COGNITIVE AND AFFECTIVE COMPONENTS OF UNDERGRADUATE STUDENTS LEARNING HOW TO PROVE

By

Visala Rani Satyam

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Program in Mathematics Education—Doctor of Philosophy

2018

ABSTRACT

COGNITIVE AND AFFECTIVE COMPONENTS OF UNDERGRADUATES LEARNING HOW TO PROVE

By

Visala Rani Satyam

Students struggle with proving, a fundamental activity in upper-level undergraduate mathematics courses. Learning how to prove is a difficult transition for students, as they shift from largely computation-based to argument-based work. In response, mathematics departments have instituted courses, introduction or transition to proof, designed to help students learn how to prove. Existing research has extensively examined students' errors, struggles, and some of their strategies at a given point in time, but we know little about students' development over a longer period of time. There is a need for longitudinal work in this area, to follow students through the transition to proof.

In addition, little is known about the affective side of proving (e.g., attitudes, beliefs, emotions). Affect plays a central role in mathematics learning, influencing students' cognitive processes while problem solving and their motivation to value and want to do mathematics. Understanding affective issues are important, as students consider their future participation in mathematical work and communities. Positive experiences at transitional junctions, such as learning how to prove, are crucial for retention of students through the STEM (Science, Technology, Engineering, and Mathematics) pipeline.

The purpose of this work was to explore the cognitive and affective factors involved in undergraduates' efforts to learn how to prove: how their proving developed during a transition to proof course and what kinds of *satisfying moments*, i.e. positive emotional reactions, they experienced. Four semi-structured interviews across a semester were conducted with eleven undergraduate students enrolled in a transition to proof course. The resulting data was analyzed using qualitative methods.

Findings indicate that students showed growth in fluency, strategy use, and monitoring and judgement over time. Four developments were frequently observed across the sample: (1) increased sophistication in students' rationales for choice of proof techniques, (2) awareness about how a solution attempt was going and managing that for their subsequent strategies, (3) intentional exploring and monitoring when unsure about what direction to pursue, and (4) checking examples in conjunction with other strategies as a way to become unstuck. The variety of developments – and the different ways in which they emerged – is significant, because it confirms that multiple developments occur in different ways, strongly suggesting that there is no one path that students take through the transition to proof.

Students' satisfying moments were largely about accomplishments both with and without struggle, understanding, external validation, as well as interacting with others. A theory for how satisfying moments are elicited was proposed. Expectations and a sense of mastery played large roles in mediating satisfying moments, but students' desire for understanding and sense-making was also prominent.

This work provides guidance for curriculum design of transition to proof courses, in considering how to support students' development in proving. In addition, examination of just what makes a satisfying moment satisfying is helpful in thinking about how to construct mathematical tasks with opportunities for positive experiences with math.

To the love of my life, no matter what – mathematics.

For every person who believed in me and my mathematical endeavors,
this is dedicated to you.

ACKNOWLEDGMENTS

This has been a long road. I knew I wanted to study mathematics since I was in fourth grade and thought cross-canceling when multiplying fractions was the most amazing thing. There are many people who have brought me then to where I am today.

They say we stand but on the shoulders of giants; the same can be said for our teachers. I thank the math teachers in my life: my fourth-grade teacher Ms. Drouin, my Intermediate Algebra II teacher Mr. Bradman (under whom I failed a math test for the first time due to the Boston Red Sox winning the ALCS), and my calculus teacher Mr. Pasquale. In college: Ben Hescott for the warmth he brought to all things theoretical, Dr. Kim for her energy and that proof about Galois theory that was a true moment of mathematical beauty to me, and Professor Quinto for tea and cookies and conversations about doing math long after. My ideals of who I'd like to be as a math teacher and professor have been shaped by all of you. To Judah Schwartz, who introduced me to mathematics education as a field - he would set students off arguing against each other, like a wonderfully ornery grandpa. And this list would not be complete without Mrs. Mooney, who while she was not my teacher, although she was a teacher (indeed vice-principal), encouraged me from my very first high honor roll in middle school to her end.

I was so fortunate to land at Michigan State University at the time I did and experience all the good things that came of my years there. I thank my affective research group (Eileen, Kevin, and Missy) for helping me think about how to study this crazy thing called emotions and my writing group for being the support I sorely needed my last semester: Amit, Andy, Amanda, Brittany, Jose, Merve, Valentin, and Younggon.

