point Scale)

5.4.3 Exam Performance

Once again, students in the experimental section showed considerable improvement. The experimental success rate (A,B,C vs. D,F,W) was 79% while the comparison success rate was only 56%. This difference in success rates (23%) was statistically significant ($\chi^2 = 6.3529, df = 1, p = 0.0117$), building on phase I's result, which was marginally significant (Phase I's difference in success rates was 13%).

Table 5.5 shows the average exam scores by condition. These scores provide the average for all of the exams simultaneously, rather than considering each exam separately (i.e., the experimental section was considered as having 139 exams total: 38 for Exam 1, 35 for Exam 2, 34 for Exam 4, and 32 for the Final Exam). I tested for significant differences in exam scores across conditions. Student in the experimental section performed significantly better than those in the comparison sections ($t = 6.5085, df = 184.817, p = 6.935 \cdot 10^{-10}$). The effect size was $d = 0.59$, which is medium, according to Cohen (J. Cohen, 1988).

	Mean Exam Score	N
Experimental	75.20	139
Comparison	64.17	1124

Table 5.5: Average Exam Scores, By Condition (Percentage Out of 100)

These differences are also evident in individual exam scores (see table 5.6). I include phase I results to facilitate comparison between phases I and II.

	Exam 1	Exam 2	Exam 3	Final Exam
Experimental	68.4%	81.7%	75.7%	75.7%
	$N = 38$	$N = 35$	$N = 34$	$N = 32$
Comparison	62.2%	66.5%	61.5%	65.9%
	$N = 303$	$N = 294$	$N = 273$	$N = 254$
Difference	**6.2**	**15.2****	**14.2****	**9.8****
t/p-values	1.71 / 0.09	5.00 / $< 10^{-5}$	4.45 / $< 10^{-4}$	2.80 / 0.007
Difference (phase I)	**3.0**	**4.8***	**8.1****	**7.1***
t/p-values	1.18 / 0.24	2.20 / 0.03	3.08 / 0.003	2.20 / 0.03

Table 5.6: Phase II, Mean Exam Scores. (significant for $p < 0.05^*$, $p < 0.01^{**}$)

Comparing the experimental section to the comparison sections for phase II, effect sizes were as follows: $d = 0.86$ (Exam 2), $d = 0.77$ (Exam 3), and $d = 0.48$ (Final Exam). These indicate large effect sizes for exams 2 and 3, and a medium effect size for the final exam, according to Cohen (J. Cohen, 1988). To test for significant changes in exam scores from the first exam to final exam in each section, I conducted a repeated measures ANOVA with between-subjects factors. Each of the four exam scores was treated as a repeated measure of the same outcome variable, and I compared the outcomes across the two groups. The assumption of sphericity was violated (Mauchly's $W = 0.915, \chi^2 = 25.171, df = 5, p < 0.001$), so the Greenhouse-Geisser correction was used. There was a significant main effect for time ($df = 2.841, F = 8.722, p < 0.001$) and group ($df = 1, F = 11.182, p = 0.001$), but interactions between group and time were not significant ($df = 2.841, F = 1.666, p = 0.176$). These results are consistent with the t-tests shown in Table 5.6.

Once again I analyzed exam scores by problem type (see table 5.7). The main story is unchanged; the experimental section mostly did better across the board. However, there are also noticeable differences. Drastic problem solving gains were present in phase II but not phase I, which may be attributable to the effective problem sessions and PAR training during phase II.

The results for computational problems were nearly replicated; students in the experimental section performed significantly better on midterms 2 and 3. Once again this prompted an in-depth analysis of the final exam (see table 5.8). Once again the experimental section did slightly worse (numerically) on the limit computation problems, which could be attributed to a variety of reasons (e.g., few PAR problems on limits, limits occurring early in the semester). Students performed considerably better than their comparison counterparts on the true/false explanation/counterexample questions, although the results were not statistically significant during phase II.

	Exam 1	Exam 2	Exam 3	Final Exam
Problem Solving	7.65*	11.16**	16.11**	11.27**
t/p-values	2.27 / 0.03	2.74 / 0.009	5.58 / $< 10^{-5}$	3.89 / 0.0003
True/False	6.15	6.50	9.6	4.82
t/p-values	1.40 / 0.16	1.84 / 0.07	1.52 / 0.14	0.85 / 0.40
Pure Computation	-0.21	11.53**	14.33**	4.83
t/p-values	-0.05 / 0.95	3.99 / 0.0002	2.88 / 0.007	1.19 / 0.24
Miscellaneous	8.26	18.20**	10.10*	-
t/p-values	1.61 / 0.11	3.15 / 0.003	2.34 / 0.02	

Table 5.7: Mean (Percentage) Differences Between Experimental and Comparison Sections by Problem Type

	Limits	Derivatives	Integrals
Experimental	53.96	88.13	64.16
Comparison	56.08	78.41	57.27
Difference	**-2.13**	**9.71***	**6.89**
t/p-values	-0.40 / 0.69	2.42 / 0.02	1.20 / 0.23

Table 5.8: Spring 2013 final exam pure computation problems (percentage out of 100)

5.4.4 Comparison Sections

To contextualize the activities that took place in the experimental section, I provide a brief description of the standard practices that were observed in the comparison sections.

5.4.4.1 Sam

Sam's section ($N = 37$) met in the early afternoon. The class sessions fell under two categories: lecture and board work. During the lecture sessions, Sam engaged in long abstract discussions of the mathematical concepts (e.g., spending 30 minutes on the meaning of *approach* when it comes to limits). During these lecture sessions, students mostly listened; for instance, during the first observed lecture (week 2) students had a total of three speaking turns during the entire session, and only one of them was a complete sentence.

In contrast to the lecture sessions, Sam also put a strong emphasis on what he called board work. On the first day of class, Sam had students form groups of four students, and the same groups remained throughout the semester. During the board work sessions each group of students would work on problems at a chalkboard (there were enough boards for each group to have its own workspace). Of both of the board work sessions observed, students worked on problems the entire time while Sam circulated around the room answering questions. Students appeared to be engaged with the material and discussing ideas with the other members of their group. The majority of the problems were procedural (e.g., students spent an entire class session practicing derivatives with the product, quotient, and chain rules).

5.4.4.2 Tom

Tom's section ($N = 31$) met in the afternoon after Sam's section. By the second week of class Tom knew most of his students' names. Class sessions always proceeded in whole-class discussion, but Tom asked lots of questions to include students in the discussions. Student responses were typically complete sentences, but students rarely elaborated their thoughts with multiple sentences. Students frequently asked questions when they were confused or didn't follow steps on the problems being worked on the board. Tom frequently asked his students whether or not they were following along the lecture, and he re-explained ideas when students expressed confusion. Tom discussed a number of minor theorems that were not discussed at all by other instructors (e.g., Darboux's theorem).

5.4.4.3 Bashir

Bashir's section ($N = 28$) met later in the afternoon. Like Sam's lecture sessions, Bashir engaged in long abstract discussions of the mathematical concepts. Attendance in Bashir's section appeared to be very high. All class sessions were whole class lecture, with students mostly listening. Bashir's section moved through the material quickly, and he was typically one-two sections in the syllabus ahead of the other instructors. When students made errors, Bashir responded to correct their thinking.

5.4.5 Implementation of Intervention

5.4.5.1 Reflective Understanding (and Reflective Questioning)

Reflective questioning was used to elicit student reasoning, beyond just IRE sequences. I provide an example (from week 4) to illustrate how both the instructor and students used the reflective questions to guide the discussion. I put a number of infinite limits problems on the board for students to work on, with the relevant problems reproduced here:

$$\lim_{y\to\infty} \frac{(y+1)(y^2+1)}{y^2+2} \qquad\qquad \lim_{h\to-\infty} \frac{h+1}{h^2}.$$

Reflective questions were intended to help class think deeply even about these procedural computations. Throughout the course I had repeatedly connected *approach* and *why did you do what you did*, so I started the discussion by asking students about their approach to cue that thinking. Another reflective question, *How do you know?*, was used to probe deeper reasoning.

Dan: Let's look at something here like number 1. What's going on here with number 1? How would you start to think about approaching this problem?
Dan: Yeah, Ivan.
Ivan: As y is approaching infinity. You can just kind of think of it as the top part is growing faster than the denominator, so it would approach a positive infinity.
Dan: How do you know it's growing faster?
Ivan: Because it has the y^2 times the extra y.
Dan: So what's that going to give us in terms of y's?
Ivan: y^3
Dan: So I agree with your general reasoning, what would we want to do to show that limit?

The discussion continued until the problem was solved. Afterwards, I asked students if there were any questions. Two students brought up questions, using the basic format of reflective questioning: *How do you know?*; and *Why did you?*. I used the third type of question: *What does that mean?* These basic questions were regularly integrated into classroom discussions.

Brook: how do you know the top is going to infinity and the bottom is going to 1?
Dan: the limit of 1 is one and the limit of 2 over y squared is 0. It's 2 over a really big number; which is...
Students: 0
Dan: And here we have a similar kind of thing. These terms go to 0, but then we have an actual y there which goes to infinity. Once you do the division you can see it's going to be the highest power that matters; everything else goes to 0.
Travis: why did you multiply by one over y squared?
Dan: what would you multiply by? What else could we multiply by? Why did we need to multiply at all?
(Silence)
Dan: What did we have when we evaluated the limit without multiplying?
Bill: infinity over infinity
Dan: Infinity over infinity. And what does that mean?
Bill: Not much.

Shortly thereafter, I used the same method to introduce the next problem.

Dan: let's look at this second one. How do we approach this?

Kevin: I'm horrible at fractions so this is a shot in the dark, but couldn't you just factor an h out of the bottom and divide out?
Dan: Could we factor out an h here and cancel the two h's?
(A number of students shake their heads no)
Dan: Why not?
DJ: You have to be multiplying all of the top.
Dan: So we have to have an h in every term in the top to do that. How else could we think about approaching this?

These above excerpts illustrate how the questions were integrated into classroom discussions. To address questions of impact, I turn to student interviews. Surprisingly, a number of students (10 of 22) stated that they either never used the poster, or that they had no idea what the questions even were. Some examples of students who never used the poster are given (Peter):

> I don't think past the first week I've even looked at it or acknowledged it (the poster). I don't even know what the questions are that the monkeys are asking, so I don't think it's beneficial. It's kinda cool though, it adds a little twist to the class; unique.

or Cory:

> They don't usually pop into my head at first. I've been solving math problems for a really long time, so to get new ways of thinking about something stuck in your head right away is kinda hard.

A number of the students who did mention using the poster or questions said they did so subconsciously (6 of 22). For example (Brook):

> I feel like I have unconsciously or subconsciously. They are like justifying stuff. I don't feel like I actually ask those questions in my head, but the path that I go down with my work gets me to the same ending.

Based on student interviews it appears that the questions were seldom used by students outside of class, at least consciously. In fact, a few students mentioned that they might use it in class but outside of class they simply forgot about it. It appears that the primary impact of the reflective questions and poster was to help facilitate discussions in class.

5.4.5.2 Opening Problems and Self-Assessments of Learning

Opening Problems

I provide an example of how students used the two-column format to answer the following opening problem:

Suppose you have the following table of values for a function $f(x)$:

$f(x)$	1	10	50	1000	500,000,000
x	3	0.0003	0.000001	0.000000005	0.00000000000000009

True or False: Based on the table above, we can conclude that $\lim_{x \to 0} f(x)$ Does Not Exist (DNE). Explain your answer.

Katherine's sample response is given in figure 5.1. Notice that Katherine's initial solution on the left side was incorrect, and the work on the right side reflects the notes she took based on the in-class discussion.

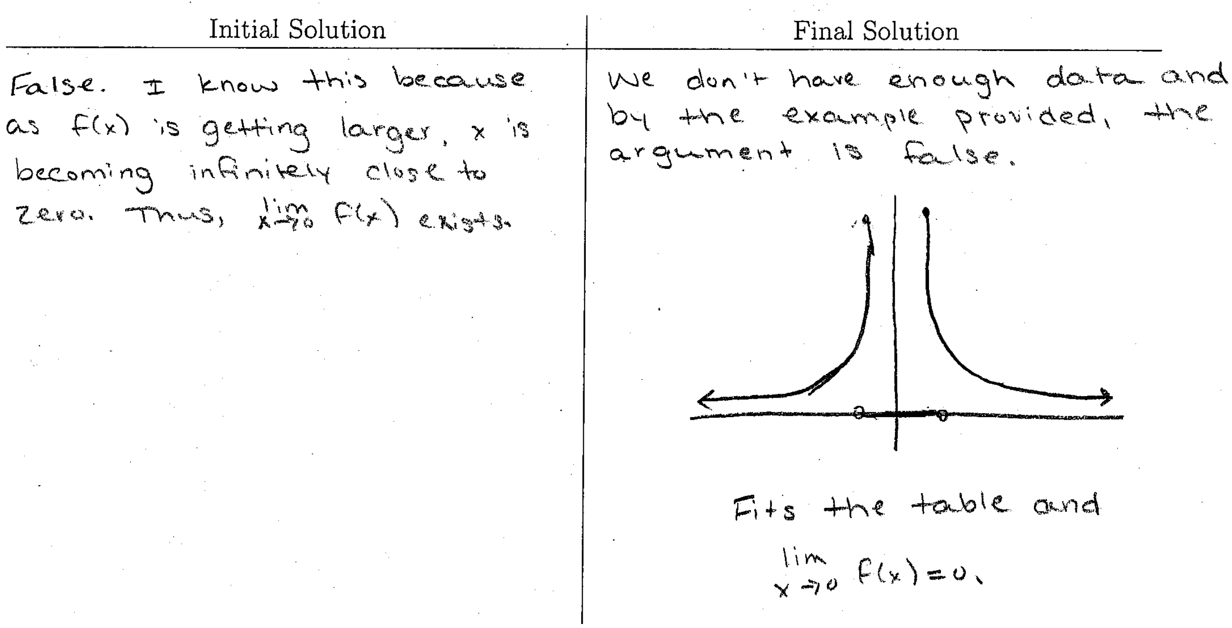


Figure 5.1: Sample opening problem response.

The right-hand column was not just copied down from the board; only the figure was drawn on the board. Whether or not it is possible to make any claims about how well the student understood the problem after the in-class discussion, the two-column format provides a clear demonstration that the student initially did not understand the problem. Forcing students to "put their cards on the table" in this way was designed to help them better self-assess.

Interviewed students virtually unanimously responded that opening problems were helpful (21 of 22). Students noted opening problems helped get into "math mode" (11 of 22). For example, DJ said:

> I like it. It gets you into the mood for math. It makes you think about math right away, so when you come from psychology and you're thinking about the human brain, you actually get into the mindset of math.

Many students (13 of 22) noted the usefulness of two-columns. Students noted how the two columns helped them connect their work to the correct solution. For instance (Tyler):

> That's probably the most helpful part, because I can look back and see my mistakes and compare it to what's right.

or Samantha:

> I think it's helpful to have a space where I can just write what I know and what I think might be helpful, and the final. It's really helpful to have this is where I started and this is where I should be. So if I had nothing, then I know I need to study that. Or if I have the right thing, then I draw an arrow to show what happened.

Students also made connections to PAR and the importance of being able to write down one's thoughts without worrying about being right (Mark):

> I find it helpful. It's like PAR, where you get a second chance. You don't feel pressured to get the right answer because you have space over here to write the right answer down. You can just spill your thoughts onto the initial column and I like it.

In combination with PAR, the two-column format helped see their solutions as a work in progress, which led to greater persistence, as I discuss in the following chapter. Nevertheless, only two students noted that they used the two-column format in their notes, and only when explicitly prompted by their instructor.

Self-Assessments of Learning

Daily self-assessments provided a wealth of information to the instructor, and the "tell me something else box" provided opportunities to build relationships with students by talking about things often not related to mathematics. To understand the extent to which self-assessments supported students directly, I turn to the student interviews. Students overwhelmingly responded that they struggled to come up with questions, particularly due to time constraints. Most students appeared to see the questions solely as a means for remedying confusions, not to learn more about specific topics. A few quotes highlight this perspective well. As Brook said:

> The second part, where we come up with a question, sometimes is really hard. People around me are all like looking around, that I can't think of a darn question to ask. But in some ways I think that's more helpful because you have to dig deeper and not think more surface level about math-related concepts. But most of the time I'll just ask a question from webwork or a conceptual one. Sometimes it's really hard to think of a question, which is kinda frustrating sometimes. Leaving it to the end of class kinda sucks, because then you're rushed and frantically trying to think of something.

Or as Todd responded:

> Sometimes the reflections are nice because you have questions you can answer in class. Other times it is tedious, because you can't ever think of a question, so you always put some dumb question you already know the answer to just to get points. But it helps sometimes.

Other students, like John, noted the usefulness of questions:

> I like those, because for me, if there's a question that I thought of that maybe I didn't really want to ask in front of the class, or maybe if there's a question that I thought about and wasn't sure and then Dan moved on, it's a good way to ask that and get feedback every day.

While asking questions was useful for some students, students might be served better if it were an optional, rather than mandatory, requirement.

The other aspect of the self-assessments that students seemed to find useful was ranking a percentage of understanding each day. As Kevin noted:

> It's helpful because I could look back at which days I struggled, and on the front side it tells you exactly what we did that day. If it's a day I struggled I could open up the text book and read or do some more problems on my own.

John had similar thoughts:

> If you really don't understand something and that percentage is low, it helps you identify that as an area to really go back and study for the test.

Although the absolute accuracy of self-assessments is questionable, being able to rate their relative understandings between lessons appeared to support students in their learning. Not a single student made a negative comment about percentage ratings.

5.4.5.3 Peer-Assisted Reflection

During phase II, PAR included regular training opportunities for students with the darts activity. I now provide an example conversation that came out of this activity. The sample comes from PAR5, which was focused on the relationship between secant and tangent lines. This particular problem was chosen because it is one of the problems that I analyze in-depth later when considering student explanations. To begin the class session students picked up the opening problem with the three sample solutions on it, and spent a few minutes working silently. The prompts are given:

Instructions: Classify the “bullseye” (correct explanation), “on the board” (a mostly correct idea that is communicated poorly or has a minor error), and “off the mark” (incorrect) solutions.

Prompt: Write a short statement explaining in terms of your graph what $f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$ means (be sure to talk about secant lines and the tangent line).

Sample Solution 1: $\frac{f(a+h) - f(a)}{h}$ is the secant line. As h goes to 0, it becomes like the tangent line.

Sample Solution 2: As the secant line becomes ever so small (to the point we can substitute its value for 0) we can use that average rate of change to generalize for the instantaneous rate of change.

Sample Solution 3: $\lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$ means that as the distance between a and $a + h$ approaches 0, the slope of the secant line through $(a, f(a))$ and $(a + h, f(a + h))$ becomes a better and better approximation of the slope of the tangent line at the point $(a, f(a))$, the derivative $f'(a)$.

In the discussion that followed, students made many useful analytic observations, even before the second midterm.

Dan: So without further ado, darts darts darts. What do you guys think about number 1? yes, Revati.

Revati: I thought it was on the board, because it defines what the secant line is, and it kinda explains what’s right, but it’s not in depth enough.

Dan: What do you guys think? Do you agree with that or disagree with that?

(A few students nod or give thumbs up.)

Dan: What else did you notice about this one, that maybe wasn’t perfect?

Kevin: It doesn’t talk about what the derivative is.

Dan: It doesn’t really talk about, well, it doesn’t say that the tangent line relates to that up front. So, there’s that. Aamir, what do you think?

Aamir: It doesn't explain the limit
Dan: It doesn't talk about the limit right, it says as h goes to 0, but it doesn't say in the limit. Ivan?
Ivan: It doesn't mention the slope at all
Dan: It doesn't mention the slope at all! What is $(f(a+h) - f(a))/h$?
Revati: That's the slope of the secant line.
Dan: It's the slope of a secant line, it's a number. So if you call a number a line that's a problem.

I now continue to the next solution.

Dan: What about number 2, what do you think? What do you guys think about number 2? All right Mark, what do you think?
Mark: Off the mark.
Dan: Why is this off the mark? Why do they hit the drywall?
Mark: Well, I was bothered that they use the words 'ever so small.' That's not a very mathematically helpful description.
Dan: Yeah, what does ever so small mean? Colton, what else do you think?
Colton: It doesn't say anything about the whole equation.
Dan: Okay, what about the equation. It doesn't refer to what?
Colton: The different parts.
Dan: Okay, it doesn't say how the different parts relate to the equation.
Peter: It doesn't say anything about the secant line
Dan: It doesn't say anything about... well it says the secant line, but not the tangent line. Anything else?
Peter: The secant line doesn't become ever so small.
Dan: Right, the secant line doesn't become ever so small, what is becoming ever so small?
Peter: h
Dan: The h, the distance between those two points. Any other problems?
Dan: I'm also not quite sure what it means to generalize to the instantaneous rate of change, but other than that, not bad.
Dan: Do we all agree the first one is on the board? What about 3?
Harry: bullseye
Dan: yeah, there weren't many choices left.

Students' improved analytic skills were evident in PAR conversations and the quality of their peer-feedback, although in-depth analyses of PAR conversations are beyond the scope of this dissertation.

Students apparently benefited from the random assignment of partners for peer-conferences; phase II had far fewer superficial conversations recorded than phase I. Nevertheless, I only recorded two instances of students discussing extension questions, so they were mostly ignored. In contrast, the students in Tom's section regularly completed the extension questions, even though they knew that they were not graded. Finally, the self-reflection form, peer-feedback form, and tasks all worked effectively, so no revisions are suggested for future use.

5.4.5.4 Student Explanations

Students frequently explained their mathematical reasoning through: reflective questioning, in-class presentations, PAR, and group work. Typically, when the syllabus afforded two class periods for a topic, the second class period was dedicated to open-ended problem solving in small groups working at chalk boards. Review sessions for exams were also run as group work sessions. The following excerpt from a review session illustrates the way in which students worked collaboratively. The prompt is given first:

> True/False. If true, explain why. If false, provide a counterexample and explain why it is false. Also, ask yourself what would make the statement true. If $f(x)$ is concave down at $x = \pi$, then $f(x)$ has a local maximum at $x = \pi$.