In terms of faculty, only time will tell me just what a bounty I had around me. I thank Russell Schwab for allowing me into the course. I am indebted to my committee - Mariana Levin, Shiv Karunakaran, Teena Gerhardt, and previously on my committee, Vince Melfi - for being so thoughtful and often offering the words to describe what was in my head that I myself struggled to find. To my advisor, Jack Smith (our fearless leader) – thank you for believing in my work when I do not. I think you read me accurately from our very first meeting and have supported me as an individual, like knowing I'll critique my work enough for the both of us. But in all seriousness, a heartfelt thank you to you.

As a factor in my coming to Michigan, I thank figure skating for providing me escape as needed – and for friends from that part of my life who were outlets as my precise mathematical career aspirations shifted through college. To Big Charlie and Jacqui White, for picking me up the first time I arrived in Michigan for a grad school visit and from there on became part of my Michigan support system. Our long conversations about passions and success and failure and how to find joy in the grind stay with me.

But this dissertation especially would not have been possible without my friends around me. Thank you to AJ, Amy, and Farheen for long phone calls and texts that felt so comfortable – but also giving me solutions. To Lucy, whose joy for my joy of math is inspiring in and of itself; your cards mean so much to me!

I am so grateful for my fellow PRIME graduate students across all the years and staff; this is a community and support that is not to be easily imitated elsewhere. (I include all the math grad students who are basically math ed also in this as well) To Amy, for being my job search and dissertation buddy, Jill for being my dilemma-master through grad school and essentially teaching me qualitative methods, Joanne for mentoring me through each step of the dissertation, Kate for our weekly Thursday lunches and asking about things in my life and hearing me, and Molade for our walks to cardio kickboxing and WhatsApp voicenotes - you always know just what to say to comfort me.

And now for the 5th floor of Wells Hall (and co), who steadily went from "math friends" to straight up friends over the years: Allison for her I believe in you's, Charlotte for her drop-by office visits and walks, Fatemah and Loura, Marowan, and Yammn for being the best neighbors one could ask for (what's more rejuvenating for the soul while dissertating than a hug from a child?), Mollee for reaching out and being kind beyond words, Tyler for always making sure to ask me how the big things in my life were going, Reshma for late night talks and godcat cuddles and check-in's, Sarah for walks to IM East and salted caramel mochas when I needed them most, and Hitesh for our morning walks from CVS to campus that I cannot imagine this year without – for being there.

I thank extended family and family friends who supported me in math over the years: my brother Ravi, Anand, Hima, Rupa and last but not least, my uncle SN Satyam, who gifted me books about space and math that nurtured my fledgling dreams. In the end, I say thank you to my parents, Venkata and Sujatha Satyam, for financially supporting me all through school and for ensuring that this was my calling. The truth is – even having these dreams would not have been possible without your hard work and struggle to come to this country and provide this level of security in life for me to do what I want. I am grateful.

I end then with a nod to all my participants, the true stars of this work. Through our conversations across the interviews, I reflected and learned so much about my own collegiate experience with mathematics. The relationships cultivated in that space are the real products of this dissertation.

PREFACE

It is a funny thing, where we get ideas from – how long they sit with us unbeknownst or when the same kind of question about the nature of things emerges in varying contexts. While working on this dissertation, I realized both of these were true.

I watch a lot of figure skating. Like many others, I watch for those special moments when the crowd would go crazy and instantly rise to their feet. I often wondered – What guarantees a standing ovation? Why is it that some performances end with good applause but others pull the crowd out of their chairs, as though electric? Perhaps it had something to do with the rise and fall of the music or the way a skater hit certain movements or maybe even a sequence of the above that would elicit that automatic rise to one's feet. Could one purposely design for this?

Without realizing it at the time, I seek to answer the same question in this dissertation but with an eye toward mathematics: What kinds of mathematical experiences bring about an internal standing ovation for an individual, i.e. feelings of satisfaction and elation? Are there features in common across individuals when these events happen and if so, can we as instructors intentionally create learning opportunities with these features embedded? This may be playing with fire; humans are delicate things and their emotions even more so, nowhere nearly deterministic as to be easily managed. But in a similar way to how Tolstoy wrote that "Happy families are all alike," perhaps satisfying mathematical experiences have threads in common too.

The above explains the genesis of the affective side of this work. The cognitive side came about in wanting to see how students' proving changed as they were learning, not just at one given snapshot in time. But soon into conducting interviews with students, this question was on my mind: why do some people interpret failure negatively and others positively?

Throughout these interviews, some students were dejected about their solutions when they felt they were wrong. But other students reacted to getting my problems wrong by asking me how it worked and outright saying that now they knew how to do it in the future. They genuinely saw failure as learning opportunities. I was shocked. Could this difference be the key?

To the reader, I urge you to keep this last question in mind as you read (or let's be honest, peruse) these chapters ahead. While this was not the research question I set out to answer and there may not be enough evidence to truly draw claims, I think it is the deep question at the heart of all this. I do believe that math educators, in all their forms, want the same things for their students, to grow and to feel good about math. I hope this work spurs some thoughts – and in keeping with the theme, feelings too - on these basic goals.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

MLC	Math Learning Center
RQ	Research Question
[word]	Additional word(s) for sake of clarity, not said by interviewee

CHAPTER 1: Introduction

The transition to proof is difficult for undergraduate students (Moore, 1994; Selden & Selden, 1987). Students struggle with learning how to prove (Iannone & Inglis, 2010; Selden & Selden, 2013). The transition to proof is a shift in the "game" of mathematics, from answering "exercises" that are largely procedural (Schoenfeld, 1992) to now writing arguments and justifying said answers.

Researchers have identified the types of errors students make (Selden & Selden, 1987) and their struggles (Harel & Sowder, 1998; Selden & Selden, 2003). Common proving errors in undergraduates' proofs are in regards to use of examples, notation and symbols, quantifiers, and general logic (Epp, 2003; Selden & Selden, 1987). Students struggle with larger issues as well, such as giving empirical rather than deductive arguments (Harel & Sowder, 2007) and having difficulty writing formal arguments (Alcock & Weber, 2010). Another strand of research has focused on students' strategies and approaches to the proving process (Karunakaran, 2014; Savic, 2012).

We know students' struggles and their strategies while proving at singular points in time, but few have looked at how these strategies change over the course of the learning process. Much existing research is about whether students understand logic and proof techniques, such as contradiction and induction. One way to interpret this work is that gaining the ability to prove statements is about the accumulation of individual techniques. But development is not necessarily about accumulating competencies; as Piaget (1964) said, "For some psychologists, development is reduced to a series of specific learned items, and development is thus the sum, the culmination of this series of specific items. I think this is an atomistic view which deforms the real state of things" (p. 38). Thinking about proving as the sum of skills and assessing whether or not students have those skills may not be enough for us to understand students' learning process. We may be able to tentatively assess their proof competencies at certain points in time, but we do not yet know *how* students put all these pieces together while they are learning how to prove nor the *order* in which these proving abilities develop. We lack models of students' cognitive development for how they learn how to prove, as a mathematical activity. There is a need for longitudinal work in work on undergraduates' proving (Smith, Levin, Bae, Satyam, & Voogt, 2017; Bae, Smith, Levin, Satyam, & Voogt, 2018), for having frequent interactions with the same students over a reasonable interval to see how they change

In addition, the affective side of learning has largely been understudied in mathematics education teaching and learning (McLeod, 1992; Sinclair, 2006). Affect plays a central role in mathematics learning but especially in problem solving (McLeod, 1994; Silver, 1985). Affect can also influence cognitive processes, such as knowing what to do next while problem solving (McLeod, 1988). McLeod (1992) claimed that any research can be strengthened by examining both affective and cognitive issues together. Within the context of proof learning, Selden & Selden (2013), called for more research on how students' affect influences their problem solving and proving work. Positive affective moments may provide the intrinsic motivation then (Middleton & Spanias, 1999) for students to continue doing and valuing mathematics. Moments of positive affect are therefore educationally desirable.

In summary, the field is currently missing developmental and affective examinations of how students learn how to prove. How do students learn how to prove? Moreover, is this an activity they wish to do more of?

Research Questions

In response to this gap, the purpose of this study is to examine both the cognitive and affective components involved in how undergraduates learn how to prove. The research questions are:

1. How does undergraduate students' proving develop over the duration of a transition to proof class?
2. What kinds of satisfying moments do undergraduate students have during the transition to proof?

This study contributes to research and practice about mathematics education teaching and learning in multiple ways. First, this work attempts to describe how undergraduate students learn how to prove. Second, this work examines the nature of affective experiences in mathematics, specifically at a transition point in students' mathematics education. These results may be of interest to mathematics education researchers with interests in proof, emotional responses, and cognitive approaches to learning in general. Lastly, findings from this work may benefit course developers, by informing the design of future undergraduate transition to proof courses, from managing expectations about the pace and depth of student understanding to engineering opportunities for positive, satisfying moments for students.