I now provide the transcript. The ability to draw and gesture to a shared space facilitated group discussions, allowing students to annotate one another's drawings (as Cory does):

> John: Is there any way we could have a concave down, where maybe it's like, not finished or something? I feel like if it's concave down then you have to have ...
> Cory: Can you have um...
> Francis: Well, I mean, there's the, uh, x (draws x^3), well, whatever. Maybe if we shift it to the right? This is concave down,
> Ivan: Uh, that's concave up.
> Francis: Yeah, for example, so if we shift it over here (draws in dotted lines with inflection point of x^3 at π), this would be π.
> Cory: But then it's an inflection point.
> Francis: yeah, because at that point it's neither concave up or down
> Ivan: If it's like this though (draws a single hump of the cosine graph, with the maximum slightly to the right of π; see figure 5.2)
> Cory: I think, so say here's the max (labels "not π")
> John: Um, technically on a whole I think this interval here would be...
> Ivan: Isn't that whole interval concave up? or concave down sorry.
> Francis: Yeah, that's down. Yeah, not necessarily. Because any point between these two intersections is concave down, but they have to define the derivative at that point has to be 0 for that to be the maximum.
> Ivan: So, is that a counterexample? If that's the max?
> Francis: Yeah.

John: That would make sense. I kinda want to ask Dan if every one of these points would technically still be concave down. I feel like the easiest way to determine if you have a maximum is to check the slope on either side of the point. I know the second derivative test is a quick way. . .

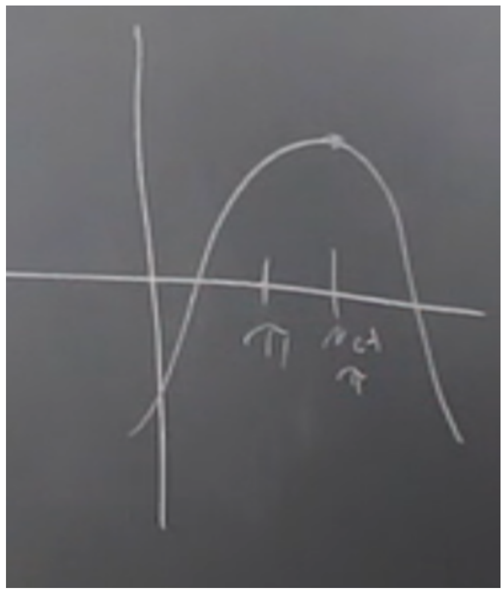


Figure 5.2: Second figure drawn during groupwork.

This excerpt highlights the potential of the shared space of the chalkboards for supporting student conversations.

5.5 Discussion

Phase II replicated and extended the results from phase I. During both phases, students had improved success rates; they were 13% higher during Phase I (marginally significant at $p = 0.0642$), and 23% higher during Phase II (significant at $p = 0.0117$). Students also had improved exam scores; for instance, final exam scores were 9.8% higher during Phase II ($p = 0.007$), and 7.1% higher during Phase I ($p = 0.03$). These differences were larger during phase II, which seems to be attributable to improvements in the design (e.g., darts, random PAR partners, and collaborative board work). In the following chapter I further explore these differences by analyzing student explanations in depth.

5.5.1 Revisions for Theory and Design

There were no major theoretical changes after this semester of implementation, but there are two main areas revision for the design: reflective questioning and daily-self assessments.

While the reflective questions seemed to help facilitate classroom discussions, few of the students remembered or used the questions on their own, so the impact is questionable. Accordingly, for future uses of the design it is likely that one could use another form of reflective questioning, such as Schoenfeld's questions (Schoenfeld, 1985) or ideas from Accountable Talk (Michaels, O'Connor, Hall, & Resnick, 2010) with a similar effect.

The other area of revision is related to the self-assessments of learning. While students seemed to find assigning percentage scores to their level of understanding to be useful, many students struggled to come up with questions. One way to take the design would be to further develop the ability to ask insightful questions through classroom activities and practice. Another alternative would be to transform the asking of questions aspect of the self-assessment into an optional component and instead focus on some other sort of reflection, such as the development of lifelong learning skills (e.g., collaboration, persistence). This is the most likely path I will take in the future. Given evidence that PAR has helped students improve their mathematical self-efficacy beliefs, having students reflect on dispositions and beliefs towards mathematics could be a powerful way to supplement those changes.

6 Analysis of Student Explanations

Analysis of students' exams showed considerable growth on traditional measures, but the exam format required students to provide short (1-2 sentences) explanations under strict time constraints. Also, recall (as explained in phase I) that exams were scored for "mathematical correctness," not the quality of communication. I turn to homework data to address the shortcomings of the traditional exams. In the following analyses I focus only on students written explanations; all other components of the problems are ignored.

6.1 Measuring Explanation Quality

I developed four criteria for measuring the growth of students' explanations over time: accuracy, use of mathematical language, clarity, and use of diagrams. In more detail:

- Accuracy: The extent to which an explanation correctly represents a mathematical truth.
- Mathematical Language: The quality of the mathematical language used to describe the idea.
- Clarity: The extent to which an explanation clearly and unambiguously describes *why* a statement is true.
- Diagrams: The extent to which supporting diagrams are used to bolster a written explanation.

The first three criteria can be applied to any explanation, whereas diagrams may not always be relevant (e.g., Euclid's proof of infinite primes is unlikely to benefit from a diagram). These four criteria are intended capture how well an explanation communicates *why* something is true. A basic measure of an explanation is its *accuracy*; i.e., whether or not it describes the mathematics correctly. But accuracy does not imply quality of expression; hence the other criteria. *Mathematical language* concerns the accuracy and precision of the words used to describe the mathematics; more technical language is not necessarily preferred, because it may obscure communication. *Clarity* relates to both the clarity of expression and the extent to which the explanation describes *why* something is true. While clarity of expression and describing why something is true could easily be separated in theory, in scoring it was easier to consider them simultaneously. Finally, *diagrams* can improve the clarity of written explanations.

These four dimensions alone were insufficient to develop an analytic scheme; they did not capture the interplay between accuracy and quality of expression. When students explained mathematical ideas incorrectly, they often drastically simplified the problem situation, making it much easier to describe their incorrect reasoning. Yet, if an explanation lucidly explained the mathematics in a wrong way, it did not seem appropriate to classify it as being a high quality expression. I illustrate by providing example responses to one of the PAR problems used in the course, which will be analyzed in depth later. Students were given the definition of a derivative of a function f at a point $x = a$ as defined by

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}.$$

Students were given three tasks: (1) illustrate the definition of a derivative graphically; (2) connect the slopes of secant lines to the slopes of tangent lines; and (3) *explain why the limit is needed to find the slope of the tangent line. (Why can't we just use arithmetic and algebra?)*

For illustrative purposes, I focus only on (3). Students were to connect the slopes of secant lines and the slopes of tangent lines, as described in (2), to explain why the limit is required. To contextualize student responses, I provide a complete explanation of the mathematics: the average rate of change between two points (the slope of a secant line), can be used to approximate the instantaneous rate of change at a single point (the slope of the tangent line). The smaller the interval over which the average rate of change is calculated, the better the approximation will be. Ideally, one would like to choose an interval of zero width, but this is not possible, because the average rate of change is defined on a finite interval. Attempting to evaluate the average rate of change at a single point would result in an undefined output:

$$\frac{f(a+0) - f(a)}{0} = \frac{0}{0}.$$

The number $f'(a)$ can be approximated as accurately as desired by calculating the average rate of change over small enough intervals; the limit represents the logical conclusion of this process of approximation.

Keeping this explanation in mind, consider the following student explanations. The first example is:

> "The limit is needed because otherwise we would have the slope of a secant line."

This explanation is written in a way that is relatively easy to follow. It correctly states that the equation written without a limit would describe the slope of a secant line. However, this information does not explain why arithmetic and algebra cannot be used to find the slope of the tangent line; this explanation does not address the mathematics of the question asked. Contrast this with the second explanation:

> "The limit is necessary because if h was zero instead of just approaching zero, it would result in a zero over zero output."

This explanation begins to address why the limit is required: if an average rate of change was calculated using the same point twice, the output would be zero over zero. Since average rate of change is a method for computing a slope that involves arithmetic and algebra, the explanation addresses the question asked. Yet, the explanation does not reference arithmetic or algebra explicitly, and it still does not explain the role of the limit; the explanation is lacking.

The first explanation clearly describes the mathematics incorrectly, while the second explanation begins to explain the correct mathematics in a less clear way. Explaining correct ideas is often much more difficult than explaining incorrect ideas, so incorrect explanations may appear to be clearer, at least at a surface level. This tension was resolved by focusing on the *core mathematical idea*. Above, the core mathematical idea was: average rate of change can only be calculated with two distinct points on the curve, so it cannot be used to find a tangent line, which (locally) intersects the function at a single point.

Rather than analyzing explanations for some general notion of communication, I used the core mathematical idea to guide analysis. The first explanation above does not express the core idea at all; it does not even reference average rate of change. Because the explanation does not express *this* idea, it cannot be considered a clear explanation of *this* idea, even if the statement is relatively easy to comprehend. In contrast, the second explanation begins to address the core idea, by referencing a computation of average rate of change. Although this explanation is still incomplete, compared to the first explanation, it is a clearer explanation of the core idea. Thus, by focusing on the extent to which explanations describe the core mathematical idea, I was able to resolve the tension between accuracy and clarity. Combined with the four analytic dimensions given previously, this provided a content-neutral analytic scheme.

6.2 Method

6.2.1 Scoring Procedures

Student work on PAR problems 5, 10, and 14 was analyzed, because these problems: (1) spanned and separated the semester into thirds, (2) required considerable written explanation, and (3) had the same explanation prompts in both Phase I and Phase II.[1] These problems are henceforth referred to as the early semester (PAR5) mid-semester (PAR10) and late semester (PAR14) problems. I chose a subset of the prompts that required meaningful student explanations; there was one prompt for the early semester problem, two prompts for mid-semester problem, and two prompts for the late semester problem.

[1]Some of the other problems received significant revisions, or were even removed entirely.

To choose a representative sample of work, I used the 10-point scoring rubrics that were used for course grading for each PAR problem; these rubrics were consistent across instructors and phases of instruction. For each PAR problem, I broke each class into quintiles according to these scores. I removed the top and bottom quintiles from the analysis, to eliminate any outliers (e.g., students who had already mastered the material or students who attended very few class sessions). The middle 60% of students remained; these quintiles represented, high, medium, and low scores on the PAR problems, respectively.

I used a random number generator to select three students from each of these groups in each section. The result was 27 samples of student work for each problem, (3 sections of the course: Phase II, Phase I, Comparison) x (3 groups of students: high, middle, low) x (3 randomly sampled students per group). Across the three problems, I had a total of 81 samples of student work. Because PAR10 and PAR14 had two prompts scored for each, there were a total of 135 prompts scored. Using photoshop I removed all grading marks, comments, and student/class identifiers; this made it impossible to tell that the problems had been graded, let alone what a grader might have written. I assigned random codes to the de-identified student work in order to hide which section each sample came from.

Michelle (the Phase I experimental instructor) and I worked together to blindly double code all of the student work. Before scoring I created a rubric and codebook that elaborated the rubrics for each prompt, using samples of student work at all levels of the rubrics (taken from student work that was not sampled for double-coding). After Michelle reviewed these materials independently, we began the blind coding process. After the first two problems were scored we had a short discussion about our coding rationale. Afterwards, we scored the remaining problems and had a discussion once scoring was complete. Through discussion each coder had the opportunity to change their initial coding to reach a consensus score.

6.2.2 Scoring Rubrics

Explanations were scored along four dimensions related to how well they expressed a core mathematical idea: accuracy, mathematical language, clarity, and diagrams. Analyses required that the core mathematical idea be established before any scoring could take place; agreeing on the core mathematical idea was unproblematic for the present analysis. Once established, the four dimensions were scored according to that core mathematical idea. Individual prompts on PAR problems all addressed only a single core idea. In situations where a single prompt addresses multiple core ideas, an explanation could be scored multiple times, once for each core idea.

Accuracy describes how correctly the explanation describes the core mathematical idea. Explanations with no accuracy don't express the core idea at all, and as a result, they tend to receive a 0 on all other dimensions as well.[2] *Mathematical language* is the accuracy and precision of the words used to describe the core mathematical idea. Mathematical language that is not related to the core idea is not scored for this dimension. *Clarity* focuses on

[2]Only a single solution that scored 0 on accuracy received a score higher than 0 on any other dimension.

the extent to which an explanation clearly and unambiguously describes *why* a statement is true. *Diagrams* that are used to bolster an argument are scored based on their effectiveness. An appropriate diagram can improve a students' scores on other dimensions as well (e.g., by making connections that weren't apparent in the written text). The scoring rubrics are given in Table 6.1.

Dimension	0	1	2
Accuracy	The core mathematical idea was incorrectly expressed or not expressed at all.	The core mathematical idea was mostly expressed, with some inaccuracies.	The core mathematical idea was expressed correctly.
Mathematical Language	The mathematical language was inappropriate or incorrect.	The mathematical language had some minor inaccuracies.	The mathematical language was appropriate and correct.
Clarity	The explanation was difficult to follow; OR the explanation was built on a misunderstanding of the core idea.	The explanation was easy to follow, but did not express "why"; OR the explanation expressed "why," but was unclear.	The reasoning was clear and easy to follow, AND expressed the why of the situation.
Diagrams	The diagram was seriously flawed and did not illustrate the core idea.	A diagram partially conveyed the core idea; OR a diagram conveyed core idea with some inaccuracy.	A diagram clearly illustrated the core idea.

Table 6.1: Rubrics for Scoring the Quality of Student Explanations.

The rubrics use a 3-point scale: (0) completely incorrect, (1) partially correct, (2) fully correct (to the level of sophistication of an introductory calculus student). All dimensions must always be scored, except diagrams, which could receive NA, because the use of a diagram was optional. Completely incorrect (but present) diagrams received a 0.

6.3 Illustrative Growth Trajectories

To illustrate student growth over time, I present an "average" response from each section for each problem. Although it was possible to compute mean scores for each section, it was not possible to find exemplars that corresponded to all of these sets of scores. Median was also not a useful measure, because it did not capture the significant improvement in the higher scores in the experimental sections. As a result, I chose the highest scored responses in the middle quintile of each section. These responses are illustrative of the types of growth found in each section, as will be evident when they are compared to the section average scores.

I provide the full problem statement from each problem (with relevant prompts in italics) as well as a brief solution to contextualize the student explanations that follow. Next, I present the three exemplar solutions: from the comparison section, from the Phase I experimental section, and from the Phase II experimental section. These average solutions are presented to illustrate student growth qualitatively; I follow up by presenting the aggregate results for each section.

6.3.1 Early Semester

The early semester problem focused on the definition of the derivative. Students were asked to illustrate the definition of the derivative graphically and explain the connection between secant lines and tangent lines. Finally, students were asked why a limit is needed to find the slope of a tangent line. The task is given in its entirety, but only part (c) was scored:

> Draw an accurate graph of your favorite nonlinear function $y = f(x)$ (no formula for $f(x)$ needed) and pick a point on the x-axis and label it "a". (Make the graph fairly large so you can clearly draw other things on it.) Recall that the derivative of a function f at a point $x = a$ is defined by
>
> $$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$
>
> (a) Illustrate and label on your graph each of the following quantities that appear in the definition. Then write a short statement explaining in terms of your graph what each quantity means (1-2 sentences for each quantity).
>
> (i) $f(a)$
> (ii) h
> (iii) $f(a+h)$
> (iv) $\frac{f(a+h)-f(a)}{h}$
> (v) $f'(a)$
>
> (b) Explain in terms of the graph what the equation $f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$ means (be sure to talk about secant lines and the tangent line).
>
> *(c) Explain why the limit is needed to find the slope of the tangent line. (Why can't we just use arithmetic and algebra?)*

The core idea for this problem was: average rate of change (AROC) can only be calculated with two distinct points on the curve, so it cannot be used to find the slope of a tangent line, which (locally) intersects the function at a single point. An *accurate* explanation should communicate that AROC cannot be used to calculate the slope of a tangent line. The *mathematical language* should describe AROC, expressing either that two points are needed to find AROC, or that using a single point to compute AROC would result in 0/0. A *coherent* explanation should express why the limit can be used to find the slope of the tangent line (e.g., slopes of secant lines can be used to approximate the tangent slope with arbitrary accuracy). No students used diagrams to support their explanations, so diagrams were scored as NA for all students. The average student responses from each section are given in Figures 6.1, 6.2, and 6.3.

Adam (see Figure 6.1) received a 1 for *accuracy*. He correctly stated that "you cannot find the slope at one point using algebra." However, he also made the incorrect statement "the limit tells you where x approaches but algebra tells you what x is," so he received a 1 rather than 2. Adam received a 1 for *mathematical language*, because he described that the slope could not be found with algebra, but did not describe how AROC could be found. Adam's explanation introduced a lot of irrelevant ideas, and did not did not connect approximation to tangent slope, so he received a 0 for *clarity*.

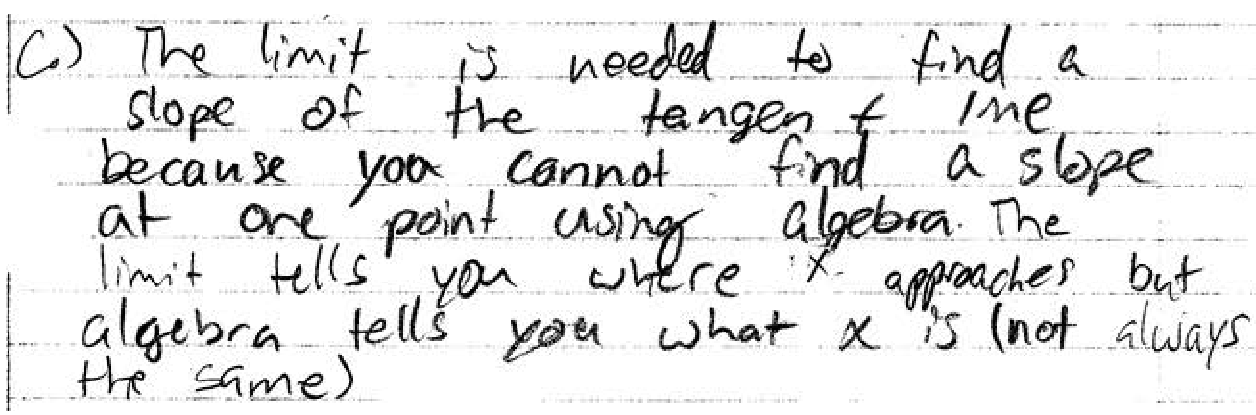
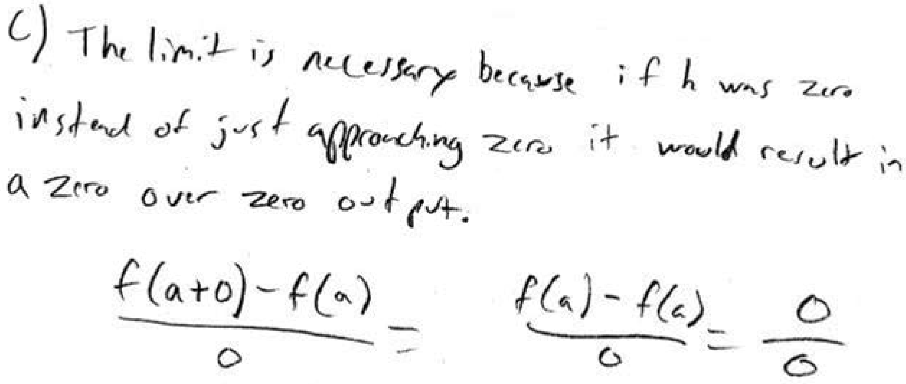

c.) The limit is needed to find a slope of the tangent line because you cannot find a slope at one point using algebra. The limit tells you where x approaches but algebra tells you what x is (not always the same)

Figure 6.1: Adam's Solution to the Early Semester Problem (Comparison Section).

Samantha's response (see Figure 6.2) received a 2 for *accuracy* and *mathematical language* because she described that "if h was zero . . . it would result in a zero over zero output," and she demonstrated this symbolically as well. This calculation and description illustrated that trying to calculate AROC using a single point (i.e., $h = 0$) would result in 0/0. Samantha received a 0 for *clarity*, because her explanation was relatively unclear and she did not introduce the idea of approximation or how to find a slope.

Brian's response (see Figure 6.3) received a 2 for *accuracy* and *mathematical language* because he stated "just trying to evaluate the tangent line at point a would result in a 0/0 slope." Brian received a 1 for *clarity*, because it was relatively clear, even though it did not explain the role of the limit. Scores are summarized in Table 6.2.

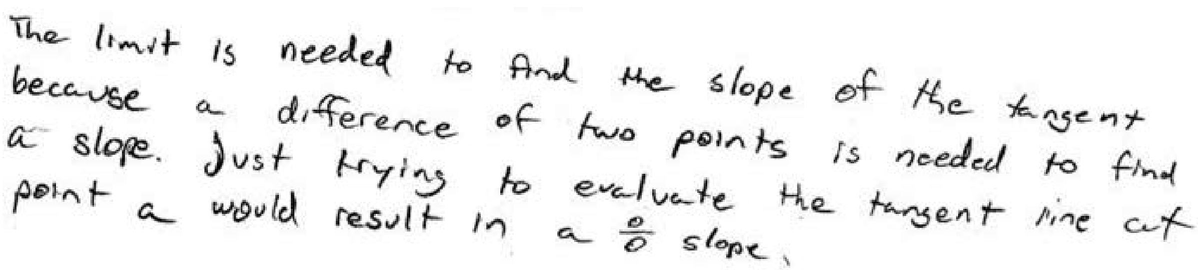
Early in the semester, none of the average student explanations explained why the limit was needed; they scored low on clarity as a result. Ambiguous and imprecise communication, like Samantha's "it would result" or Brian's "evaluation of the tangent line" (when he was referring to calculating the slope) also reduced the clarity of the explanations. The com-



c) The limit is necessary because if h was zero instead of just approaching zero it would result in a zero over zero output.

$$\frac{f(a+0)-f(a)}{0} = \frac{f(a)-f(a)}{0} = \frac{0}{0}$$

Figure 6.2: Samantha's Solution to the Early Semester Problem (Phase I Experimental Section).