CHAPTER 2: Literature Review

In this chapter, I review literature in order to unpack the two major phenomena in my study, proving and emotions in regards to mathematics. First, I provide an overview of what we know about students' proving, the need for understanding development, and then proving from a problem solving perspective. Second, I discuss what we know about affect in mathematics education before focusing on emotion. My conceptual framing of the constructs used in this study will be discussed in Chapter 3.

Proof and Proving

What does it mean to prove? It is difficult to pin down a definition of what it means to prove but many have tried. One can think of proving in terms of creating a product, a proof. Stylianides (2007) provides us with one definition of a *proof*:

> *Proof* is a *mathematical argument*, a connected sequence of assertions for or against a mathematical claim, with the following characteristics:
>
> 1. It uses statements accepted by the classroom community (*set of accepted statements*) that are true and available without further justification;
>
> 2. It employs forms of reasoning (*modes of argumentation*) that are valid and known to, or within the conceptual reach of, the classroom community; and
>
> 3. It is communicated with forms of expression (*modes of argument representation*) that are appropriate and known to, or within the conceptual reach of, the classroom community. (p. 291; emphasis in original.)

A proof can generally be thought of then as an *argument* with certain norms of expression. A proof also has generality, distinguishing it from computations which are tied to specific instantiations of variables.

We can also consider what activities constitute proving. “Given this definition of proof, we define *proving* broadly to denote the activity in search for a proof." (Stylianides, Stylianides, & Weber, 2016). Indeed, covering all that constitutes proving is difficult. In terms of reasoning, deductive reasoning is often associated with proofs, but inductive and abductive reasoning are at play as well. Some specific activities that constitute proving include:

- Constructing a proof
    - Estimating the truth of a conjecture
    - Justifying a statement estimated to be true
- Presenting a proof
    - Taking audience conviction into account
    - Explaining to an audience
    - Demonstrating validity
    - Demonstrating understanding
- Reading a proof
    - Proof Comprehension
    - Proof Evaluation

(Mejia-Ramos & Inglis, 2009, p. 90).

Why should we care about proof? Proving is often thought of as a foundational activity in mathematics (Harel & Sowder, 2007). Some purposes for why we should produce proofs are listed here:

- Verification (demonstrate truth)
- Explanation/illumination (why it is true)

- Discovery (discovering new things in process of proving)
- Systematization (organizing results into a system)
- Intellectual challenge (affective feeling of self-realization and fulfillment)
- Communication

(Bell, 1976; de Villiers, 1990).

A proof serves many functions then, beyond just a way of verifying the correctness of mathematical statements. It is important to note that mathematics did exist prior to proof; proof as a notion is attributed to Euclid's *Elements*. However, Euclid's form of argumentation has been so useful that it has become a staple of mathematics as a discipline and remains today (Harel & Sowder, 2007).

Proof and Mathematics Education

Within the field of mathematics education, there was some attention to proving in the early 21st century (Fawcett, 1938), but most of the work has come in recent times. Recent educational standards have staked the importance of proving at all ages, e.g. Common Core State Standards in Mathematics (NGA & CCSSO, 2010) in the United States. Indeed, there is the notion that proof should play an important role in all students' mathematical education (e.g., Hanna & Jahnke, 1996; Mariotti, 2006).

Proving is a difficult activity, however, and students have a hard time learning how to prove (Baker & Campbell, 2004; Moore, 1994; Selden & Selden, 2013). This is not surprising; in everyday life, people use examples as verification for truth. We are not accustomed to general, formal arguments. It is sensible that a foreign skill would take time to learn. One reason is that they have little experience with proving (Jones, 2000). Students are used to computations or following algorithms, as laid out by the curriculum, and precursors to proof like explaining one's work is not overly common in the curriculum. For students in the United States, high school geometry is typically the first place they encounter the word proof, by way of two-column proofs. But the highly constrained nature of two-column proofs make it not an adequate introduction to proving (Herbst, 2002).

Students struggle with proving at all ages: mathematical justification and proof in middle school (e.g. Bieda, 2010; Knuth, Choppin, & Bieda, 2009; Staples, Bartlo, & Thanheiser, 2012) and geometry proofs in high school (e.g. Senk, 1989). Introducing young children to the ideas of proof is a developing topic of interest (e.g. Bieda, Drwencke, & Picard, 2014; Stylianides, 2007).