The limit is needed to find the slope of the tangent because a difference of two points is needed to find a slope. Just trying to evaluate the tangent line at point a would result in a $\frac{0}{0}$ slope.

Figure 6.3: Brian's Solution to the Early Semester Problem (Phase II Experimental Section).

parison explanation (Adam) was unclear and added a lot of irrelevant information. These explanations were typical for early-semester explanations, which generally scored low.

	Accuracy	Mathematical Language	Clarity	Diagrams
Comparison (Adam)	1	1	0	-
Phase I (Samantha)	2	2	0	-
Phase II (Brian)	2	2	1	-

Table 6.2: Scores of the Average Student Solutions to the Early Semester Problem.

6.3.2 Mid-Semester

The mid-semester problem was assigned to students as they began to study integration, before a formal treatment of Riemann sums. Students were asked to trace their hand and use simple shapes approximate the area enclosed in the shape they created. Finally, students were asked to estimate the error of their approximation and describe how to reduce the error. The task is given in its entirety, but only parts (d) and (f) were scored:

> In this problem, you will trace the shape of your hand and approximate the area of the picture that you create. Your main tasks are to devise a method for approximating the area and to show that your approximation is very close to the actual area.
>
> (a) Put your hand flat on the grid provided (with fingers touching, no gaps) and trace the shape of the outline of your hand. Make sure that the shape you trace is a function (if not, erase the parts of the shape that would make it not a function).
> (b) Devise a method to approximate the area of the region inside the curve you have traced. Explain your method in detail, and explain why it should work. (Don't perform any calculations yet.)
> (c) Use the method you described above to approximate the area of the outline of your hand. (Show your work.)
> *(d) Describe a method for estimating the error in your method of approximation. (Error is something you would like to make small! Thus an estimate for the error means being able to say the error is less than some value.)*
> (e) Calculate an estimate the error for your method.
> *(f) Explain (in principle) how you could improve your method to make your estimate as accurate as one could want (i.e., minimize the error). (You do not actually have to perform the calculations, just explain what you would do.)*

The core idea for part (d) was: an error estimate must be an overestimate (as requested by the problem statement), not an estimate of the actual error. An *accurate* explanation should provide a method for overestimating error. A student's *mathematical language* should unambiguously describe how error is actually calculated. A *coherent* explanation should clearly explain why the estimate provided is an overestimate. *Diagrams* were only scored when the error calculation was made explicit in the diagram. When necessary, work from part (e) was used to determine how the estimate was calculated.

The core idea of part (f) was: smaller shapes result in a better estimate. An *accurate* explanation should focus on decreasing the size of the intervals or shapes used. A student's *mathematical language* should describe the "small" shapes. Students needed to communicate that the process can go on forever (e.g., by referring to a limiting process). A *coherent* explanation should describe why smaller shapes reduce error; they result in less unaccounted for area inside or outside the original figure. The overlap between the original figure and shapes used to estimate area could be illustrated by a *diagram*. Three solutions (Figures 6.4, 6.5, and 6.6) are now provided.

Sanjay (Figure 6.4) stated: the error "can be approx by adding together the spaces in between the rectangles and the curve." These spaces are shaded on Sanjay's diagram.

Looking at part (e) of the solution (where he performs calculations to support (d)), some of the areas were added while others were subtracted; Sanjay attempted to estimate the exact error value. The core mathematical ideas was to overestimate error, not try to estimate it exactly, so Sanjay received a 0 for *accuracy*. Because his accuracy was 0, he received a 0 for the other dimensions for part (d) as well. In part (f), Sanjay stated: "I should make smaller rectangles/subintervals to have less error," but he also added "and use midpoint method instead," which would not help him reduce his error indefinitely. Because both of these ideas were stated but not elaborated, he received a 1 for *accuracy*. Sanjay received a 1 for *mathematical language* because he did not describe how to increase the accuracy indefinitely. He received a 1 for *clarity* because he did not describe why smaller rectangles would reduce error. Sanjay did not provide a diagram for part (f).

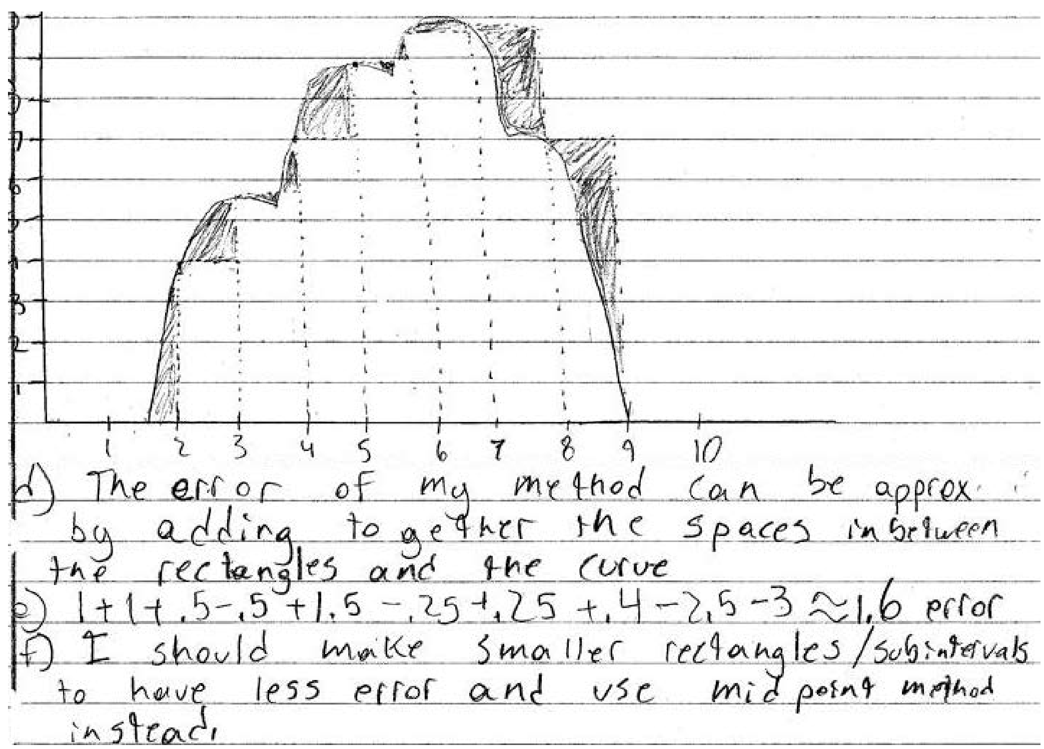


Figure 6.4: Sanjay's Solution to the Mid-Semester Problem (Comparison Section).

Tim (Figure 6.4) explained, to "count the 1x1 grid units that were in the rectangles and not under the function," and "to add the amount of space under the curve that was not in the rectangles." Since both of these sources of error are added, this provides an overestimate, so Tim received a 2 for *accuracy* for part (d). He received a 1 for *mathematical language* because his exact calculation was unclear; his diagram illustrated numbered boxes corresponding to error, but the assignment of partial boxes appears somewhat unsystematic. Tim received a 2 for *clarity* because it was clear that all sources of error were added so it should be an overestimate. Finally, Tim received a 1 for *diagrams* because the diagram supported the explanation, but did not show exactly how to find the error. In part (f), Tim said, "we would need to increase the amount of rectangles," so he received a 2 for *accuracy*. He received a 1 for *mathematical language*, because he said we need "an infinite amount of rectangles." While this demonstrates that the size of rectangles could be reduced indefinitely, it is not possible to have "infinite rectangles." Tim also received a 1 for *clarity* because he did not describe why smaller rectangles result in less error. Tim did not provide a diagram for part (f).

Joe (Figure 6.4) received a 1 for *accuracy* for part (d). Joe "outlined a section of the graph that always overestimates the error of the hand as a "Max" error tolerance." This provides an overestimate of the actual area, but Joe did not take additional steps necessary to overestimate the error (e.g., by computing an underestimate of the area and using the two estimates together). Joe received a 2 for *mathematical language*, because his diagram made it clear exactly how he calculated his overestimate. He received a 1 for *clarity* for his confusion between bounding error and the estimate itself. Finally, Joe received a 1 for *diagrams* because he showed how to calculate the overestimate, but not exactly how that would relate to error. Joe's response to part (f) received a 2 for *accuracy*, because he said "make more boxes to be more accurate." Joe also stated "maybe use midpoint as well," but since he emphasized iteratively choosing smaller intervals ($0.5 \to 0.25 \to 0.125 \to \ldots$), he still received a 2 rather than a 1. He received a 2 for *mathematical language* because he illustrated the iterative nature of choosing smaller interval sizes, indicating that the process could go on indefinitely. Joe received a 1 for *clarity* because he did not describe why smaller intervals would result in less error. Joe did not provide a diagram.

	Accuracy	Mathematical Language	Clarity	Diagrams
Comparison (Sanjay)	0	0	0	0
Phase I (Tim)	2	1	2	1
Phase II (Joe)	1	2	1	1

Table 6.3: Scores of the Average Student Solutions to the Mid-Semester Problem (Part d).

	Accuracy	Mathematical Language	Clarity	Diagrams
Comparison (Sanjay)	1	1	1	-
Phase I (Tim)	2	1	1	-
Phase II (Joe)	2	2	1	-

Table 6.4: Scores of the Average Student Solutions to the Mid-Semester Problem (Part f).

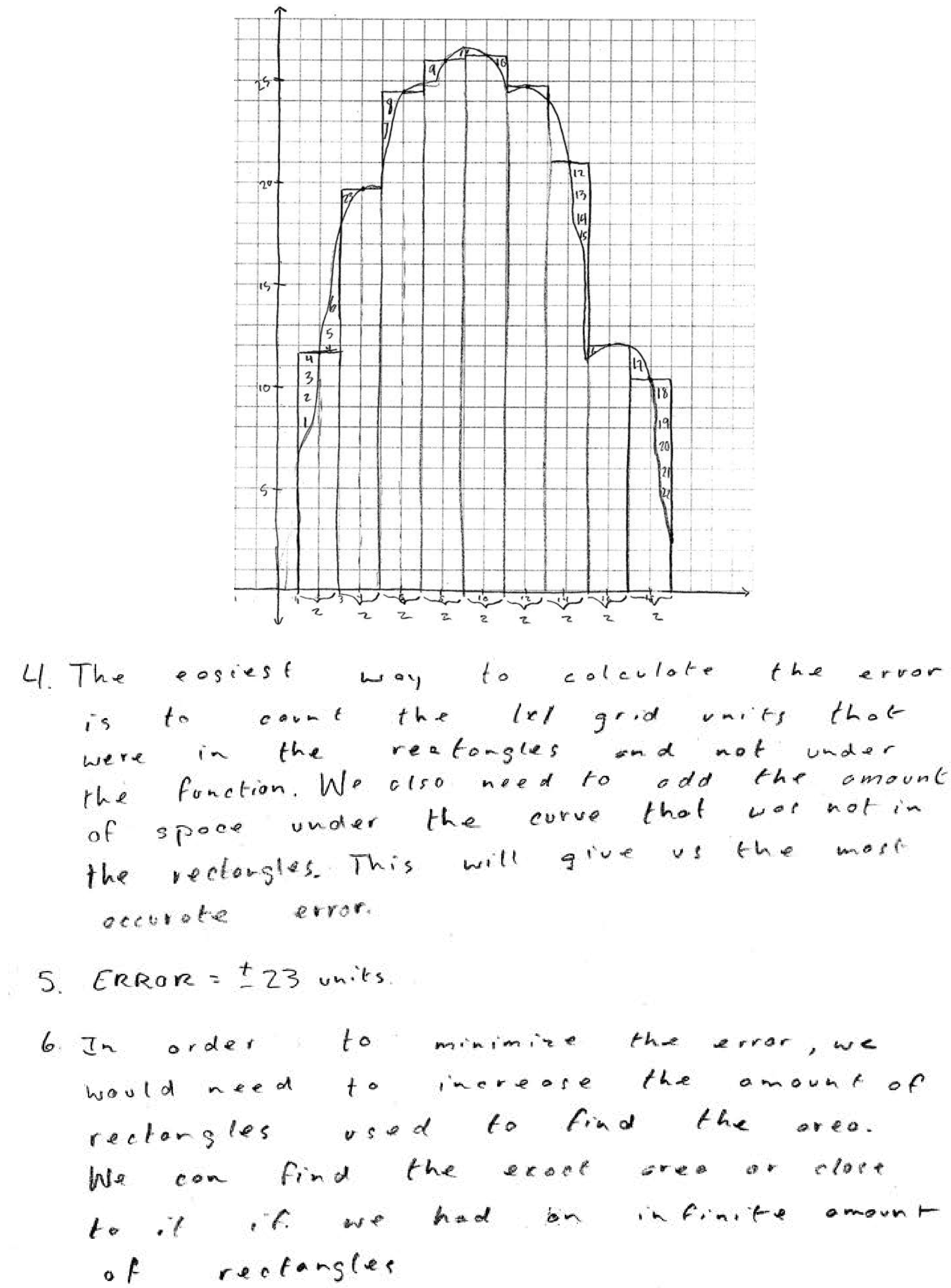


Figure 6.5: Tim's Solution to the Mid-Semester Problem (Phase I Experimental Section).

The scores for the mid-semester problem are summarized in Table 6.3 and Table 6.4. Sanjay's explanation (comparison section) missed the core mathematical ideas, and his ex-

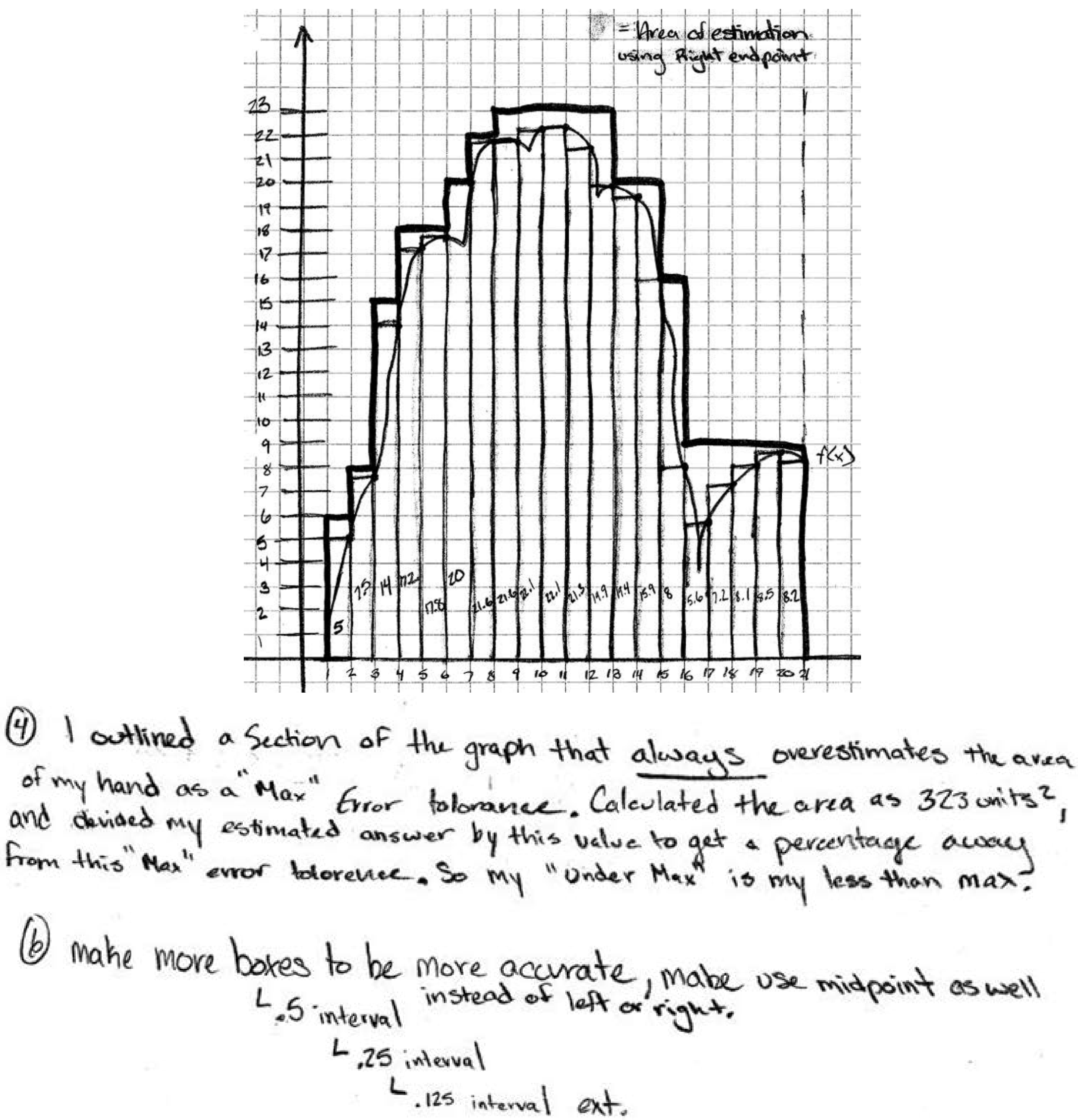
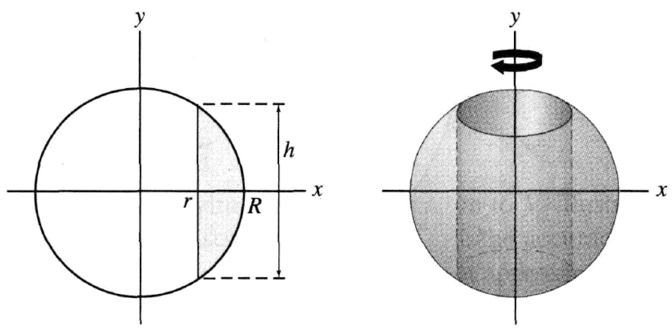


Figure 6.6: Joe's Solution to the Mid-Semester Problem (Phase II Experimental Section).

planations were relatively incomplete. In contrast, the average students in the experimental sections used multiple sentences to more clearly express their ideas. They also used diagrams to support their written explanations, in contrast to Sanjay's response that featured an incomplete diagram. The differences in the average explanations were representative of the other sampled mid-semester explanations, which showed an increasing divide between the three groups of students.

6.3.3 Late Semester

The late semester problem was the final PAR problem that students completed. This problem described the creation of a bead designed by drilling a cylindrical hole through the center of a sphere; this problem is also known as the napkin ring problem. Students could represent the bead as a volume of revolution, and rewrite their result to express the fact that volume only depends on the height, not the radius of the original sphere. Students then explained two prompts related to the volume and the surface area of the bead. The task is given in its entirety, but only parts (d) and (e) were scored:

> A bead can be formed by removing a cylinder of radius r from the center of a sphere of radius R (see the figure below).
>
> 
>
>
> A bead is a sphere with a cylinder removed.
>
> (a) Use calculus to find the volume V of the bead with $r = 1$ and $R = 2$.
> (b) Use calculus to find the volume V of a bead in terms of the variables r and R.
> (c) The bead's height h is labeled in the figure. Rewrite your formula from (b) to show that $V = \frac{\pi}{6}h^3$.
> *(d) Since your answer in part (c) expresses the volume entirely in terms of h (and not r or R), it means that all beads of the same height have the same volume. In other words, if you started with a sphere the size of an orange and a sphere the size of a basketball and made them each into beads a height of 2 inches, the beads would have the same volume. Explain how this can be true. (Hint: think about the shape of the beads)*
> *(e) Do all beads of the same height h also have the same outside surface area (not including the surface area of the cylindrical hole inside)? (Note: you do not need to use an integral to compute the surface area, just discuss it intuitively.)*

The core idea of part (d) is: as the outer radius of the sphere increases, the area of the cross-section that is rotated is smaller (i.e., the bead is thinner); this allows the volume of the beads to remain constant. An *accurate* explanation should explain the changing nature of the cross-sectional area. When a cylinder is removed from the small sphere, the cross-sectional areas of the remaining portions are relatively large (in comparison to the total area), which results in a thick bead with a small outer circumference. In contrast, when the cylinder is removed from a large sphere, the cross-sectional areas of the remaining portions are relatively small, resulting in a thin bead with a large outer circumference. *Mathematical language* relates to how the shapes of the beads are described. Ideally, students should use the terms cross-sectional area or thickness to describe aspects of the beads. A *coherent* explanation should connect the increasing circumference and decreasing cross-sectional area and why that matters. The differences in thickness and circumference cancel one another to result in two beads with the same volume. *Diagrams* can illustrate graphically the difference in cross-sections; a diagram is provided for clarity (see Figure 6.7).[3]

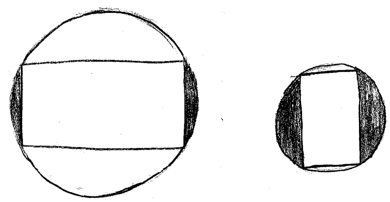


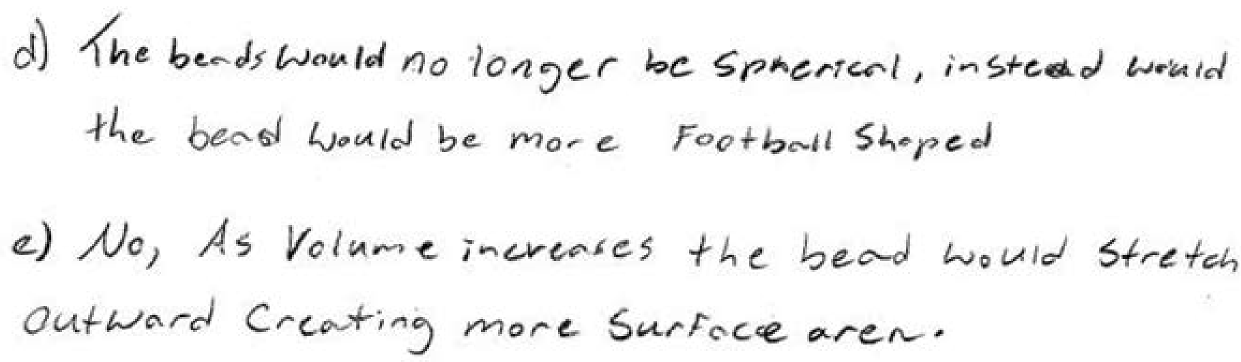
Figure 6.7: The Large Sphere Makes a Thinner Bead. The Small Sphere Makes a Thicker Bead.