The majority of research on students proving has been at the undergraduate level. There is lots of research on students' difficulties, and the difficulties are many. One common issue is in using empirical instead of deductive arguments (Harel & Sowder, 1998; Recio & Godino, 2001). Harel & Sowder (1998) proposed the idea of a proof scheme to be a person's conception of proof, of what counts as ascertaining (remove one's own doubts) and persuading (removing others' doubts). Another issue is in translating informal to formal arguments (Alcock & Weber, 2010; Pedemonte, 2007; Pedemonte & Reid, 2011). If the "distance" between the informal and formal arguments are too wide, students struggle to produce a proof (Pedemonte, 2007). Other student difficulties are around proof-specific writing, such as using quantifiers and notation (Epp, 2003; Selden & Selden, 1987), proof methods (Stylianides, Stylianides, & Philippou, 2004, 2007), using theorems (Selden & Selden, 1987), generalization (Selden & Selden, 1987) and understanding and working with definitions (Dubinsky, Elterman & Gong, 1988; Moore, 1994).

Undergraduate students also struggle to tell whether a proof verifies a mathematical fact as true, i.e. they are not persuaded by proofs (Alcock & Weber, 2005; Inglis & Alcock, 2012; Ko & Knuth, 2013; Selden & Selden, 2003; Weber, 2010). This effect has been seen in preservice secondary teachers (Bleiler, Thompson, & Krajcevski, 2014) and also inservice secondary teachers (Knuth, 2002).

Selden and Selden (2007) distinguished between the *problem-centered* versus *formal-rhetorical* parts of proving. *The problem-centered* aspect of proving involves the decisions and key insights that are made in order to solve the embedded problem in the proof, oftentimes with no set procedure. The *formal-rhetorical* aspect of proving involves the logical structure of the proof. Students learning how to prove encounter difficulties of both of these types. Both aspects are necessary in order to interpret mathematical statements and try to prove them, although students may favor one approach to proving over the other (Weber & Alcock, 2004). Selden & Selden have worked on helping students with the formal-rhetorical difficulties of proving, through the use of their *proof frameworks.* While difficulties with formal-rhetorical aspects of proving hinder students especially in the beginning, the problem-centered aspect may pose a longer, more on-going struggle to students. There is still much left to be learned in the problem-centered aspect of proving, with its emphasis on strategies and decision-making, especially in terms of how students develop this sense in regards to proof.

In summary, research has established many of the ways in which undergraduates struggle. Some work has focused on how to help students (e.g. Blanton, Stylianou, & David, 2003). What we need more work on, however, is in what the learning process of proving looks like. What are students able to do and what does the learning process look like?

Learning how to prove is more than just accumulating individual skills or techniques, so analyzing just what they struggle to do is not enough to understand their learning. Development is not necessarily just about accumulating competencies. What more, despite the numerous difficulties, students somehow still learn how to prove through experience and with the help of instructors. Students may not gain full mastery of proving quickly but they do make progress. How can we understand the developments students go through in learning how to prove? For this, we go to the closest cousin of proving for which we have an abundance of research: problem solving.

Proving as Problem Solving

Research on proving has been conducted in a myriad of ways, with new approaches emerging especially over the last couple decades (Stylianides, Stylianides, & Weber, 2016). One way of looking at proving is as a form of problem solving (Savic, 2012).

For these reasons, I draw on the literature of problem solving, as well as that of proof. Problem solving as a research area was a common theme among mathematics education researchers of the 1980s and early 1990s (Schoenfeld, 1992; Silver, 1985). Non-routine mathematical problem solving may be thought of as situations "in which possessed knowledge of algorithms, facts, and procedures do not guarantee success" (Malmivuori, 2001, p. 7). I briefly describe the evolution of theory on mathematical problem solving below.

Polya (1945), the forefather of mathematical problem solving, described the problem solving process in a linear fashion: understanding the problem, devising a plan, carrying out the plan, and then reflecting back on one’s work in order to extend it for future problems. A number of theoretical frameworks for investigating problem solving have been created since then, building off Polya's work.

Garofalo & Lester (1985) brought into focus the importance of *metacognition* in problem solving, of having knowledge of one's own cognition and regulation of it. They identified four categories activities people engage in when working on a task - orientation, organization, execution, and verification – and how metacognition is involved in each. Schoenfeld (1985b; 1992)'s work on problem solving identified five components of problem solving: cognitive resources, strategies or heuristics, monitoring and control, beliefs and affect, and practices.