The core idea for part (e) is: beads with a larger outer radius have a larger outside surface area. An *accurate* explanation should explain that the outer radius increases the surface area. The bead made from the larger sphere has a larger surface area than the bead made from the smaller sphere, because the outside circumference is the most significant factor in determining surface area (the sphere's curvature matters but not as much). In the limit as $R \to \infty$, the outside of the bead approaches a cylinder, so the surface area approaches $2\pi Rh$, which diverges to ∞ (while the smaller spheres have finite surface area). *Mathematical language* relates to how students describe the regions of the sphere that have an impact on surface area; students could use terms like "outer radius" or "outside circumference." A *coherent* explanation should make clear how the outer radius affects surface area (e.g., by including an equation). Not a single student mentioned the outside curvature of the bead, so this aspect of the solution was not included in analysis. *Diagrams* could be used to help illustrate graphically the surface areas of different beads. Student solutions are provided in Figures 6.8, 6.9, and 6.10.

Will (see Figure 6.8) stated: "the beads would no longer be spherical, instead would the bead would be more football shaped." This statement indicates a misunderstanding of the problem situation, so Will received a 0 for all dimensions on part (d) except for diagrams, which was NA. In part (e), Will stated: "as volume increases the bead would stretch outward creating more surface area." The problem statement indicates that spheres

[3]This is a diagram that Katherine (Phase II experimental section) submitted as a part of her solution.

have equal volumes, so once again Will received a 0 for all dimensions other than diagrams, which was NA.


d) The beads Would no longer be Spherical, instead would the bead Would be more Football Shaped

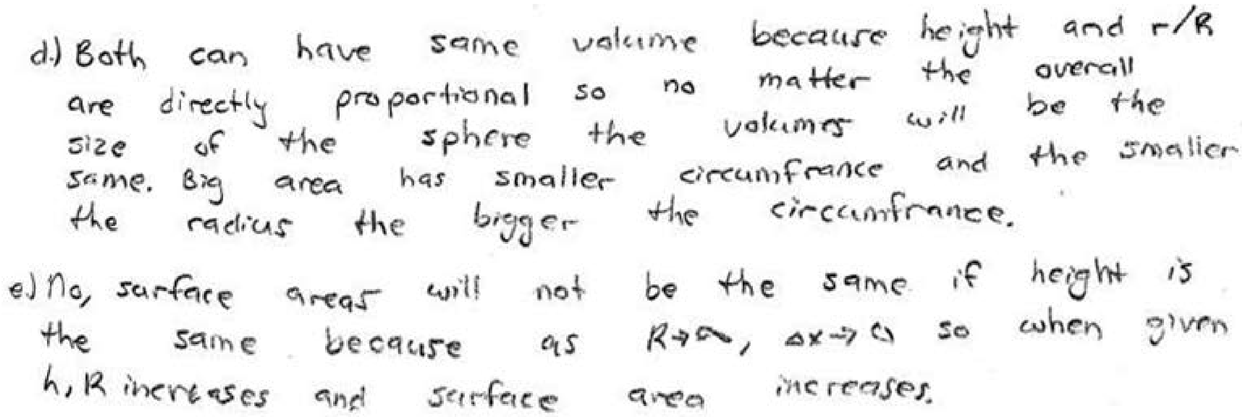
e) No, As Volume increases the bead Would Stretch outward Creating more Surface area.

Figure 6.8: Will's Solution to the Late Semester Problem (Comparison Section).

Silvia (see Figure 6.9) states: "big area has smaller circumference and the smaller the radius the bigger the circumference." Following Silvia's parallel construction it is likely that she meant "the smaller the area" but accidentally wrote radius instead of area. This would indicate a correct description of the relationship between cross-sectional area and circumference. However, she only received a 1 for *accuracy*, because she said "r/R are directly proportional." If r and R were directly proportional, the thickness of the beads would increase with size of the sphere, rather than decrease. Because directly proportional contradicts the special relationship between cross-sectional area and circumference, Silvia received a 0 for *mathematical language* and *clarity*. For part (e), Silvia stated "the surface area increase," but it's unclear if she's attributing it to a growing R, a shrinking Δx, or something else; she received a 1 for *accuracy*. Despite these difficulties, she still received a 1 for *mathematical language*, because using R rather than just saying "radius" clearly denotes the outer radius. Her explanation was unclear, so she received a 0 for *clarity*. She received an NA for *diagrams*.

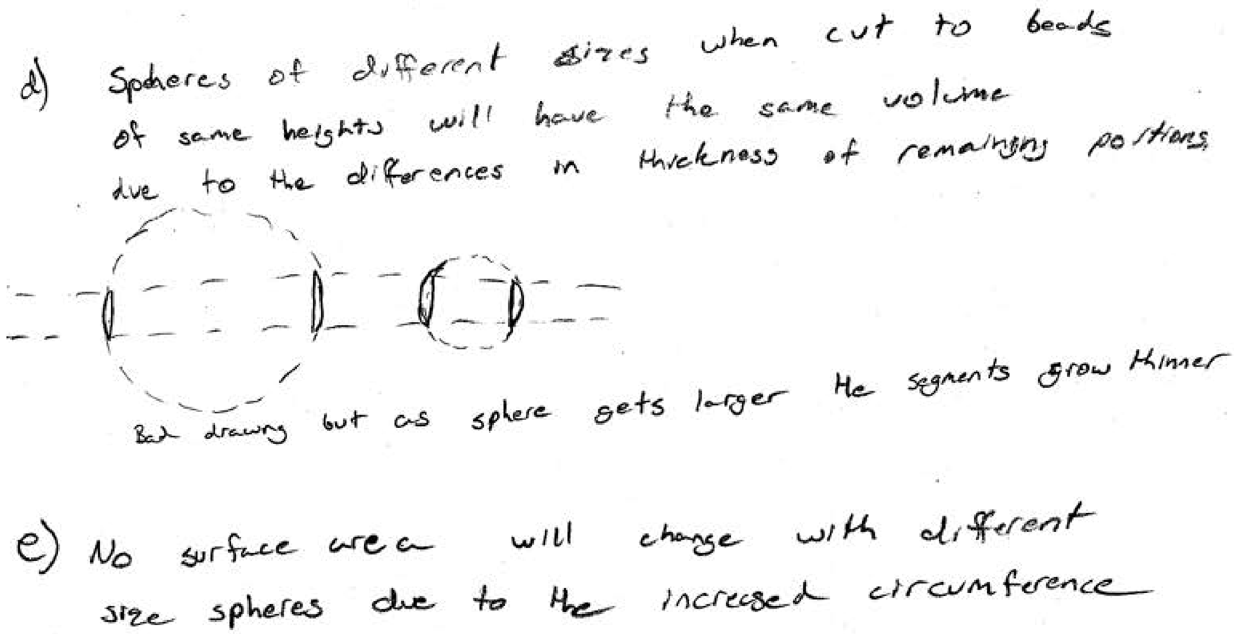
Kevin (Figure 6.10) stated: "as sphere gets larger the segments grow thinner," and illustrated the relationship between bead thickness and outer circumference in his diagram. Due to the clarity of his written explanation and diagram, he received a 2 for all dimensions on part (d). For part (f), Kevin received a 2 for *accuracy*, stating "surface area will change with different size spheres due to the increased circumference." This is also illustrated in his diagram. Kevin does not clearly refer to outer circumference so he received a 1 for *mathematical language*. He also did not state how exactly circumference relates to surface area, so he received a 1 for *clarity*. His *diagram* received a 2, because it illustrated how the surface area would increase with a larger outer circumference.

The late semester scores are summarized in Tables 6.5 and 6.6. Will's solution was representative of the other sampled solutions ($N = 9$) in the comparison section; all solutions sampled received 0 scores for all dimensions. In contrast, explanations in the experimental sections were of much higher quality.


d.) Both can have same volume because height and r/R are directly proportional so no matter the overall size of the sphere the volumes will be the same. Big area has smaller circumfrance and the smaller the radius the bigger the circumfrance.

e.) No, surface areas will not be the same if height is the same because as R→∞, Δx→0 so when given h, R increases and surface area increases.

Figure 6.9: Sivia's Solution to the Late Semester Problem (Phase I Experimental Section).


d) Spheres of different sizes when cut to beads of same heights will have the same volume due to the differences in thickness of remaining portions

e) No surface area will change with different size spheres due to the increased circumference

Figure 6.10: Kevin's Solution to the Late Semester Problem (Phase II Experimental Section).

6.4 Overall Results

Solutions were scored with a very high agreement of 94.1%. There were only 8 disagreements out of 135 explanations scored, and all disagreements were resolved after discussion. I began my analyses by computing the correlation matrix for the four dimensions that were scored. The goal of this analysis was to determine how closely related the various dimensions were to one another. I expected that the dimensions would be closely related, because students were likely to improve on all of these aspects of explanations simultaneously, rather

	Accuracy	Mathematical Language	Clarity	Diagrams
Comparison (Will)	0	0	0	-
Phase I (Silvia)	1	0	0	-
Phase II (Kevin)	2	2	2	2

Table 6.5: Scores of the Average Student Solutions to the Late-Semester Problem (Part d)

	Accuracy	Mathematical Language	Clarity	Diagrams
Comparison (Will)	0	0	0	-
Phase I (Silvia)	1	1	0	-
Phase II (Kevin)	2	1	1	2

Table 6.6: Scores of the Average Student Solutions to the Late-Semester Problem (Part e)

than independently.[4] I use pairwise deletion for correlations with the diagrams dimension, which were mostly NA. The results can be found in Table 6.7. Given that students in the comparison section nearly uniformly received all 0's on the late semester problem, I suspected the results of that section may be inflating the levels of correlation between the various dimensions. To provide a more nuanced picture, I pooled together the results from the Phase I and Phase II experimental sections into a single group, and compared this with the results of the comparison section. See Tables 6.8 and 6.9.

	Mathematical Language	Clarity	Diagrams ($n = 21$)
Accuracy	0.8123	0.7171	0.8368
Mathematical Language	-	0.6990	0.6491
Clarity	-	-	0.8438

Table 6.7: Correlation Matrix of the Four Dimensions, $N = 135$

As suspected, the four dimensions of analysis were highly correlated. In the experimental sections there was a range of correlations values; for instance, mathematical language was most correlated with accuracy ($r = 0.7829$) and less so with clarity ($r = 0.6558$) and diagrams ($r = 0.4580$). In contrast, all dimensions were correlated above $r = 0.7379$ in the comparison section. This indicates that the development of skills was less differentiated in the comparison

[4]For comparison, consider the mechanics of students' writing. Grammar, punctuation, spelling, etc., are all independent dimensions, but one would expect for them to be highly correlated.

	Mathematical Language	Clarity	Diagrams ($n = 15$)
Accuracy	0.7829	0.6894	0.7082
Mathematical Language	-	0.6558	0.4580
Clarity	-	-	0.8016

Table 6.8: Correlation Matrix of the Four Dimensions (Experimental, $N = 90$)

	Mathematical Language	Clarity	Diagrams ($n = 6$)
Accuracy	0.8417	0.7379	1
Mathematical Language	-	0.8139	1
Clarity	-	-	1

Table 6.9: Correlation Matrix of the Four Dimensions (Comparison, $N = 45$)

section; students either did well overall, or not so well. In the comparison section, there were only six instances in which diagrams were used. In these cases, the use of diagrams was perfectly correlated with all other dimensions. In five of six cases, this was because the students scored zero on all four dimensions. Because there were so few students who used diagrams, scoring would need to be conducted again with a larger sample size for a more meaningful correlation to be found. The use of student diagrams is summarized by Figure 6.11. The early semester problem was not included because no students sampled used diagrams in their explanations.

The scores in Figure 6.11 show that only the Phase II experimental section was likely to correctly use diagrams to support their explanations (on problems later in the semester). Students used diagrams even though it was not an explicit part of instruction. Evidently, students spontaneously came to value the use of diagrams; this idea may have been spread throughout the class as a result of peer-conferences, but I do not have sufficient data to make this claim.

Next I analyzed student scores by computing the aggregate scores for each section. Each dimension (accuracy, mathematical language, clarity, and dimensions) was scored out two points, so it was possible to receive a maximum of eight points for each explanation. Table 6.10 shows the average scores across sections. The first column of scores provides the average score with all four dimensions included. Numerically, the Phase II section scored more than 4.5 times higher than the comparison section, and more than 1.5 times higher than the Phase I section. The second column of scores provides the average for diagrams removed. This score is also provided because students were not expected or encouraged to include diagrams in any of the sections. The differences were still numerically large, but not as much so (Phase II / Comparison = 4.1, Phase II / Phase I = 1.5). The final row shows the average

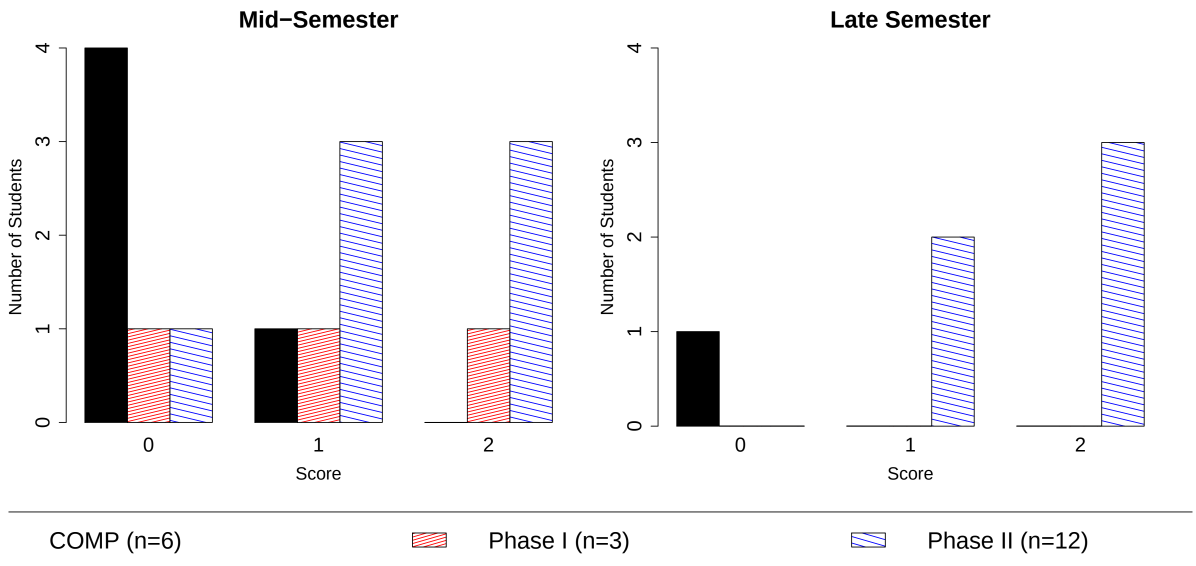


Figure 6.11: Frequency of Diagram Scores Received.

scores for the aggregate of the two experimental sections.

I tested for significant differences across sections in Tables 6.11 and 6.12. The first test was between the comparison section and the aggregate of the experimental sections, to see if the experimental design had a significant impact. The second test was between the two experimental sections (Phase I and Phase II), to see if the changes in the design had a significant impact. Both comparisons were significant when diagram scores were included in the analyses (Table 6.11). However, when diagrams were removed (Table 6.12) the differences between Phase I and Phase II was only marginally statistically significant.

	Total Average (8 Possible)	Diagram-free Average (6 Possible)
Comparison	0.57	0.56
Phase I	1.60	1.53
Phase II	2.66	2.29
Experimental	**2.13**	**1.91**

Table 6.10: Average Explanation Scores By Section ($N = 45$ Per Section)

The final analysis of aggregate scores consisted of looking at average scores by dimension. For these averages all NA's for diagrams were converted to 0's. As a result, the scores for diagrams represent a combination of the frequency with which diagrams were used and the quality of the diagrams used. See Table 6.13. As with Table 6.10, the table exhibits full

Average Explanation Scores	Independent samples $t-$test	Sign.
Comparison vs Experimental	$t = 4.6813, df = 128.767, p = 7.12 \cdot 10^{-6}$	**
Phase I vs Phase II	$t = 2.1424, df = 81.963, p = 0.0351$	*

Table 6.11: Statistical Comparisons of Average Section Scores

Diagram-free Average Scores	Independent samples $t-$test	Sign.
Comparison vs Experimental	$t = 4.3828, df = 118.335, p = 2.55 \cdot 10^{-5}$	**
Phase I vs Phase II	$t = 1.7708, df = 85.468, p = 0.0802$	

Table 6.12: Statistical Comparisons of *Diagram-Free* Average Section Scores

monotonicity: Phase I is higher than the comparison section on all dimensions, and Phase II is higher than Phase I on all dimensions. Moreover, there were considerable numerical differences between the sections. For instance, the Phase II section scored more than five times as Phase I's section for diagrams, and more than 17 times as high as the comparison section.

	Accuracy	Mathematical Language	Clarity	Diagrams
Comparison	0.2889	0.1778	0.0889	0.0222
Phase I	0.7333	0.4444	0.3555	0.0667
Phase II	1.0000	0.7333	0.5556	0.3778
Experimental	**0.8667**	**0.5889**	**0.4556**	**0.2222**

Table 6.13: PAR Scores by Dimension (Out of 2 Possible)

The rest of the analyses that follow focus on change over time. To create a single score for each time period, I averaged the scores for the two prompts on the mid-semester problem and the two prompts on the late semester problem. Table 6.14 shows that students in the experimental sections performed numerically better than the comparison section, and these differences grew over time.

Students in all sections scored lower on the late semester problem than the mid-semester problem; this reflects the difference in difficulty level of these two problems. To account for problem difficulty, I decided to compare scores across sections rather than look at raw scores. To compare scores I divided all scores by the Phase II experimental scores (i.e., each score was divided by the Phase II score corresponding to that same problem). These scaled scores

	Early Semester	Mid-Semester	Late Semester
Comparison	1.11	0.89	0.00
Phase I	1.66	2.61	0.55
Phase II	1.66	3.33	2.50
Experimental	**1.66**	**2.97**	**1.53**

Table 6.14: Average Explanation Scores by Problem (Out of 8 Possible)

represent the percentage of the Phase II experimental score that each section achieved (see Table 6.15).

	Early Semester	Mid-Semester	Late Semester
Comparison	0.67	0.27	0.00
Phase I	1	0.78	0.22
Phase II	1	1	1

Table 6.15: Average Scaled Explanation Scores by Problem (Raw Score / Phase II Experimental Score)

Table 6.15 shows considerable numerical improvement for the experimental sections. As the semester progressed, the ratio of comparison scores to the Phase II experimental section continued to shrink. Simultaneously, the numerical distance between the two experimental sections also grew, suggesting that the revised intervention had a greater impact on learning in the Phase II experimental section. These scores are consistent with the patterns of growth illustrated by the average explanations in the previous section.

Table 6.16 shows statistical comparisons of growth over time between the experimental and comparison sections, and Phase I and Phase II experimental sections. The results were computed by examining the change in scores for each section from the early semester to late semester problems. To do so, I subtracted the average early semester score for each section from the late semester scores. These results show that the experimental sections learned significantly more than the comparison section during the semester, and that the Phase II section learned significantly more than the Phase I section.

Changes along individual dimensions are given in Figure 6.12. At the beginning of the semester, student performance was relatively similar across sections. Over time, students in the experimental sections began to perform numerically better across all dimensions. As with the aggregate scores, these numerical differences grew over time, with the Phase II experimental section performing even better than the Phase I experimental section. In

Diagram-free Average Scores	Independent samples $t-$test	Sign.
Comparison vs Experimental	$t = 2.624, df = 35, p = 0.01279$	*
Phase I vs Phase II	$t = 2.8856, df = 22.696, p = 0.008416$	**

Table 6.16: Statistical Comparisons of Average *Change* From Early to Late Semester

general, students scored numerically higher on accuracy than on mathematical language and clarity. This highlights the difficulty of communicating clearly.

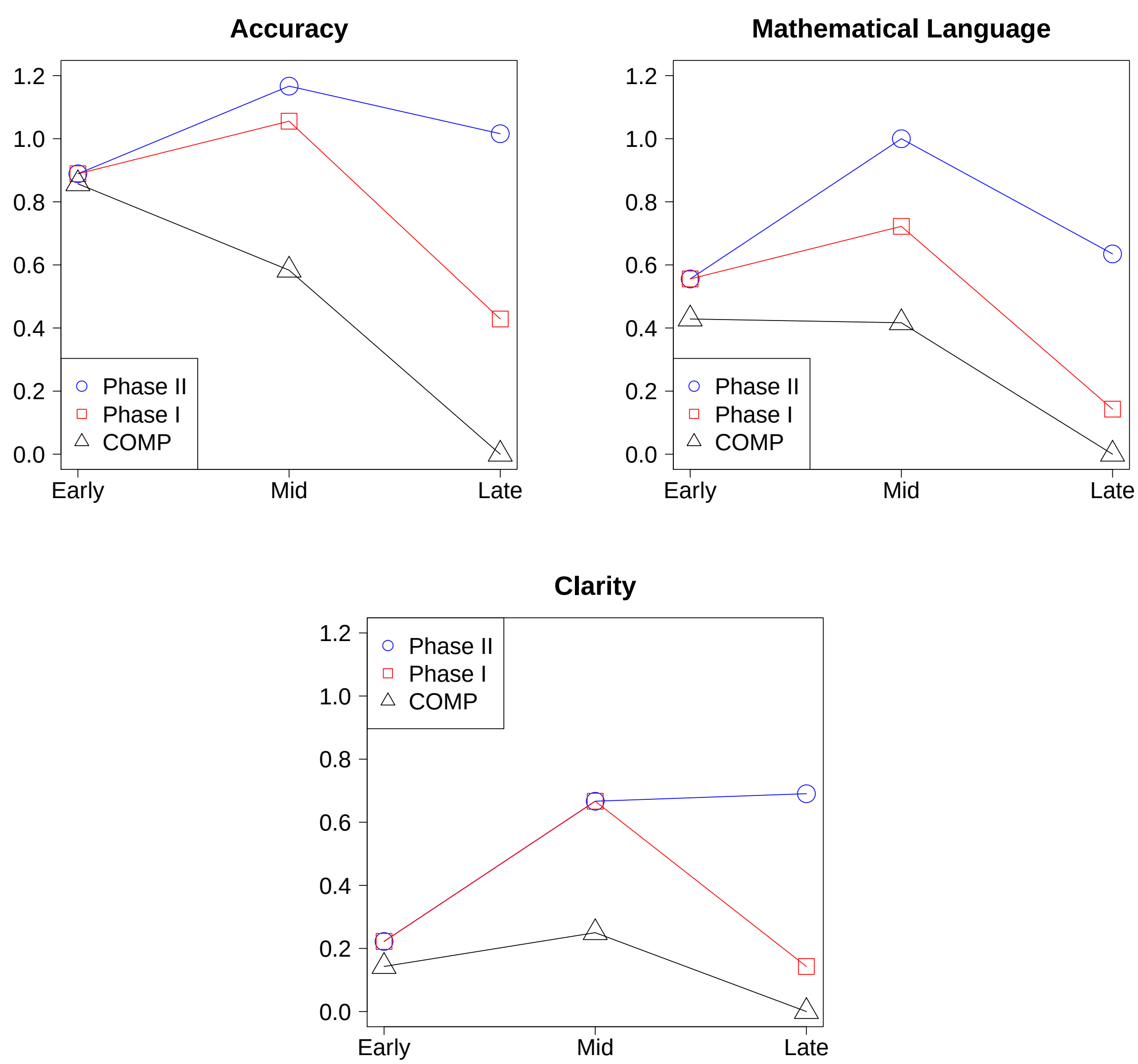


Figure 6.12: Average Explanation Scores by Dimension.

6.5 Discussion

The analysis of student explanations further supported the positive impact of Peer-Assisted Reflection (PAR). At the beginning of the semester, student explanations were relatively similar across sections. As the semester progressed, students in the experimental sections improved the accuracy and clarity of their explanations considerably, while the students in the comparison section did not. Overall, students in the Phase I experimental section scored over twice as many points as the students in the comparison section, while the Phase II experimental students scored over four times as many points as students in the comparison section; these scores highlight the impact of the intervention, as well as the considerable design improvements that were made between Phase I and Phase II. These improvements were the result of careful study and analysis of student engagements during Phase I. Through PAR, students engaged in peer-analysis to learned to self-reflect on the quality of their own explanations. Systematic training and regular practice analyzing a variety of examples supported students to construct their own high-quality explanations.

Students in all sections scored relatively low at the beginning of the semester, especially in terms of mathematical language and clarity. Even though the students were enrolled in a college level mathematics course, it was difficult for them to clearly express their mathematical ideas. Unfortunately, the comparison section of students who received standard instruction developed relatively little in terms of explanation skills. Not a single one of these students received any points on the late semester problem; their explanations were largely incoherent. This suggests that students may make little progress over the course of a semester in learning to explain mathematics unless it is made a specific focus of the teaching and learning environment.

In contrast, students in the experimental sections made considerable improvements in their explanations. Not only did they increase the accuracy of their explanations, but their clarity and use of language was also much greater than the baseline of the comparison section. Growth was most dramatic in the Phase II experimental section, due to iterative design improvements.

Accuracy scores were consistently higher than mathematical language and clarity in all sections; a certain level of accurate understanding seems prerequisite to effective communication. This highlights the challenge of separating communication from understanding (cf., Sfard, 2007).

Only students in the Phase II experimental section effectively used diagrams to support their written explanations. Curiously, the use of diagrams emerged mid-semester, and was not an explicit focus of instruction. It may be that the use of diagrams emerged spontaneously as a means of improving communication, and that it spread as a result of peer interactions; without more data it is not possible to tell. Regardless, it is clear that explaining mathematics is difficult for students, and that students are unlikely to learn how to explain effectively without considerable support. Peer-Assisted Reflection (PAR) appears to provide such support.

7 Mechanisms for Learning

Previous chapters provided evidence of learning gains for students engaging in PAR. I now delve into the specific mechanisms of PAR that seemed to support learning. PAR required students to work in iterative cycles: students made a preliminary attempt at a problem, receive feedback and thought about the problem more deeply, revised, and turned in their final solution. Within these iterative cycles, students encountered new ideas to support their learning: by spending extra time with a problem, by discussing with peers, by explaining and hearing explanations, and by seeing the work of others. These four sources can be consolidated into the acronym IDEA, meaning **Iteration, Discussion, Explanation, Alternatives**. In contrast to other activities, PAR seemingly allowed students to better close the feedback cycle; in more general learning contexts, students often receive feedback that is never used, limited in usefulness.

The data I draw upon in this section are primarily for illustrative purposes. Given the nature of my data collection, it is not possible to make claims that certain mechanisms had greater impacts than others, or that these mechanisms impacted all students. However, the evidence I collected indicates that at least some students benefited in these ways, and provides a tentative theoretical framework for understanding the impact of PAR.

7.1 Iteration

The iterative cycles of PAR seemed to change students' perceptions of problem solving. Rather than viewing homework as something that is attempted once, turned in, and forgotten about, PAR forced students to revisit their work. As a result, students seemed to increasingly view their first draft of the problem as a work in progress, and didn't expect it to be correct. As Andy said:

> With the PAR problems I honestly think we're not supposed to get it correct the first time around.

Other students also made note of this. As Tony noted one day during office hours:

> I like the PAR. It's like we get to come to class and be wrong, and that's okay. Then later we get to revise our work and be right.

Mike made a similar remark:

> PAR is good. I like how we can put our initial solution down, and even if it's wrong it doesn't really matter, because we can just talk about it with a group member the next day, and figure it out together. And generally you don't get stuck on a wrong solution, you figure it out.

These quotes from students (and others not shown) indicate that PAR helped some students view homework in a different way. Students worked on a problem, received new information from a variety of sources, and tried again. This activity structure seemed to increase students' persistence. Rather than giving up when they couldn't solve a problem on their own, students realized that getting input from peers, the instructor, or other resources was often sufficient to help them solve challenging problems. Persistence in the face of failure is an extremely important skill that students seemingly developed through PAR.

PAR also influenced the way that students perceived mathematics and their role as students in the classroom. In the interviews, students noted that both explanation and justification were skills that were emphasized in the course and that they learned to develop. Andy captured this sentiment well in his interview:

> What this class has really showed me, now that I think about it, is the importance of justification. Because Dan talks about it all of the time. So we're always looking to justify stuff. So if I look back on the written homework assignments I did well on, you can see this evolution of here's my answer, *and* here's why I think it's right.

As students developed these skills, they learned to analyze the work of others and make assessments as to whether or not they thought it was correct. This positioned students as authorities and liberated them from complete reliance on the instructor's judgments of correctness.

PAR provided students with crucial practice in how to collaborate in an honest and collegial way. In the interviews, students emphasized that they could be honest with their partners, and that as long as they communicated in an appropriate, tactful way, it was not problematic. As Maria notes:

> I feel like all of the students are pretty open about their work and are not upset if I say something's wrong. Normally if I see something I don't understand or don't think is right I point it out and ask them about it. I'll point it out and ask them how do you know it's right?

The conclusion that students learned to collaborate better through this process garners evidence in the types of group work that took place in the class and the study groups that students formed outside of class to solve problems together.

By the end of the semester, more students wrote solutions containing complete, coherent sentences describing mathematical ideas compared to students who did not receive PAR training. Students learned how to better express their ideas both verbally and in writing, and they seemingly began to realize the importance of communication. Sometimes time alone, without even working with other students, was enough to support students to revise their work. As Chris notes:

> The PAR stuff works out the second time I go over it; then I usually understand it. Like the first time, like today, just before I came here, I was looking over the PAR that we have due on Tuesday, and I found that since I've already been doing stuff for awhile and I didn't get much sleep last night and stuff, I'll have to come back to it this weekend. If I try to do it the same way as I've always done math before, just do it and knock it out, it's not gonna be good enough. After the feedback part, I usually understand it, but not because of the feedback, which I found interesting. I have to go over it, have some time to not think about it, and then think about it again. When I think about that now it makes sense, in terms of trying to memorize stuff for history class or something. I can go over it and keep memorizing for 2 hours, and I'll forget it again. If instead, I go over it for 20 minutes, leave it for half a day, go over it for 20 minutes, and go over it 20 minutes the next day, I've only spent 1 hour on it, but I've actually memorized it instead of just seeing it over and over again at the same time.

Chris' quote illustrates that sometimes students were able to conquer problems because they were given additional time to think about them. In traditional homework cycles students don't have the same period after working on the problems (interviews revealed that students used to work on it, finish up, and turn it in a single period). PAR prevented students from doing this. During these time periods, students were exposed to new ideas, as highlighted by Kevin:

> If we are learn something in the class that day which applies to the PAR I'll go back and put that in. Or if another problem on the written homework ties into the same ideas I'll go back and revise the PAR.

Iteration supported students considerably. On PAR days a large number of students would come to office hours after class (typically 5-10) to work on the PAR problems. Students worked collaboratively to solve the problems, only asking for help as needed. Because students had already worked on the problems, each student had ideas to contribute to group discussions. As students worked through these iterative cycles, they came to recognize that even though they couldn't solve it on their first try, they could solve it eventually.

7.2 Discussing the Problem Together

As students noted in the interviews, often both partners in a peer-conference didn't know how to solve some part of the problem, and they were forced to figure it out together. Even given a short amount of time, this was sometimes effective given that students had already done a lot of thinking about the problem previously. Consider Lance and Peter working together on PAR10. Both students were struggling with part (d), which required them to come up with an estimate for the error in their method. Initial solutions are given (see figures 7.1 and 7.2).

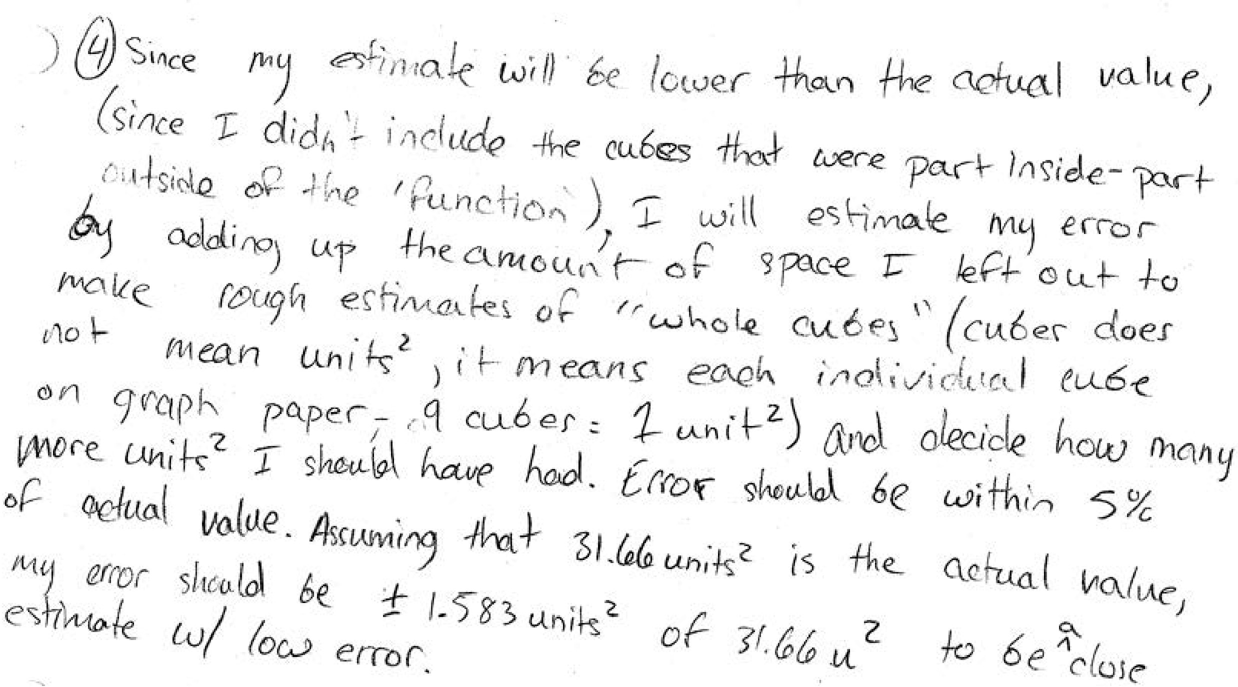

4) Since my estimate will be lower than the actual value, (since I didn't include the cubes that were part inside-part outside of the 'function'), I will estimate my error by adding up the amount of space I left out to make rough estimates of "whole cubes" (cubes does not mean units², it means each individual cube on graph paper - 9 cubes = 1 unit²) and decide how many more units² I should have had. Error should be within 5% of actual value. Assuming that 31.66 units² is the actual value, my error should be ± 1.583 units² of 31.66 u² to be a close estimate w/ low error.

Figure 7.1: Peter's Initial Solution to PAR10, Part d.

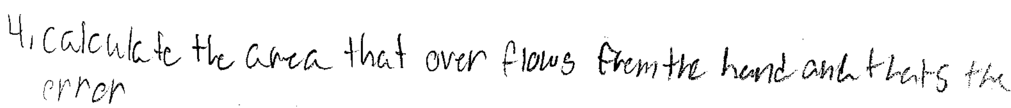

4. calculate the area that over flows from the hand and thats the error

Figure 7.2: Lance's Initial Solution to PAR10, Part d.

Both of these initial solutions lacked an adequate method for estimating error. Peter's method was focused on estimating the error from the inside, while Lance focused on estimating from the outside. These different opinions meshed together nicely in the PAR conversation:

Peter: One thing, I just wasn't sure how to do the error either, really.

Lance: The error?
Peter: Yeah
...
Lance: Were you trying to... get an under?
Peter: Yeah, initially.
Lance: What I was thinking was you could make an over-approximation. Take this right here and create an over-approximation and then subtract what you got from here with your over-approximation and it should get you this space that you were... that you didn't have filled in initially.
Peter: So it's showing it has to be between those two values. That's the error.
Lance: Right, that actual value is going to be between your low approximation and your high approximation.
Peter: Yeah, that makes sense.
Lance: And by subtracting them you'll find the error between the approximations.
Peter: Okay, cool.
Lance: Or at least I think that's what it's supposed to be.
Peter: Yeah
Lance: Maybe
Peter: I read this thing while I was doing yours and it... that made it seem like... it says you want to think about bounding your error with some larger value. So that would make sense then.
Lance: Right. Overestimate.
Peter: Alright.

The final solutions show how the students revised their solutions accordingly. Lance's solution is given in figure 7.3.[1]

2. I will draw blocks inside my hand and get the area of these blocks and add them together for my approximation.
4. You could find the area of the hand using an overestimate. Then subtract the estimate you just got from your first estimate. This will show you the difference between the two which will give you the error.

While the idea is not fully expressed, Lance draws a diagram that shows all of the blocks inside the hand. Then, he talks about getting an over approximation, and that the difference between the two would be the error. This method if executed properly would indeed calculate an overestimate of the error. Peter also improved his solution (see figure 7.4).

In his solution Peter fully explains the idea of overestimate and has an accompanying diagram that makes his work clear. While this is only one example, it shows that by working together students were able to make new insights into the PAR problems. A number of

[1]Lance's solution is typed due to the poor image quality of the scan and difficulty of reading the handwriting.

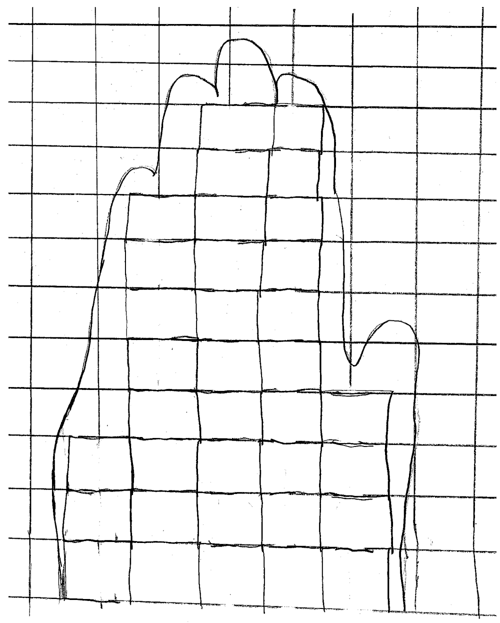


Figure 7.3: Lance's final solution to PAR10.

students noted this in the interviews as well. For instance, Mike notes that having the opportunity to discuss and get new ideas is very helpful:

> It's also just talking to the person you're working with. Every time, unless I've gotten it spot on, I've always added something from talking to the peer. Generally, most students are helpful. It doesn't really matter on intelligence level. If you talk it out with anyone you can figure it out.

Andy gave a similar response:

> I do spend time before class working on these things, but it always seems like I get something wrong. I feel like the hard part is the justification. It would be almost impossible if I had to do it myself. It's when I get to talk to other people, and when I get to bounce ideas off of them and hear their ideas that I can see why it's right or wrong, and how to justify that.

7.3 Explanation

Explanation is a powerful tool for promoting learning. In this context explanation works in two ways - students explain their ideas and are exposed to the explanations of others. Sometimes when students try to explain their own ideas, new insights are revealed or inconsistencies are uncovered. Explanations also allow students to provide one another with individualized feedback. Consider Revati and Zane's discussion of PAR5:

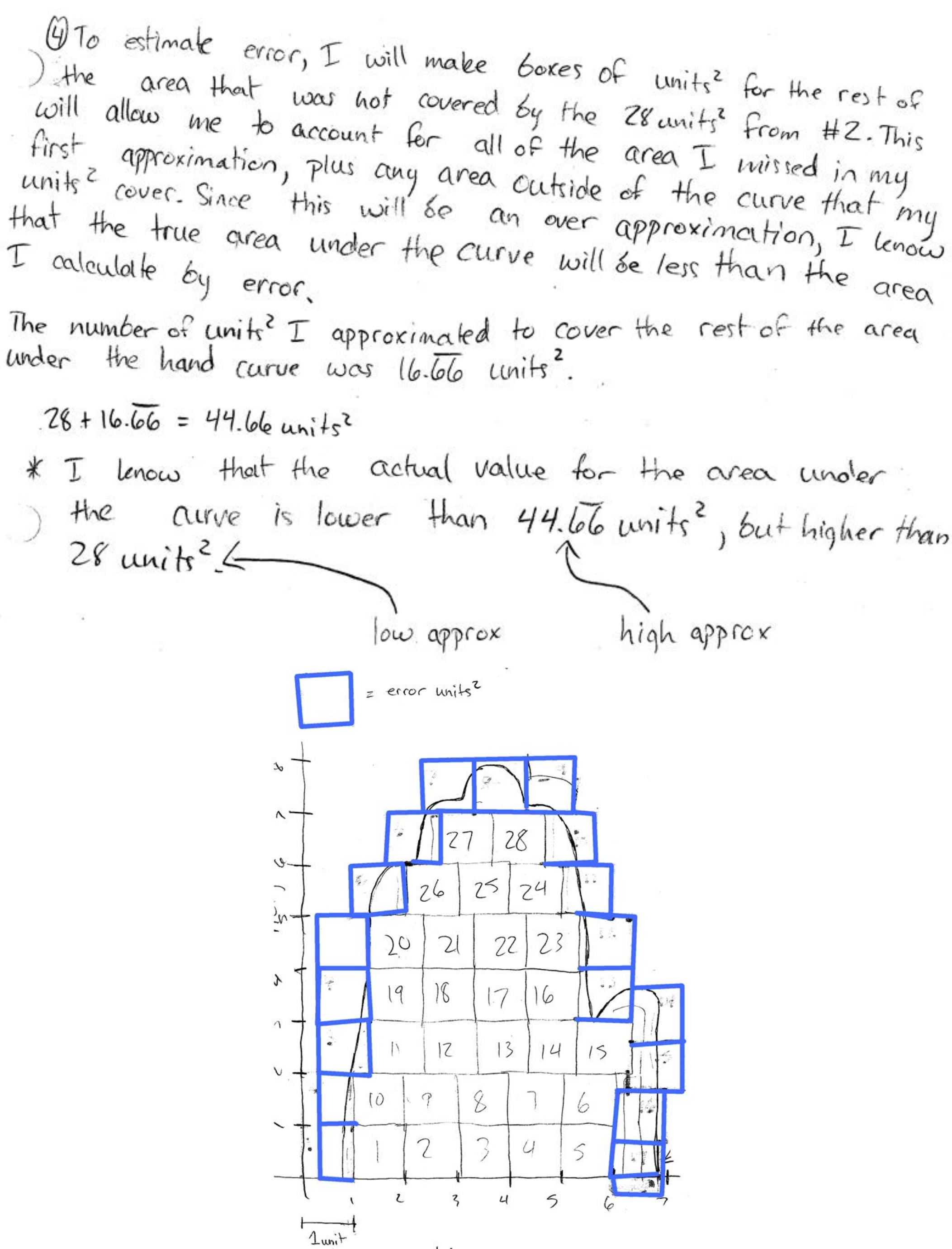


Figure 7.4: Peter's final solution to PAR10.

Revati: So explain to me what you meant in 'b' and 'c'. Cause I feel like you have the right idea, you just didn't explain it as well as you could.

Zane: So the equation. . . isn't that equation used to find the slope.