But problem solving need not be sequential; it can be a cyclical process. Carlson & Bloom (2005) found that subjects often go through cycles of reasoning when problem solving: making a plan, executing the plan, checking if the plan continues to work, and then creating a new plan if issues arose. This framework has been used to analyze proving as well (Savic, 2012), due to the similarities between problem solving and proving processes. In summary, various theories of mathematical problem solving have been developed and have built on each other, leading to the refined work we have today.

Transition to Proof Courses

Now I turn to a discussion of transition to proof courses. Considering all the difficulties inherent to proving, it is not surprising that mathematics departments have responded with courses designed to help students learn. The formation of introduction or transition to proof courses can be seen as a departmental response to students' struggles. These courses can take on many names, but I call all courses of this nature t*ransition to proof* for the sake of simplicity.

There is great variety in the design of transition to proof courses across the United States. David & Zazkis (2017) conducted a syllabus study to categorize the variety of designs. One common design is to teach proving as a stand-alone skill, with instruction on formal logic, quantifiers, proof methods, and propositions. The content of these courses is often around sets, functions, etc. A variation of this is for the majority of the course to be about logic and grammar, with an introduction to an advanced mathematical topic they will encounter in the future near the end. The other course design is to teach proving through a content area to provide some context, with oftentimes little explicit instruction to formal logic. In these courses, students are often expected to pick up how to prove along the way. On the other hand, there is an advantage to proof in the context of a content area, where proof as a means of discovery of new results is better motivated.

Multiple transitions taking place. Transition to proof courses are transitions in terms of content – proof-based work in place of computation. There is transition then in terms of cognitive aspects. But transition can also refer to a transition in terms of experience. Mathematics as many students are used to in K-12, of computations and algorithms, has now been exchanged for mathematical argumentation and writing. This constitutes a shift in students' mathematical experience, at a socio-emotional level (Smith, Levin, Bae, Satyam, & Voogt, 2016).

Affect

Affect is generally thought as the domain involving emotions (Middleton, Jansen, Goldin, 2017). McLeod (1992) defined the *affective domain* as "the wide range of beliefs, feelings, and moods that are generally regarded as going beyond the domain of the cognition" (p. 576).

One way to think of affect is as *a representational system*:

> *Affect* includes changing states of emotional feeling during mathematical problem solving (local affect)...and more stable, longer-term constructs (global affect), which establish contexts for local affect and which local affect can influence. Our hypothesis is that affect is fundamentally *representational*, rather than a system of mostly involuntary, physiological side-effects of cognition. (DeBellis & Goldin, 2006, p. 133)

For example, frustration while working on a problem serves as an indicator that something is not working (DeBellis & Goldin, 2006). Thus, frustration serves as an encoding of this cognitive noticing that current strategy is not working – and trying a new strategy should be taken.

Major Types of Affect in Relation to Mathematics Education

Three major types of affect include beliefs, attitudes, and emotions (McLeod, 1992). I provide definitions of each, using McLeod (1992)'s dimensions and Middleton, Jansen, & Goldin (2017)'s state vs. trait distinction to discuss these constructs and how they relate.

Attitudes. Attitudes are "orientations or predispositions toward certain sets of emotional feelings (positive or negative) in particular (mathematical) contexts." (DeBellis & Goldin, 2006, p. 135). Some examples of attitudes in mathematics education are being bored by algebra, curious about geometry, and disliking story problems. Attitudes are seen as traits, in that they are long-term and relatively stable to an individual, thus difficult to change.

Beliefs. *Beliefs* are "the attribution of some sort of external truth or validity to systems of propositions or other cognitive configurations" (DeBellis & Goldin, 2006, p. 135).

One pervasive example of a belief in mathematics education is believing in one is bad at mathematics. Other beliefs include self-efficacy and other motivational variables. Beliefs are often highly stable and perhaps the most difficult to change among attitudes, beliefs, and emotions.

Emotions. Emotions are "rapidly-changing states of feeling experienced consciously or occurring preconsciously or unconsciously" (DeBellis & Goldin, 2006, p. 135)**.** Emotions are generally thought of as responses to events. Emotions tend to be short in duration but can reach high intensity, in contrast to attitudes and beliefs tending to be long in duration but low in intensity. Emotions are local and oftentimes bound up in the context at hand. Emotion is the state (rapidly changing) of affect vs. attitudes/beliefs as traits (stable).

Emotions can also function as representations of the consequence of goals, thereby communicating information about the situation. For example, a person feels happy when they make progress or sadness when noticing a lack of progress (Middleton, Jansen, & Goldin, 2017). In addition, emotions do not sit in a vacuum away from attitudes and beliefs but are influenced by them: students' long-term interests and beliefs about a situation at hand can manifest themselves through their emotions (Middleton, Jansen, & Goldin, 2017).