Revati: Wait, what? Say that again.
Zane: This equation? It's used to find the slope. So,
Revati: So explain, say it out loud, explain to me what this, in terms of your graph, what that means. And if you don't know that's totally fine. I can help you. That's what we're here for.
Zane: I dunno
Revati: That's fine. So what this means... you understand what this is, right? That's $f'(a)$, which is the derivative.
Zane: Yeah
Revati: This is the really long way to find the derivative, okay?
Zane: Okay

In practicing their explanations, students learned a lot about communication. As Maria said, PAR helps her learn:

> how to make it easier to read from another person's perspective. It's one thing if I think it looks good, but other people look at it and say it doesn't make sense to me. So it helps me figure out how to communicate better. It helps me to explain things in a way that is readable to others and not just myself.

As Bob said,

> I like the PAR. I'd take it over the other homework. You do about the same amount of work, and I think I actually learn more from it. You do have to explain what you did, rather than just say here, I got this magical number. You actually understand the process and I think that helps more in learning than just getting the magical number. I noticed I've been getting better grades on PAR than written homework, so I'm trying to do the written homework like I do the PAR, and hopefully that will help my grade go up. In my physics class we have these problem sets. I've actually started to use the PAR thing on those. Not the checklist, but the general method.

7.4 Exposure to Alternatives

On many occasions students recognized errors in their own solutions simply by looking at one another's work. Consider the following excerpt from PAR7. As soon as the students finished silently reading one another's work, Revati exclaimed out that she saw her errors:

Revati: I know I did it all wrong. I was reading yours and was like, 'Oh my goodness. How did I miss this'. Okay, so. You did a really good job explaining, so you have all that right. And your math is all correct so... good job! You could have turned this in as your final and gotten 100%

> Federico: Okay, thank you. Em... Well I think now you know the errors?
> Revati: I feel so dumb because I know I should have done that. I don't know why I didn't do that. Can I see mine?
> Revati: I don't even know what I was doing. I was trying to do this this morning. Probably not the best idea in the world.

Mike summarized this same experience:

> If I got stuck at a part of the PAR, and I've looked at theirs, and their steps make more sense than mine I'll consider it, look over what they did, and change what I did wrong.

As did Harry:

> I really like looking at other people's initial models. I can see what they are thinking, it puts me in their head, and I can see that. A lot of times I'm really wrong and I can see different ways to do the same thing.

In his interview, Kevin also spoke a lot about learning from reading other students' work:

> All of the time I look at how other people work the problems, because everyone works it differently. Algebra was really weak for me coming into this class, and just seeing how other people approach things has helped me out quite a bit.

Kevin continued on, elaborating about darts:

> Darts gives us 3 examples of how the PAR should have been, or at least a part of the PAR. It shows you what you need to do to justify a little better. All in all it helps, in PAR especially, but really the whole course because we're supposed to justify a lot of our answers. All in all the darts it shows you how you should work things, or at least an example of how you should, and ways you shouldn't, so it's helpful.

Other students gave similar responses, noting that the darts activity was a useful means of gaining access to sample working in order to assess their own work.

8 Summary and Impact

The main purpose of this dissertation was to better understand how reflection could promote explanations and improve performance in mathematics. Introductory calculus is a notoriously difficult area of mathematics. This made it an excellent context for studying a novel approach to learning: I was unlikely to encounter ceiling effects on student growth and the intervention could have a meaningful impact on students and their future career goals.

8.1 Impact on Learning

The design had two primary goals: (1) improve lack of success in introductory calculus, and (2) deepen shallow understandings of concepts in introductory calculus. Through iterative cycles of problem solving, reflection, feedback and analysis, and revision, students learned to explain and reflect on mathematics. PAR was implemented in a context in which there were classroom discussions involving reflective questioning, student self-assessments of their work on a daily basis, and students regularly working collaboratively in groups and explaining their mathematical reasoning to others. Although the context of this study was introductory calculus, PAR is a flexible method and can be applied to a variety of contexts.[1]

During phase I success rates (A,B,C) were improved by 13%, and during phase II they were improved by 23%. Exam scores also echoed this improvement, with differences of up to 10-15% on individual exams. Students significantly improved their ability to explain mathematical ideas; students improved on their use of language (mathematical terms and avoidance of pronouns), use of diagrams, organization of work, and explaining why, not just what. Considering the students who were sampled for qualitative analysis, the Phase I explanation scores in the experimental section were more than twice as large as those in the comparison section; during Phase II, the scores in the experimental section were over four times as large as in the comparison section. Students learned how to look through the work of others, make sense of it, and provide meaningful feedback. These skills were then seemingly transferred to self-reflection. Through engaging in PAR students learned to work collaboratively. The iterative nature of PAR helped students develop the persistence required to solve challenging problems.

[1]Since the end of this study, it has also been used in ordinary differential equations and introductory mechanics.

8.2 Design

8.2.1 Design Products

This dissertation was premised on the importance of high-quality tasks. Throughout the pursuit of this project, I collected and modified a large number of mathematical tasks (see appendix G). In particular, I found that having additional opportunities for explanation and generating examples supported student interactions, because the students had different solutions to discuss with one another. This set of tasks is now in use by instructors at the math department in which the study took place. I believe the tasks would also be of great applicability to others teaching introductory calculus. The task design principles are broadly applicable.

I also generated questions that could be useful for problem solving or reflecting on solutions (see appendix A.3). Student interviews showed that these questions were not used in a way that optimally supported student learning. Nevertheless, with additional work, these could potentially be incorporated into a successful intervention. For instance, I believe there may be utility in integrating these questions into an online learning environment, which often doesn't provide many opportunities for reflection.

The PAR activity structure and darts training exercise are very general methods that should be broadly applicable. I believe that PAR techniques could be used in a wide variety of mathematical and scientific domains (e.g., physics and chemistry) with little to no modification.

8.2.2 Design Evolution

I refined the experimental intervention over the course of three stages. I drew primarily on the insights that: (1) explanation promotes learning (Chi et al., 1994); (2) peer-analysis and self-reflection are closely related (Black et al., 2003), and (3) self-reflection requires practice analyzing a large number of examples of various quality (Sadler, 1989). These insights guided me to create a design in which students had lots of opportunities to analyze work and practice explaining mathematics, yet, they provided little guidance how to implement them in the classroom.

The first stage took of the study place in an elementary algebra course. Initial attempts were largely ineffective. However, they led to two important insights: the expert barrier, and the impact of an authentic audience. The expert barrier prevented students from engaging meaningfully with the tasks, because the students perceived the tasks as ones that only experts could engage in, even though they were exactly the types of tasks that could help the students gain expertise. This was addressed by reframing the tasks as having students provide feedback, but not being expected to make explicit judgments about correct and incorrect work. The other insight was that for students to learn to better communicate and explain their ideas, they needed an authentic audience to explain to. It was difficult for the instructor to provide feedback, because the students simply thought the instructor was being overly critical. However, when other students responded that they couldn't understand the explanations, it provided visceral feedback that the explanations were lacking. Relatedly, giving students an authentic audience for their analyses (helping a peer rather than just completing a mathematical task), they were more likely to see the utility of the activity.

These insights were embodied into a standard tool for analysis and reflection that served the basis for the intervention in phase I. Simultaneously, I created a regular activity structure called Peer-Assisted Reflection (PAR), that provided students with opportunities to: work on difficult problems, self-reflect, analyze a peer's work and provide/receive feedback, and revise and turn in a final solution. The logic behind this structure was that it would help students make connections between self-reflection and peer-analysis, and that the regular structure would make it easier to implement into regular classroom practice.

The iterative nature of the PAR cycle turned out to be crucial to its effectiveness. Students were forced to regularly revisit their work on a single problem, in contrast to their standard practice in which they worked on a problem once, turned it in, and forgot about it. PAR problems were difficult, and most students could not solve them on their first try. However, between peer-conferences and other collaborations that took place before their final revisions, students were able to make significant progress and solve many of the problems. This helped them see that through sustained work it was often possible to solve problems that they couldn't solve initially on their first try. Evidently, the iterative nature of PAR made it a powerful intervention for helping students develop persistence in problem solving.

The communicative aspects of explanation also played an important role in PAR's effectiveness. Thinking through another person's communication helped students think about various aspects of each other's solutions. The focus on communication was well-aligned with the framing of helping other students be more successful, so students were willing to put in effort to provide meaningful feedback.

As powerful as the activity structure was, it also required considerable training for students to engage with meaningfully. During phase II a regular training activity, darts, was used to great effect. By having students analyze sample work for problems they had just completed, it removed the difficulties and time constraints of working out a new problem before analysis. Also, because students had just spent many days working through the PAR problem, they were generally interested in knowing the correct answer the problem, so they actively engaged with the sample solutions, trying to determine what was correct.

These key features of PAR are summarized as follows:

- The PAR cycle provided opportunities for students to revisit and revise their work multiple times. This promoted persistence in problem solving.
- Students analyzed one another's work silently before conferencing. This helped students focus on each other's reasoning, not just the problem itself.
- Only the final solution was graded; this helped students see the initial solution as a learning opportunity.
- A small number of targeted prompts helped students focus on important aspects of their solution.

- Students explained their work to their peers; they received visceral feedback about communication from peers who could not understand their solutions.
- PAR focused on providing feedback, not grading. This helped students overcome the need to provide a "correct" answer.
- Students received regular training and instruction on how to analyze others' work and provide meaningful feedback. This seemed to improve the quality of their conversations in peer conferences.

The two other components of the design, (a) opening problems and self-assessments, and (b) the reflection framework, also evolved throughout the study. Reflective questions appeared to have little impact on students' mathematical learning practices outside of the classroom; their utility is unclear, but most likely limited. Opening problems were well-accepted by students and helped facilitate class sessions. Some students connected the use of two columns to the revision cycles of PAR, but otherwise, they were mostly independent of PAR, and could likely be modified without diminishing the impact of PAR. Finally, some students found rating their understandings using percentages to be useful, and self-assessments provided information to the instructor, but I was unable to demonstrate that they had a significant impact on students' reflection skills. Even during phase II students seemed to struggle to construct meaningful questions.

8.3 Toward a Theory of Peer-Analysis

There is still relatively little literature on peer-assessment and peer-marking. Unfortunately, this area suffers from a lack of conceptual clarity. A number of distinct activities are often lumped under the heading of peer-assessment, yet the various activities have different theoretical underpinnings and potential for supporting student learning. In my theoretical framing, I argued for the use of the term peer-analysis rather than peer-assessment, to make a clear distinction between these types of activities.

In particular, I found that framing the analysis tasks in a way that didn't make students feel as though they had to make judgments of correct and incorrect was an important part of engagement. In contrast, the explicit purpose of peer-marking is to judge correctness. This overemphasis on grades seems to negate some of the benefits that would be achieved by supportive feedback (R. Butler, 1988). Through peer-analysis students provided feedback to one another that was ultimately used to revise and improve their solutions. In peer-marking, this is generally not possible because the grades are being assigned for the purpose of evaluating the quality of a student's work, not allowing for revision. Finally, peer-analysis pushed students to think about one another's reasoning in depth. Peer-marking is unlikely to promote such thinking, because a salient feature is classification and sorting, not necessarily deep understanding. For all of these reasons, peer-analysis is a very distinct activity from peer-marking, and should be recognized as such. The literature would benefit from further conceptual clarity on this matter.

To further theories of peer-analysis, I have identified a number of potential research areas based on the present study.

- There is a need to better understand the impact of student feedback compared to instructor feedback. While instructors are more likely to have more accurate feedback, students seemed much more likely to listen to other students regarding matters such as communication.
- The darts training activity (analyzing three example solutions) supported students to improve the quality of their feedback, but it is only one of many possible training activities. Better understanding what types of training, how often they should occur, and when they should be sequenced in a course would be beneficial.
- I found value in students having opportunities to practice explanations. In particular, there seemed to be great utility in having students practice verbal explanations in addition to written. However, there are many ways students could be asked to explain their reasoning (e.g., how they thought about a task, how they understand a concept, how they might approach a proof) and a number of different formats for providing an audience (e.g., partners, small group, whole class). Which of these formats might be most effective and how they would support students to take risks is still unknown.
- The iterative design of PAR seems to be a promising intervention for promoting persistence. However, explicit conversations or student reflections were never targeted at developing this skill. It seems worthwhile to study how PAR might be used in combination with other interventions to impact student dispositions.
- I identified mechanisms through which PAR promoted learning (Iteration, Discussion, Explanation, and exposure to Alternatives). These mechanisms are not necessarily unique to PAR and could be used in other contexts. It would be worth study to better understand how some of these mechanisms (in particular iteration and exposure to alternatives) contribute to learning.

As with any new insight, the potential avenues for further inquiry are even greater than what seemed evident before the project began. These are just a few areas that I feel might be productive, but there are many other unanswered questions (e.g., how to best use reflective questioning, helping students self-assess and ask questions, promoting student justifications).

8.4 Future Directions

Given the success of the PAR intervention, it is currently being implemented in all sections of introductory calculus at the university the study was conducted at. This opens up a number of productive research paths: how to train relatively inexperienced teaching assistants to implement the activities, how suboptimal implementations might still impact learning, etc.

Another area I am currently pursuing is better understanding how to influence student dispositions towards learning. Collaborating with a number of physics educators I have been developing a Guided Reflection Form (GRF) to help students reflect on a number areas that are crucial to disciplinary practice. Combining the impact of PAR and the GRF together, I am interesting in studying how such dispositions develop.

References

AAAS. (1993). *Benchmarks for science literacy.* New York: Oxford University Press.

Andrade, H. (2010). Students as the definitive source of formative assessment: Academic Self-Assessment and the Self-Regulation of learning.

Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., & Tahan, K. (2011). *The condition of education 2011* (Tech. Rep. No. 2011-033). Washington, DC: National Center for Education Statistics.

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

Black, P., Harrison, C., & Lee, C. (2003). *Assessment for learning: Putting it into practice.* Open University Press.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in education*, *5*(1), 774.

Black, P., & Wiliam, D. (2009). Developing the theory of formative assessment. *Educational Assessment, Evaluation and Accountability*, *21*(1), 531.

Boaler, J., & Greeno, J. G. (2000). Identity, agency, and knowing in mathematics worlds. In J. Boaler (Ed.), *Multiple perspectives on mathematics teaching and learning* (pp. 171–200). Westport, CT: Ablex Publishing.

Boaler, J., & Staples, M. (2005). *Creating mathematical futures through an equitable teaching approach: The case of railside school.*

Bookman, J., & Blake, L. (1996). SEVEN YEARS OF PROJECT CALC AT DUKE UNIVERSITY APPROACHING STEADY STATE?*. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *6*(3), 221234. Retrieved 2013-02-03, from `http://www.tandfonline.com/doi/abs/10.1080/10511979608965825`

Bookman, J., & Friedman, C. P. (1999). The evaluation of project calc at duke university, 1989-1994. *MAA NOTES*, 253256. Retrieved 2013-02-03, from `http://www.www.maa.org/saum/maanotes49/253.html`

Bransford, J. D., Brown, A., & Cocking, R. (2000). Learning and transfer. In *How people learn: Mind, brain, experience and school, expanded edition.* Washington DC: National Academy Press.

Bressoud, D. M., Carlson, M. P., Mesa, V., & Rasmussen, C. (2013). The calculus student: insights from the mathematical association of america national study. *International Journal of Mathematical Education in Science and Technology*, *44*(4), 685–698.

Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. *Metacognition, motivation, and understanding*, *65*, 116.

Brown, A. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences*, *2*(2), 141–178.

Burmeister, S. L., Kenney, P. A., & Nice, D. L. (1996). Analysis of effectiveness of supplemental instruction sessions for college algebra, calculus, and statistics. *Research in collegiate mathematics education II*, 145154.

Butler, D., & Winne, P. (1995). Feedback and self-regulated learning: A theoretical synthesis. *Review of educational research*, *65*(3), 245.

Butler, R. (1988). Enhancing and undermining intrinsic motivation: The effects of task-involving and ego-involving evaluation on interest and performance. *British Journal of Educational Psychology*, *58*(1), 1-14.

Carpenter, T., Lindquist, M., Matthews, W., & Silver, E. (1983). Results of the third NAEP mathematics assessment: Secondary school. *The Mathematics Teacher*, *76*(9), 652659.

CCSSI. (2010). *Common core state standards for mathematics* (Tech. Rep.). Retrieved from `http://www.corestandards.org/assets/CCSSI\_Math\%20Standards.pdf`

Chi, M., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive science*, *13*(2), 145182.

Chi, M., De Leeuw, N., Chiu, M., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive science*, *18*(3), 439477.

Cobb, P., Confrey, J., diSessa, A. A., Lehrer, R., & Schauble, L. (2003). Design experiments in educational research. *Educational Researcher*, *32*(1), 9–13.

Cohen, E. G., & Lotan, R. A. (1997). *Working for equity in heterogeneous classrooms: sociological theory into practice*. New York, NY: Teachers College Press.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciencies*. Routledge. Retrieved 2013-11-07, from `http://books.google.com/books?hl=en&lr=&id=Tl0N2lRAO9oC&oi=fnd&pg=PR11&dq=cohen+statistical+power+analysis&ots=dp5DZhlXTr&sig=D9wHJZwfiOVDrMdlTBq05eELPrE`

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the learning sciences*, *13*(1), 1542.

Davis, R. B., & Vinner, S. (1986). The notion of limit: Some seemingly unavoidable misconception stages. *The Journal of Mathematical Behavior*.

Douglas, R. G. (1986). *Toward a lean and lively calculus: conference/workshop to develop alternative curriculum and teaching methods for calculus at the college level, tulane university, january 2-6, 1986* (Vol. 6). Mathematical Assn of Amer.

Dunlosky, J., & Lipko, A. (2007). Metacomprehension: A brief history and how to improve its accuracy. *Current Directions In Psychological Science*, *16*(4), 228–232.

Dunlosky, J., & Thiede, K. W. (2013, April). Four cornerstones of calibration research: Why understanding students' judgments can improve their achievement. *Learning and Instruction*, *24*(0), 58–61. Retrieved 2012-12-19, from `http://www.sciencedirect`

`.com/science/article/pii/S0959475212000345` doi: 10.1016/j.learninstruc.2012.05.002

Engle, R. (2011). The productive disciplinary engagement framework: Origins, key concepts, and developments. *Design research on learning and thinking in educational settings: Enhancing intellectual growth and functioning. New York: Routledge*.

Featherstone, H., Crespo, S., Jilk, L. M., Oslund, J. A., Parks, A. N., & Wood, M. B. (2011). *Smarter together! collaboration and equity in the elementary math classroom.* Reston, VA: National Council of Teachers of Mathematics.

Fullilove, R. E., & Treisman, P. U. (1990). Mathematics achievement among african american undergraduates at the university of california, berkeley: An evaluation of the mathematics workshop program. *The Journal of Negro Education*, *59*(3), 463–478.

Ganter, S. (1999). An evaluation of calculus reform: A preliminary report of a national study. *MAA NOTES*, 233236. Retrieved 2013-02-03, from `http://vwwv.maa.org/saum/maanotes49/233.html`

Hanna, G., & Barbeau, E. (2008). Proofs as bearers of mathematical knowledge. *ZDM*, *40*(3), 345353.

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of educational research*, *77*(1), 81.

Hurley, J. F., Koehn, U., & Ganter, S. L. (1999). Effects of calculus reform: Local and national. *American Mathematical Monthly*, 800811. Retrieved 2013-02-03, from `http://www.jstor.org/stable/10.2307/2589613`

Kahneman, D. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux.

Lombrozo, T. (2006). The structure and function of explanations. *Trends in cognitive sciences*, *10*(10), 464470.

Michaels, S., O'Connor, M., Hall, M., & Resnick, L. (2010). ACCOUNTABLE TALK SOURCEBOOK.

Moreno, S. E., Muller, C., Asera, R., Wyatt, L., & Epperson, J. (1999). Supporting minority mathematics achievement: The emerging scholars program at the university of texas at austin. *Journal of Women and Minorities in Science and Engineering*, *5*(1).

NCTM. (2000). *Principles and standards for school mathematics.* Reston, VA: The National Council of Teachers of Mathematics.

NRC. (1996). *National science education standards.* Washington, DC: The National Academies Press.

Oehrtman, M., Carlson, M., & Thompson, P. W. (2008). Foundational reasoning abilities that promote coherence in students' function understanding. *Making the connection: research and teaching in undergraduate mathematics education*, 2742.

Osborne, J., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*.

Palinscar, A., & Brown, A. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and instruction*, *1*(2), 117175.

Pilgrim, M. (2010). *A CONCEPTS FOR CALCULUS INTERVENTION: MEASURING STUDENT ATTITUDES TOWARD MATHEMATICS AND ACHIEVEMENT*

IN CALCULUS. Unpublished doctoral dissertation, Colorado State University.

Pintrich, P. R. (2004). A conceptual framework for assessing motivation and self-regulated learning in college students. *Educational Psychology Review*, *16*(4), 385–407.

Raman, M. (2003). Key ideas: what are they and how can they help us understand how people view proof? *Educational Studies in Mathematics*, *52*(3), 319325.

Ranney, M., & Schank, P. (1998). Toward an integration of the social and the scientific: Observing, modeling, and promoting the explanatory coherence of reasoning. In S. Read & L. Miller (Eds.), *Connectionist models of social reasoning and social behavior* (pp. 245–274). Mahwah, NJ: Lawrence Erlbaum.

Ranney, M. A., Clark, D., Reinholz, D. L., & Cohen, S. (2012a). Changing global warming beliefs with scientific information: Knowledge, attitudes, and RTMD (reinforced theistic manifest destiny theory). In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th annual meeting of the cognitive science society* (pp. 2228–2233). Austin, TX: Cognitive Science Society.

Ranney, M. A., Clark, D., Reinholz, D. L., & Cohen, S. (2012b). Improving americans' modest global warming knowledge in the light of RTMD (reinforced theistic manifest destiny) theory. In K. J. van Aalst, M. Thompson, M. M. Jacobson, & P. Reimann (Eds.), *The future of learning: Proceedings of the 10th international conference of the learning sciences.* (pp. 2–481 to 2-482). International Society of the Learning Sciences, Inc.

Reinholz, D. L. (2013a). *Designing instructional supports for mathematical explanations.* Paper presented at the annual meeting of the American Educational Research Association, San Franscisco, CA.

Reinholz, D. L. (2013b). *That's Nice...But is it worth sharing?* Poster presented at the annual conference on research in undergraduate mathematics education, Denver, CO.

Roddick, C. D. (2001). Differences in learning outcomes: Calculus & mathematica vs. traditional calculus. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *11*(2), 161-84. Retrieved 2013-02-03, from `http://www.tandfonline.com/doi/abs/10.1080/10511970108965986`

Roddick, C. D. (2003). Calculus reform and traditional students' use of calculus in an engineering mechanics course. *Research in collegiate mathematics education V*, 56-78.

Sadler, D. (1989). Formative assessment and the design of instructional systems. *Instructional science*, *18*(2), 119144.

Saxe, G. B., de Kirby, K., Le, M., Sitabkhan, Y., & Kang, B. (in press). Understanding learning across lessons in classroom communities: A multi-leveled analytic approach. Retrieved 2013-08-05, from `http://www.culture.rcmeberkeley.com/sites/default/files/Saxe\_et\_al\_Methodology1.pdf`

Schoenfeld, A. H. (1985). *Mathematical problem solving.* New York: Academy Press.

Schoenfeld, A. H. (1988). When good teaching leads to bad results: The disasters of "well taught" mathematics courses. *Educational Psychologist*, *23*(2), 145–166.

Schoenfeld, A. H. (1991). What's all the fuss about problem solving. *ZDM*, *91*(1), 48.

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. *Handbook of research on mathematics teaching and learning*, 334–370.

Schoenfeld, A. H. (1995). A brief biography of calculus reform. *UME Trends: News and Reports on Undergraduate Mathematics Education*, *6*(6), 35. Retrieved 2013-02-03, from `http://www.eric.ed.gov/ERICWebPortal/recordDetail?accno=EJ536613`

Senk, S., & Thompson, D. (2003). *Standards-based school mathematics curricula: What are they? what do students learn?* Lawrence Erlbaum.

Sfard, A. (1991). On the dual nature of mathematical conceptions: Reflections on processes and objects as different sides of the same coin. *Educational studies in mathematics*, *22*(1), 136.

Sfard, A. (2007). When the rules of discourse change, but nobody tells you - making sense of mathematics learning from commognitive standpoint. *Journal of Learning Sciences*, *16*(4), 567615.

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research*, *78*(1), 153–189.

Smith, J., diSessa, A. A., & Roschelle, J. (1993). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *Journal of the Learning Sciences*, *3*(2), 115–163.

Songer, N., & Gotwals, A. (2012). Guiding explanation construction by children at the entry points of learning progressions. *Journal of Research in Science Teaching*.

Star, J. R., & Smith, J. P. (2006). An image of calculus reform: Students experiences of harvard calculus. *Research in collegiate mathematics education. VI*, 126.

Steen, L. (1988). The science of patterns. *Science*, *240*(4852), 611616.

Steen, L. A. (1988). *Calculus for a new century: A pump, not a filter: National colloquium: Papers*. Mathematical Association of America. ((US))

Tall, D. (1990). Inconsistencies in the learning of calculus and analysis. *Focus on Learning Problems in mathematics*, *12*(3), 4963.

Tall, D. (1992). Students difficulties in calculus. *Proceedings of Working Group 3 on Students Difficulties in Calculus, ICME*, *7*, 1328.

Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational studies in mathematics*, *12*(2), 151169.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249.

Treisman, U. (1992). Studying students studying calculus: A look at the lives of minority mathematics students in college. *The College Mathematics Journal*, *23*(5), 362372. Retrieved 2013-02-03, from `http://www.jstor.org/stable/10.2307/2686410`

White, B., Frederiksen, J., & Collins, A. (2009). *The interplay of scientific inquiry and metacognition: More than a marriage of convenience.*

Wilson, R. (1997, February). "Reform calculus" has been a disaster, critics charge. *The Chronicle of Higher Education.* Retrieved from `http://chronicle.com/article/`

Reform-Calculus-Has-Been-a/77411/

Young, A. (2010). *Explorations of metacognition among academically talented middle and high school mathematics students.* Unpublished doctoral dissertation, UNIVERSITY OF CALIFORNIA, BERKELEY.

Zimmerman, B. (2002). Becoming a self-regulated learner: An overview. *Theory into practice*, *41*(2), 6470.

A Frameworks

A.1 Pilot Study

The execution, explanation, justification framework was populated primarily by entries suggested by the students in the experimental section. It was designed for the analysis of written work.

Execution (what did you do?)	**Explanation (why did you do it?)**	**Justification) (did you do it correctly?)**
Show all solution steps in order	State any assumptions made	Check units
Define all variables in the problem	Explain why you chose to solve the problem a certain way	Estimate or solve a simpler problem
Write down important information from the problem statement	Draw a picture or diagram	Interpret answer using the problem context
Include units on all quantities	Use words to explain meaning of arithmetic operations in problem context	Solve the problem using another method
Answer all questions asked in the problem	Explain choice of representations	

A.2 Phase I

The framework for phase I revolved around reflective questions in three categories: reasoning, justification, and conceptualization. These questions were designed both to aid problem solving and for reflecting upon work.

About the **reasoning** involved in solving the problem:

- *Why would you ...?*
- *What was the purpose of...? How does it help?*
- *How would someone know to...?*

About the mathematical **justification** involved in solving the problem:

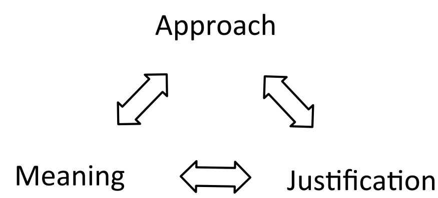
- *Why can you ...? Why does it work?*
- *How do you know that...?*
- *Do things always work this way? What if ... was different?*

About the **conceptualization** of the problem:

- *What does it mean that ...?*
- *What's really going on here?*
- *How does one visualize, interpret, or think about ...?*
- *What other situations might ... apply to?*

A.3 Phase II

The reflective questioning framework for phase II was simply a streamlined version of the phase I framework.



In order to develop reflective understanding, it is important to practice asking reflective questions during and after problem-solving. A number of such reflective questions are given below:

- **Approach:** *How would you approach . . . ?*
  - *Why would you . . . ? How would someone know to. . . ?*
  - *What was the purpose of. . . ? How does it help?*
- **Justification:** *How do you know that . . . ?*
  - *Why can you . . . ? Why does it work?*
  - *Do things always work this way? What if . . . was different?*
- **Meaning:** *What does it mean that . . . ?*
  - *What's really going on here?*
  - *How does one visualize, interpret, or think about . . . ?*

If you really understand something, you should be able to answer these questions.

B Self-Reflection Forms

B.1 Pilot Study

Big Picture: What was the approach to solving the problem?

Accuracy: Do you think the problem was solved correctly? How do you know? (Note any errors you found.)

Questions: What are you unsure about in the problem or response? What do you want to know more about?

B.2 Phase I

Note: This reflection form was implemented starting with sixth PAR problem. For the first five PAR problems the form from the pilot study was used.

Communication: *Was your solution communicated clearly?*

Did you clearly explain your reasoning (i.e. why you did what you did)? yes __ no __

Did you avoid the use of pronouns? yes _____ no _____

Did you give a written explanation of your calculations/graphs? yes _____ no _____

Did you label all graphs, include units, etc.? yes _____ no _____

Accuracy: *What evidence do you have that you solved the problem correctly?*

Was the problem solved in multiple ways? yes __ no __

Was the solution checked in any way? yes __ no __
If so, how? ________________________________

Did you explain (in writing) how you know your solution is correct? yes __ no __

Questions: What are you unsure about in the problem or response? What do you want to know more about?

Extension: What did you learn from this problem that you can apply to solving other problems? (Did you learn something about a concept, a problem-solving technique, etc.?)

B.3 Phase II

Hwk0X PAR Problem Problem Solver's Name: ______________________________

On a scale from 0% to 100%, how confident do you feel in your solution? __________

Completeness, Organization, and Labeling
Did you answer all questions asked, showing all steps, in the proper order? yes __ no __

(If applicable) Did you label and explain all graphs, include units, etc.? yes __ no __

Explanations
Did you explain why (not just what)? yes __ no __

Use of Language
Did you avoid the use of pronouns (and other ambiguous language)? yes __ no __

(If applicable) Did you consult definitions of mathematical terms you used? yes __ no __

Justification
Did you justify your solution (in at least 1 of the following ways): yes __ no __

- by checking if answers to different parts of the question are consistent?
- by explaining (in writing) how you know your solution is correct?
- in some other way? If so, how? ______________________________________

Note: Show explicitly on your solution *how* you justified your solution.

(Optional:) Is there anything in particular you'd like to discuss with your partner?

C Peer Feedback Form

C.1 Pilot Study

Note: The peer- and self-reflection forms were identical in the pilot study.

Big Picture: What was the approach to solving the problem?

Accuracy: Do you think the problem was solved correctly? How do you know? (Note any errors you found.)

Questions: What are you unsure about in the problem or response? What do you want to know more about?

C.2 Phase I

Note: This reflection form was implemented starting with PAR06. For PAR01-PAR05 the form from the pilot study was used.

Communication: Was their solution communicated clearly? (Give at least one suggestion to improve the presentation of the solution.)

Accuracy: What evidence did they provide that the problem was solved correctly? (Push your partner to justify "how they know;" also, note any errors you found.)

Questions and Feedback: What are you unsure about in their solution? How could it be improved?

C.3 Phase II

Hwk0X PAR Problem Peer-Assister's Name: ______________________

Communication: Give at least one suggestion to improve the communication of the solution. (Focus on explanations, imprecise use of language, organization, labeling, etc. Be specific: don't say "it was hard to follow" or "part 2 was unclear;" say *why* it was hard to follow, *what* was unclear, and *how* to improve it.)

Correctness: Note any errors you found. (Focus on misunderstanding of concepts, misuse of mathematical language, calculational errors, incomplete answers, etc. Be specific: don't just say "part 2 is was wrong;" say exactly *what* is wrong, *why* it is wrong, and *how* to improve it.)

(Optional:) What other feedback do you have? How else could the solution be improved?

D Yellow Paper

D.1 Pilot Study

At the end of each class period, students were asked to answer a number of written reflection questions. They are given:

1. Answer the day's guiding question.
2. "Traffic light" how well you understood the lesson.
3. Explain something you learned today as if you were explaining to a classmate who missed the lesson.
4. Ask one question about mathematics.
5. Tell me something else you think I should know.

Examples of guiding questions that were used are given:

1. What is the role of parentheses in algebra?
2. How does recognizing expressions help us solve equations?
3. How does "balance" relate to solving equations?
4. How do you know if an equation is a line?
5. Why focus on equations where the right hand side is 0?

D.2 Phase I

Assessment of Learning Name: ______________________

On a scale from 0 to 100%, rate how well you understand today's class, where 100% = understood everything, 50% = understood half of the class, and 0 = understood nothing.

What questions do you have? (What was unclear? How does the day's lesson relate to other math concepts?; **Write at least 2 questions.**)

Tell me something else you think I should know.

D.3 Phase II

Assessment of Learning Name: ____________________

On a scale from 0 to 100%, rate how well you understand today's class, where 100% = understood everything, 50% = understood half of the class, and 0 = understood nothing.

Write at least 1 math content-related question (denoted with a Q). If you can't think of any questions, you may write a conceptual exam question (denoted with an E next to it).

(Optional:) Tell me something else you think I should know (or ask non-content questions).

E Survey Questions

1. How do you define "problem solving" in mathematics?
2. How do you know when you've "solved" a mathematics problem? (ie. how do you know when you're "done?")
3. How does explaining mathematics relate to understanding mathematics?

For questions 4-7, imagine that you and your classmates have completed your homework assignment (independently), and have written your solutions on paper.

4. What might you learn from having fellow students look at your work?
5. What might you learn from looking at the work of fellow students?
6. How might giving and getting feedback from other students be *better* than from the teacher?
7. How might giving and getting feedback from other students be *worse* than from the teacher?

The following questions were borrowed (with permission) from the Modified Indiana Mathematics Belief System (MIMBS). Note that Belief 2 has been reverse coded in this study, in order to make it consistent with the other scales; all of these scales represent positive beliefs.

Belief 1 (Student's Self Confidence): I can solve time-consuming mathematics problems.

+ Math problems that take a long time don't bother me.
+ I feel I can do math problems that take a long time to complete.
+ I find I can do hard math problems if I just hang in there.
- If I can't do a math problem in a few minutes, I probably can't do it at all.
- If I can't solve a math problem quickly, I quit trying.
- I'm not very good at solving math problems that take a while to figure out.

Belief 2 (The Nature of Mathematics): Mathematics problems cannot be solved by simply finding the correct procedure.

- Learning to solve math problems is mostly a matter of memorizing the right steps to follow.
- Most math problems are easy to solve once you figure out what type of problem they are.

- \- Any math problem can be solved if you know the right steps to follow.
- \+ Many math problems cannot be solved by following a predetermined sequence of steps.
- \+ Some math problems aren't like any of the common types of problems.
- \+ There is no procedure to solve many math problems.

Belief 3: Understanding concepts is important in mathematics.

- \+ Time used to investigate why a solution to a math problem works is time well spent.
- \+ A person who doesn't understand why an answer to a math problem is correct hasn't really solved the problem.
- \+ In addition to getting a right answer in mathematics, it is important to understand why the answer is correct.
- \- It's not important to understand why a mathematical procedure works as long as it gives the correct answer.
- \- Getting a right answer in math is more important than understanding why the answer works.
- \- It doesn't really matter if you understand a math problem, if you can get the right answer.

F Interview Protocol

Here I provide the final version of the interview protocol that was used during phase II of the study. There were some minor changes from the phase I protocol, but most of the questions were the same. The script and questions follow.

As you know, the math department is piloting some new ideas to help students in math 160 develop deeper understandings of calculus. It's my job to find out if they're working, and how they might be improved. So I'd really appreciate it if you could be completely honest with me. If something just isn't working, the department needs to know it. If something can be improved, learning that would be great too. And if something has been helpful, it's important to know that, too. On the basis of what students like you tell me, the ideas the department is trying will either be abandoned or expanded. For that reason, it's really important for you to be honest with me. Anything you say is strictly confidential, and will not affect your grades in any way.

We'll begin by talking a little bit about the course in general, then the PAR, working with other students, daily reflections, and finally any other thoughts you have. Do you have any questions about the purpose of this interview or what I'm asking from you?

1. Opening Questions
    a) What is your major and mathematics background? Have you taken calculus before?
    b) How has this class been different or the same as other math classes you've taken in the past?
        i. Same?
        ii. Different?
    c) In this class, what's working well and what's working not so well for you?
        i. Well?
        ii. Not so well?
    d) How much time do you spend outside of class working on this course?
2. PAR
    a) Let's discuss the PAR. What's working well and what's working not so well for you?

   i. Well?
   ii. Not so well?

b) After you complete the PAR independently, you do a self-assessment. What works well and not well about that?
   i. How do you typically do the self-assessment?
   ii. When do you do it?

c) Do you feel like the self-assessment portion of the PAR is related or unrelated to the peer-assessment portion?
   i. how is it related?
   ii. how is it unrelated?

d) Does PAR have any influence on how you do your regular homework assignments or not? (For example, the self-reflection or peer-conferencing)

e) What type of PAR feedback have you received from other students?
   i. To what extent has it helped, or not?
   ii. Have you ever revised your work based on the feedback? How? Could you tell me a little more about that?

f) Do you feel like some students are more helpful than others, or are they all about the same?
   i. Do you have any way of determine whether or not you think a student's feedback is correct, or not really?
   ii. Were there any times where you disagreed with feedback or didn't take it into account, or do you always agree with / use the feedback? Can you tell me more about such a situation?
   iii. Do you feel like you can tell whether or not another student knows what they are talking about, or not really? If so, how would you do it? If not, why not?
   iv. Have you ever looked at another student's work and felt like you did something wrong, or has that never happened to you? How do you determine whether or not you think they are correct?

g) What type of feedback have you tried to give to other students?
   i. Did other students indicate it was helpful, or not? What did they say?
   ii. Do you feel like you are able to be honest with other students about their work, or not really? (please elaborate)
   iii. Have you ever noticed something that you didn't talk to the other student about? (please give an example)
   iv. Do you ever have feedback that you don't share because you're not sure about it, or do you always say what you're thinking?

h) Do you ever revise your work for reasons other than peer-feedback? What might those reasons be and what would it look like?

i) Do you feel like the way you engage in PAR has changed over the course of the semester, or has it stayed the same? (could you elaborate with an example?)

j) To what extent has the "darts" activity supported or not supported you to be successful in PAR, and in the course more generally?
   i. Supported you?
   ii. Not supported you?

3. Working with other students

   a) How do you feel about working with other students in this class? What are some positive and not-so-positive reactions you've had?
   b) Outside of the PAR, have you worked with other students in this class? for instance, on homework or labs. in study groups? How about TILT tutoring?
      i. If so, how did you form your study group(s)?
   c) Is working with other students something you do in other classes? in study groups?
      i. If so, is PAR similar or different from the way you work with other students outside of class?
      ii. If not, do you think you feel differently about PAR because you don't usually work with other students?

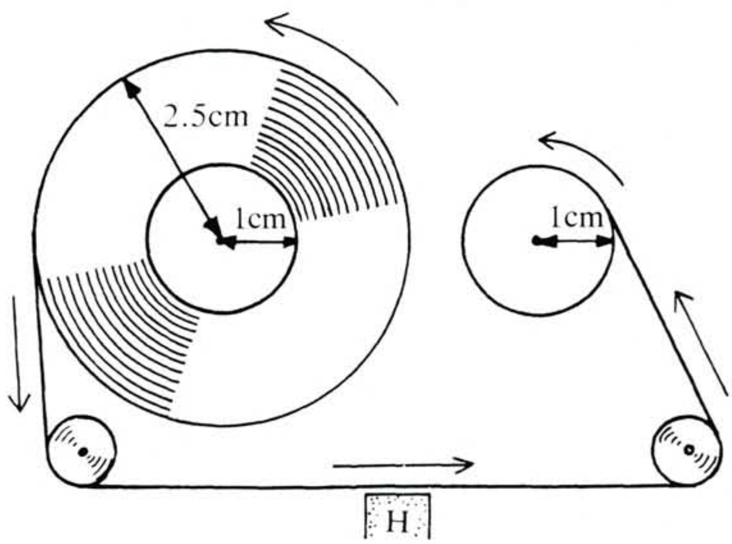
4. Reflections and questions

   a) In this class, you write reflections and ask questions every day. What are your positive and not-so-positive reactions to doing the daily reflections?
      i. To what extent have reflections and questions supported your learning, or not?
      ii. Do you feel like the instructor has taken your feedback into account, or not really?
      iii. How do you determine the percentage of what you knew? What is helpful and not helpful about this?
      iv. How do you come up with questions? Are they "legitimate?" When do you tend to write them down? How does it fit in with other activities of participating in class?
      v. Do you feel like you have enough time to complete the daily reflection in class, or do you feel rushed?
   b) What do you think about the opening problems? Have they been helpful or not really? Why?
      i. On the opening problem, your paper is split into two columns for an initial and final solution. Do you find that helpful or not really? (please elaborate)
      ii. Have you ever split your paper like this in your own work, or not really? If so, has it helped? If not, why not, do you think?
   c) Let me ask you about the poster and the "3 questions" your instructor uses. What's useful and what's not so useful?
      i. Have you ever used these questions to guide your learning? If so, can you tell me more? If not, why not, do you think?

5. Closing
   a) Do you have any comments about the PAR and this math class that I haven't asked?
   b) If you could change one thing in this course to help yourself learn more, what would it be?

G PAR Tasks

G.1 Cassette Tape



This diagram represents a tape recorder just as it is beginning to play a tape. The tape passes from the "head" (labeled H) at a constant speed and the tape is wound from the left-hand spool on to the right hand spool. At the beginning, the radius of the tape on the left-hand spool is 2.5cm. The tape lasts 45 minutes.

1. Sketch a graph to show how the *length* of the tape on the left-hand spool changes with time (label your graph as appropriate).
2. Explain (with words) why the graph you drew above has the shape that it does.
3. Sketch a graph to show how the *radius* of the tape on the left-hand spool changes with time (label your graph as appropriate).
4. Explain (with words) why the graph you drew above has the shape that it does.
5. Describe (with words) how the radius on the tape of the *right-hand* spool changes with time. Explain why it changes this way.

Extension Questions: How would the graphs change if you pressed "fast forward" rather than play on the cassette player? What would happen to the graph if you pressed pause while the tape was playing? How does the radius of the spools influence your graphs?

G.2 Examples Library

Suppose that $f(x)$ denotes a function defined for all real numbers. Each of the statements below is true *sometimes*. For each of the statements below give an example of a function for which it holds true and an example of a function for which it does not hold true.

1. $\lim_{x\to 3} f(x) = f(3)$
2. If $\lim_{x\to 0} \frac{f(x)}{x} = 1$, then $f(0) = 0$.
3. If $\lim_{x\to -7} f(x) = a$, then $\lim_{x\to -7} \frac{1}{f(x)} = \frac{1}{a}$. (You may choose any value of a that you wish for your examples.)
4. If $\lim_{x\to 5^+} f(x) > 0$ and $\lim_{x\to 5^-} f(x) > 0$, then $\lim_{x\to 5} f(x) > 0$.

Extension Questions: Look at the solutions you gave. How many more examples can you come up with for which the statements are true and untrue? Is it possible to create an infinite number of examples?

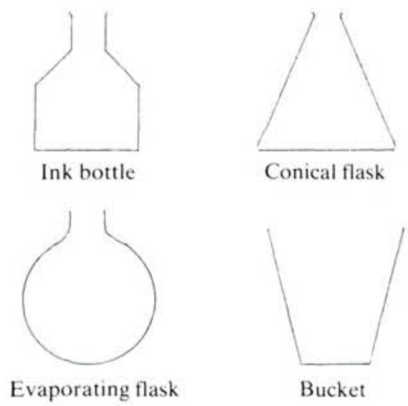
G.3 Proof and Counterexamples

Indicate whether each of the following statements is **True** or **False**. If the statement is true, explain how you know it's true. If it is false, give a counterexample. (A counterexample is an example that shows the statement is false.)