Affective Work in Mathematics Education

Early work on attitudes. I provide here a brief overview of the history of studying affect in mathematics education. Early research in mathematics education regarding affect focused on attitudes, specifically students' attitudes towards mathematics (Higgins, 1970). This work in the 1970s was largely quantitative, administering questionnaires to large groups to measuring attitudes pre- and post- some intervention. Well-known attitude scales include the Fennema & Sherman (1976)'s mathematics attitudes scales, specifically meant to study gender differences but used by many researchers for general research on students' attitudes in mathematics.

Students' beliefs about math. The second wave of development in affective work in mathematics education came from a focus on problem solving. Attention was on students' beliefs about mathematics and how their beliefs influenced their problem solving. Teacher beliefs was also a large avenue of research, but since the focus of this review is on students, I do not discuss this more.

Studies of emotion as rare. Emotions are difficult to study. Emotions are much shorter in duration and thus fleeting, and thus hard to capture, compared to attitudes and beliefs. Trait-like variables are more easily measurable, due to their stability; survey work is an appropriate method gold this. This stability means they are not easily alterable, for good and for bad.

Studies focusing on emotion in mathematics education are far fewer than that of beliefs and attitudes (e.g. Gómez-Chacón, 2000; Op 't Eynde, De Corte, & Verschaffel, 2006, 2007). Historically, those that existed were typically around math anxiety (e.g. Buxton, 1981) but some recent studies have examined how emotions influence mathematical thinking and learning (e.g. Op 't Eynde, De Corte, & Verschaffel, 2007). Careful observation of students with detailed interviews can help researchers analyze emotional states of mathematics learners (McLeod, 1988, 1992).

Why care about studying emotions? Why does studying emotions matter, especially if they are fleeting in nature? One, emotions are the vehicle for changing attitudes and beliefs. Repeated emotional responses may lead to student having different attitudes, which then may be able to alter beliefs. Two, emotions themselves as in-the-moment and states *are the most responsive to change*. Middleton, Jansen, and Goldin (2017) asserted that "In-the-moment engagement, on the other hand, is more easily susceptible to immediate influence by the teacher" (p. 691). I assert the same is for emotions, as the affective construct with the shortest duration. There is a push for more work on in-the-moment affective constructs (Evans, 2002; Hannula, 2002; McLeod 1992).

CHAPTER 3: Conceptual Framing

In this chapter, I present some of the concepts used in my study which influenced its design. This is a separate chapter from the literature review for the sake of reader clarity. First, I conceptualize proving as problem solving, specifically what a person does when stuck with a focus on strategies and monitoring and judgment. I also briefly discuss my conceptualization of development which influenced the study design. Lastly, I define a new construct, *satisfying moments*, and relate it to existing constructs about intense positive emotions. Analytical frameworks will be discussed in a separate chapter.

Conceptualizing Proving as Problem Solving

I defined students' proving as students' problem solving in the context of proof, i.e., the work of constructing a proof for a given statement. I chose this particular conceptualization of proving for multiple reasons. First, I purposely wanted to keep the phenomenon of focus broad by using the term proving rather than narrowing my focus to a particular skill, e.g. deductive reasoning. Second, I wanted to focus on proving as a process, rather than the product (Karunakaran, 2014), to look at what students do and their strategies. Third, because the focus is on students' process, their objective performance on the tasks – whether they produced a successful proof at the end of the allotted time – was not so important in this research; what they attempt to do was more vital.

To consider proving to be a subset of problem solving, we must define what is meant by problem solving, given the rich research tradition about problem solving in mathematics. To keep things simple, I take problem solving to be what a person does when stuck. This is equivalent to what activity a person engages in when reaching an impasse (Savic, 2012). Under this definition, a task may elicit problem solving in one student but not another, depending on whether or not they become stuck at any point in the proving process. The operationalization of what is meant by stuck will be discussed in Methods.

When looking at what a person does when stuck, I focused on the components of strategies (heuristics) and monitoring and judgement of problem solving (Schoenfeld, 1985b; 1992). Strategies are “techniques for making progress on unfamiliar or nonstandard problems” (Schoenfeld, 1985b, p. 15). Monitoring and judgment can be thought of as self-regulation and fall under the umbrella of metacognition (Schoenfeld, 1992), knowledge of and regulation of one’s own thinking. I include these here in the conceptual framing because while I did not strictly adhere to Schoenfeld (1985b; 1992)’s frameworks regarding strategy and monitoring and judgment, it did highly influence my thinking and the design of this study (see discussion in Methods chapter about think-aloud).