1. If $f(-1) = -2$ and $f(1) = 2$, then there is a point c between -1 and 1 where $f(c) = 0$.
2. If $f(a) = 0$ and $f(b) = 0$, and f is continuous on $[a, b]$, then there is a point c between a and b (ie. $a < c < b$) such that $f(c) = 0$.
3. The graph of a function with a horizontal asymptote can never touch that asymptote.
4. The graph of every rational function has a vertical asymptote.
5. If $\lim_{x \to a} f(x) = \infty$ then $f(x)$ is not continuous at a.

Extension Questions: For each of the untrue statements, is it possible to add additional hypotheses to make it true? (e.g., #1 would be true if you also assume that f is continuous).

G.4 Filling Bottles



For a science experiment, you are filling up various bottles with liquid, pictured above. The liquid comes from a tap that pours water at a constant rate.

1. For each bottle ((a) ink bottle, (b) conical flask, (c) evaporating flask, and (d) bucket), sketch a graph of the height of liquid in the bottle as a function of time. Label your graphs appropriately.
2. For each bottle ((a) ink bottle, (b) conical flask, (c) evaporating flask, and (d) bucket), sketch a graph of the rate of change of the height of liquid in the bottle as a function of time. Label your graphs appropriately.
3. Look at the graph you created in part (2) for the conical flask. Imagine that this graph described the height of liquid in a bottle as a function of time. Draw a sketch of the bottle.

Extension Questions: Draw a graph of your favorite function. Is it possible to create a bottle so that the graph of your function represents the height of water in that bottle? What characteristics must a function have so that it could represent the height of the water for some bottle?

G.5 Favorite Non-linear Function

Draw an accurate graph of your favorite nonlinear function $y = f(x)$ (no formula for $f(x)$ needed) and pick a point on the x-axis and label it "a". (Make the graph fairly large so you can clearly draw other things on it.) Recall that the derivative of a function f at a point $x = a$ is defined by

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$$

(a) Illustrate and label on your graph each of the following quantities that appear in the definition. Then write a short statement explaining in terms of your graph what each quantity means (1-2 sentences for each quantity).

(i) $f(a)$
(ii) h
(iii) $f(a+h)$
(iv) $\frac{f(a+h)-f(a)}{h}$
(v) $f'(a)$

(b) Explain in terms of the graph what the equation $f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h}$ means (be sure to talk about secant lines and the tangent line).

(c) Explain why the limit is needed to find the slope of the tangent line. (Why can't we just use arithmetic and algebra?)

Extension Questions: Why does this question ask you to use a nonlinear function? How would things be different if you chose h to be positive versus negative? (Does it matter?)

G.6 Radians vs Degrees: Who Wins?

When you work with $\sin(x)$ in calculus, it is assumed that x is given in radians. Let $\sin(x)$ refer to the sine function that takes inputs in radians, and let $S(x)$ refer to a **different** function, the sine function that takes inputs in **degrees** (you might verify that $S(x)$ and $\sin(x)$ are indeed different functions). Define $\cos(x)$ and $C(x)$ in a similar fashion for the cosine function in radians and degrees. It follows that $S(x)$ and $C(x)$ are given by:

$$S(x) = \sin(\frac{2\pi}{360}x) \qquad C(x) = \cos(\frac{2\pi}{360}x)$$

1. Find $S'(x)$ and $C'(x)$. Express your final answers in terms of $S(x)$ and $C(x)$ (i.e. after you find $S'(x)$, rewrite your final answer so that it no longer contains $\sin(x)$ or $\cos(x)$).
2. How does the relationship between $\sin(x)$ and its derivative compare to the relationship between $S(x)$ and its derivative?
3. How does the relationship between $\sin(x)$ and its fourth derivative compare to the relationship between $S(x)$ and its fourth derivative?
4. Why do you think we use radians in calculus?

Extension Questions: How would you take the 100^{th} derivative of $\sin(x)$? What about $S(x)$? Does $[S(x)]^2 + [C(x)]^2 = 1$? (i.e., does the Pythagorean identity hold true for both degrees and radians?) Why or why not?

G.7 Cone of Power

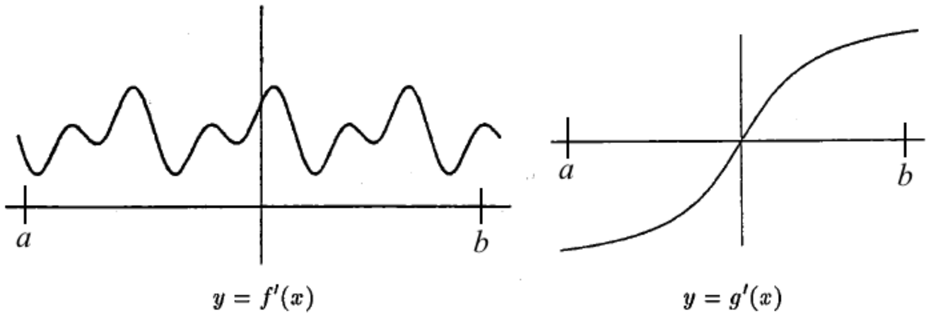
A right circular cone has radius r and height h. The radius r of the cone is increasing at a rate of 3 cm/sec while its height h is decreasing at a rate of 2 cm/sec. (The volume of a right circular cone is given by $V = \frac{1}{3}\pi r^2 h$.)

(a) Find $\frac{dV}{dt}$ as a function of r and h.
(b) How fast is the volume of the cone changing when $r = 10$ cm and $h = 20$ cm?
(c) Is the volume increasing or decreasing at this time? (Explain how you know.)
(d) For which values of r and h is the volume increasing? (Write your answer as an inequality involving both r and h.
(e) For which values of r and h is the volume decreasing? (Write your answer as an inequality involving both r and h.
(f) Is it possible to choose initial values for r and h for which the volume of the cone is always increasing? Explain why or why not.

Extension Questions: How would things change if the radius were decreasing and the height were increasing? Is it possible to choose rates of change for the radius and height so that the volume always remains constant?

G.8 Derivative Solutions

The graphs of the *derivatives* of two functions f and g are given below.



1. How many solutions can the equation $f(x) = 0$ have on the interval $a \leq x \leq b$? Explain how you know, elaborating each possible case.
2. How many solutions can the equation $g(x) = 0$ have on the interval $a \leq x \leq b$? Explain how you know, elaborating each possible case.
3. Suppose $g(x) = 0$ has two solutions on $a \leq x \leq b$. What can you say where about these solutions lie? Explain.

Extension Questions: Would would the graph of a derivative $h'(x)$ for a function with $h(x) = 0$ having 3 solutions look like? 4 solutions? n solutions?

G.9 O.I.L.

An oil drilling rig located 14 miles off of a straight coastline is to be connected by a pipeline to a refinery 10 miles down the coast from the point directly opposite the drilling rig. Laying underwater pipe costs twice as much as laying pipe on land. (Hint: You may want to draw a few pictures to get an idea of the possibilities. Scoring criteria for optimization problems are on the MATH 160 web site under Strategy for Optimization Problems.)

1. Which combination of underwater and land-based pipe will minimize the total cost of the pipeline? What is the minimum cost if pipe costs $50,000 per mile on land?
2. Suppose that underwater pipe costs α times as much as pipe on land ($\alpha > 1$). Which combination of underwater and land-based pipe will minimize the total cost of the pipeline now? (Write your answer in terms of α.)
3. How does the ideal amount of pipe under water and on land change as α increases? Does this seem to make sense to you?

Extension Questions: What would happen to your results of the refinery were 20 miles down the coast, rather than 10 miles?

G.10 Hand Area

In this problem, you will trace the shape of your hand and approximate the area of the picture that you create. Your main tasks are to devise a method for approximating the area and to show that your approximation is very close to the actual area.

1. Put your hand flat on the grid provided (with fingers touching, no gaps) and trace the shape of the outline of your hand. Make sure that the shape you trace is a function (if not, erase the parts of the shape that would make it not a function).
2. Devise a method to approximate the area of the region inside the curve you have traced. Explain your method in detail, and explain why it should work. (Don't perform any calculations yet.)
3. Use the method you described above to approximate the area of the outline of your hand. (Show your work.)
4. Describe a method for estimating the error in your method of approximation. (Error is something you would like to make *small!* Thus an estimate for the error means being able to say the error is **less than** some value.)
5. Calculate an estimate the error for your method.
6. Explain (in principle) how you could improve your method to make your estimate as accurate as one could want (i.e., minimize the error). (You do not actually have to perform the calculations, just explain what you would do.)

Extension Questions: Think about implementing your method on a computer. Is it something that could reasonably be done? If not, how much you modify your approach to calculating area? Compare your method with your partner's; which do you think will be more accurate? Would your methods work well for any arbitrary function?

G.11 An Odd Function

Let $f(x)$ be differentiable on $[-3,3]$, with $f(x) = 0$ for $x = -2, 0,$ and 2 (and nowhere else). Let $f(x)$ have critical points at $x = -1$ and $x = 1$ (and nowhere else). Also, let $f(-1) = 1$. Finally, let $f(x)$ be an odd function, meaning that $f(x) = -f(-x)$ for all x (i.e. f is symmetric about the origin). Define a new function G on $[-3,3]$ by the formula

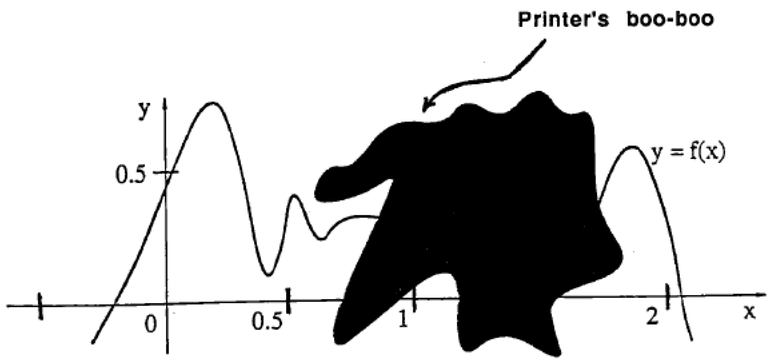
$$G(x) = \int_{-2}^{x} f(t)dt$$

(a) Sketch a rough graph of $f(x)$. (Because there are multiple acceptable graphs for $f(x)$, use your particular sketch for the rest of the problem.)
(b) Find the values of $G(-2)$ and $G(2)$. Estimate a maximum possible value for $G(0)$. (Explain how you made your estimate.)
(c) Find the critical points and possible inflection points of $G(x)$ on $[-3,3]$.
(d) Sketch a rough graph of $G(x)$ on $[-3,3]$.
(e) Interpret the points found in (c) in terms of the graph of $f(x)$.

Extension Questions: How would things change if you started the integral from $t = 0$ rather than $t = -2$? How would things change if you had an even function ($f(x) = f(-x)$) rather than an odd function?

G.12 Ink Blot

Suppose a function f is twice differentiable over $[0, 2]$ and that its graph is given below. As you see, the printer was sloppy and spilled a lot of ink on the graph. Decide, if possible, whether each of the following definite integrals is positive, negative, or equal to 0. Explain your answers.



(a) $\int_0^2 f(x)dx$

(b) $\int_{1/2}^1 f\left(x+\frac{1}{2}\right) dx$

(c) $\int_0^1 xf(x^2)dx$

(d) $\int_0^2 f'(x)dx$

(e) $\int_0^2 f''(x)dx$

(bonus) $\int_0^{1/4} f''(x)dx$

Extension Questions: Suppose the ink blot covered the whole area of $[0.25, 1.75]$. Which integrals could you still estimate? How large could the ink blot be before you could no longer estimate those integrals?

G.13 Gabriel's Horn

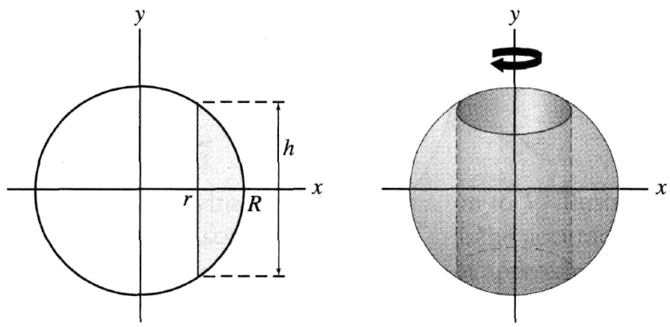
All parts of this problem are based on the function $f(x) = \dfrac{1}{x^{2/3}}$.

1. Find an expression for the area between $f(x)$ and the x-axis for $x = 1$ to $x = ☺$ for some constant $☺ > 1$.
2. Take the limit as $☺ \to \infty$ of your answer to (1) to find the area under the curve $f(x)$ for x-values from 1 to ∞.
3. Find the volume of the solid formed by rotating the region bounded by $f(x)$ and the x-axis around the x-axis for x-values $1 \leq x \leq ☹$, for some constant $☹ > 1$.
4. Take the limit as $☹ \to \infty$ of your answer to (3) to find the volume of this infinitely long "trumpet" (known as Gabriel's Horn).
5. How can your results to (2) and (4) both be true? Explain.

Extension Questions: Suppose your lower limits of integration were some arbitrary value N rather than 1. How would this affect your answers above?

G.14 Beads

A bead can be formed by removing a cylinder of radius r from the center of a sphere of radius R (see the figure below).



A bead is a sphere with a cylinder removed.

(a) Use calculus to find the volume V of the bead with $r = 1$ and $R = 2$.
(b) Use calculus to find the volume V of a bead in terms of the variables r and R.
(c) The bead's height h is labeled in the figure. Rewrite your formula from (b) to show that $V = \frac{\pi}{6}h^3$.
(d) Since your answer in part (c) expresses the volume entirely in terms of h (and not r or R), it means that all beads of the same height have the same volume. In other words, if you started with a sphere the size of an orange and a sphere the size of a basketball and made them each into beads a height of 2 inches, the beads would have the same volume. Explain how this can be true. (Hint: think about the shape of the beads)
(e) Do all beads of the same height h also have the same outside surface area (not including the surface area of the cylindrical hole inside)? (Note: you do not need to use an integral to compute the surface area, just discuss it intuitively.)

Extension Questions: What if the shape removed was something other than a right circular cylinder. Would your results still be the same? What if the rotated region was something other than a portion of a circle?

H Sample Exam

The following exam was the third midterm for phase I of the study. This exam is typical of the other exams that were used during the study.

1. (15 points) A rectangle is inscribed with its base on the x-axis and its upper corners on the parabola $f(x) = 3P - x^2$ (where $P > 3$). Complete the following parts of this question to find the dimensions of such a rectangle with the greatest possible area.

   a) Draw a diagram of the situation, labeling the variable dimensions of the rectangle.

   b) Write an equation for a function A that expresses the area of the rectangle in terms of the

      variable(s) identified in (a). $A =$ ______________________________

      What is the domain of this function? ______________________________

   c) Find the dimensions of the inscribed rectangle that will maximize its area (write your answers

      in terms of P). Width = ____________________ Height = ____________________

      The maximum area of the inscribed rectangle is $A_{\text{max}} =$ _______________

   d) Explain how you know that the value you found in (c) is the maximum area.

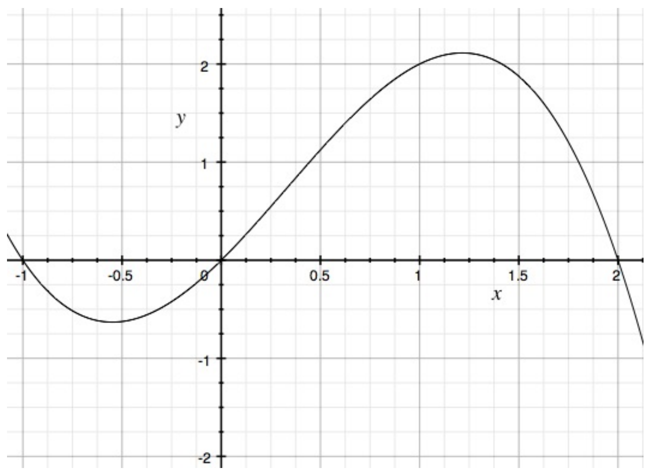
   e) Describe what happens to the shape of the inscribed rectangle as P increases. (Relate your answer to how both the length and width of the rectangle depend on P.)

2. (12 points) The Riemann definite integral $\int_a^b f(x)dx$ is defined as a limit of Riemann sums. The definite integral can be *approximated* by a finite sum:

$$\int_a^b f(x)dx \approx f(c_1)\Delta x_1 + f(c_2)\Delta x_2 + \ldots + f(c_{n-1})\Delta x_{n-1} + f(c_n)\Delta x_n$$

where $\Delta x_1 = (x_1 - x_0), \Delta x_2 = (x_2 - x_1), \ldots, \Delta x_{n-1} = (x_{n-1} - x_{n-2}), \Delta x_n = (x_n - x_{n-1})$.

A function $y = f(x)$ defined on the interval $[a, b] = [-1, 2]$ is shown in the figure.



a) Explain in words how to interpret $x_0, x_1, x_2, \ldots, x_{n-1}, x_n$. (Illustrate in the figure with $n = 6$.)

b) Explain in words how to interpret $f(c_1), f(c_2), \ldots, f(c_{n-1}), f(c_n)$. (Illustrate in the figure with $n = 6$.)

c) The definite integral is defined as a limit of Riemann sums. What does the limit mean, and what purpose does it serve?

d) Is $\int_{-1}^{2} f(x)dx$ positive or negative? (Explain why in terms of the function $f(x)$ and the x-axis.)

3. (15 points) Aaron and LeAnn are discussing $y = F(x)$ defined by $F(x) = \int_0^x \sqrt{16 - t^2}dt$ for $-4 \le x \le 4$.

   a) Aaron said, "Look LeAnn, using my calculator I found that $F(4) = 12.5664\ldots$" LeAnn interrupted, "Wait! You didn't need to do that! It's easy to get the exact value of $F(4)$ from the graph you have on your calculator screen!" The exact value of $F(4) =$ ______________.

   b) What graph did LeAnn see on the screen of Aaron's calculator? How could she get the exact value of $F(4)$ from what she saw?

   c) For what values of x is $F(x)$ negative? Explain why $F(x)$ negative for these values of x. ("Because I graphed $F(x)$ on my calculator and can see where $F(x)$ is negative" does not explain why!)

   d) Are there any values of x for which $F(x) = 0$? If not, explain/show how you know there are not. If so, find all such values of x and explain why the function is zero at the x value(s) you found.

   e) The derivative of the function $y = F(x)$ is $F'(x) =$ ____________________

4. (5 points) Let $f(x) = x^2 + 1$ on the interval $[1, 5]$. If left-endpoints are used with $n = 100$ subintervals to approximate the area of the region between $f(x)$ and the x-axis on the interval, then
   a) The approximation will be less than the actual area.
   b) The approximation will be greater than the actual area.
   c) The approximation will be equal to the actual area.
   d) We cannot determine if the approximation is greater than or less than the actual area.
5. (6 points) Suppose you are trying to find a zero of the function $f(x) = (x-3)^3 + x^2$ using Newton's Method with an initial guess of $x_0 = 4$.
   a) Write the equation of a line (in point-slope form) that you could use to the find the next approximation, x_1.

   b) How could you use the equation you wrote in order to find x_1? (Explain why this makes sense.)

6. (20 points) Evaluate the following integrals. Answers must be accompanied by supporting work that shows how you evaluated the integral. Write coefficients as fractions reduced to lowest terms. Simplify numerical answers as directed. Answers will be scored right or wrong. Dont expect partial credit for incorrect answers or answers with no supporting work.

   a) $\displaystyle\int \left(\frac{4}{\sqrt[3]{x^2}} + 2x^3 \right) dx$

   ---

   b) $\displaystyle\int \left(20x + \frac{5}{x^3} + \sec^2(8x) \right) dx$

   ---

   c) $\displaystyle\int \left(\frac{2t^2 + \pi t^{2/3}}{3t} \right) dt$

   ---

   d) $\displaystyle\int_{\pi}^{2\pi} (\sin(2x) - \cos(2x))\, dx$

   ---

   e) $\displaystyle\int_{-1}^{2} (|y| - 2)\, dy$

7. (16 points) Indicate whether each of the following statements is **True** or **False**. If the statement is true, explain how you know it's true. If it is false, give a counterexample. (A counterexample is an example that shows the statement is false.)

   a) If $\int_0^1 f(x)dx = 4$, then $\int_0^{1/2} f(x)dx = 2$

   b) If $\int_0^1 f(x)dx = 4$, then $\int_0^1 \frac{f(x)}{2}dx = 2$

   c) If $f'(x) > 0$ for $x < -3$ and $f'(x) < 0$ for $x > -3$, then there must be a local maximum at $x = -3$.

   d) If $f'(4) = 0$ and $f''(4) > 0$, then there must be a local minimum at $x = 4$.

8. (15 points) Sketch the graph of a function that has all of the following properties:

	$\boldsymbol{f(x)}$	$\boldsymbol{f'(x)}$	$\boldsymbol{f''(x)}$
	$\lim_{x\to-\infty} f(x) = -\infty$		
$x < -2$	$\lim_{x\to-2^-} f(x) = 3$	$f'(x) > 0$	$f''(x) < 0$
$x = -2$	$f(-2) = 3$	$f'(-2)$ DNE	$f''(-2)$ DNE
$-2 < x < 0$	$\lim_{x\to-2^+} f(x) = -\infty$	$f'(x) > 0$	$f''(x) < 0$
$x = 0$	$f(0) = 2$	$f'(0) = 0$	$f''(0) < 0$
$0 < x < 1$		$f'(x) < 0$	$f''(x) < 0$
$x = 1$	$f(1) = 0$	$f'(1) < 0$	$f''(1) = 0$
$x > 1$	$\lim_{x\to\infty} f(x) = -2$	$f'(x) < 0$	$f''(x) > 0$