Conceptualizing Development

Development refers to change over time, but even that can be thought of in multiple ways. For example, one way of thinking about development is in terms of stages, in which a person presumably passes through each stage on their way to full mastery (Piaget, 1971). One famous example of development is the Van Hiele (1959) levels of geometry thinking. Conceptualizing students learning by way of levels remains to this day (Cobb & Wheatley, 1986; Lo, Grant & Flowers, 2008). I conceptualize development as taking a “snapshot” - a characterization of some construct at a point in time - and looking across these at multiple timestamps for change. Figure 3.1 illustrates this idea.

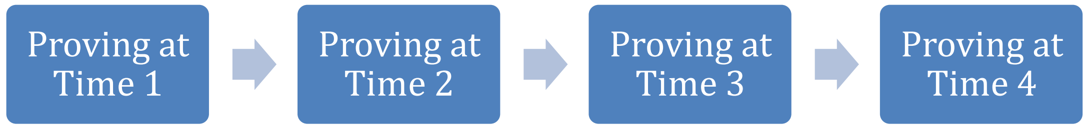


Figure 3.1. Conceptualization of development in students' proving by capturing snapshots of student's proving and compare across time.

Defining Satisfying Moments

I define a satisfying moment to be an emotional response to a particular moment in time, characterized by intense positive feelings. I think of a satisfying moment as being located within an experience, which serves as the context or situation which leads up to the satisfying moment. The use of the word moment is meant to suggest this event holds an instantaneous feeling to the individual, regardless of whether it is in reality. However, the distinction between a moment vs. an experience, the latter of which implies a duration, is not important.

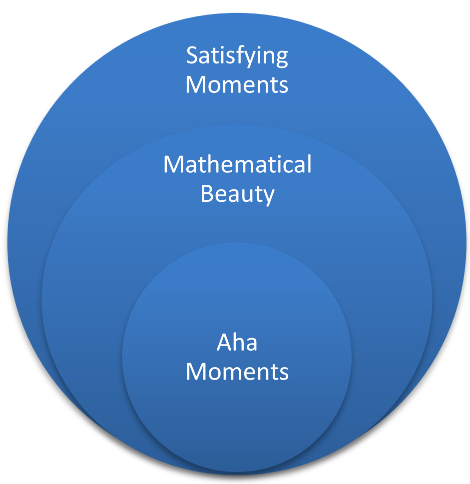


Figure 3.2. How satisfying moments relate to existing concepts regarding intense positive emotion.

Figure 3.2 shows how I conceptualize satisfying moments, as acting as a superset for other existing constructs in the literature regarding intense positive emotions. One concept that falls under the umbrella of satisfying moments is the idea of mathematical beauty (Hardy, 1940; Sinclair, 2006). Another example is the aha or eureka moment (Barnes, 2000; Liljedahl, 2004). I discuss these related ideas below.

Related Constructs

Mathematical beauty. There is a well-documented phenomenon of mathematicians writing and talking about beauty in mathematics (Hadamard, 1945; Hardy, 1940; Lockhart, 2002; Poincaré, 1952; Thomsen, 1973). Mathematicians say statements like "That is an elegant solution" or "This is a beautiful proof" when talking about mathematics they admire and talk about math as being comparable to art in certain ways. Zeki, Romaya, Benincasa, and Atiyah (2014) showed that when mathematicians experience mathematical beauty, this correlates with activity in the same part of the brain associated with enjoying art.

There are philosophical differences over whether mathematical beauty is an objective characteristic of the piece of mathematics or a projection from the observer (Sinclair, 2006, 2009). However, I take the approach of the latter and conceptualize mathematical beauty as an emotional response to mathematics.

In the same way it is difficult to define beauty, it is difficult to define mathematical beauty. G.H's Hardy's (1940) *A Mathematician's Apology* is one of the texts most associated with the idea of mathematical beauty. Hardy claimed that theorems that are beautiful tend to exhibit a triumvirate of inevitability, economy, and unexpectedness. Some commonly stated features of mathematical beauty include simplicity, brevity, inevitability, economy, enlightenment, understanding, and surprise, among others (Blåsjö, 2012; Cellucci, 2015; Hardy, 1940; Rota, 1997; Satyam, 2016; Sinclair, 2006).

Mathematical beauty is a driving force for doing mathematics, playing a crucial part of engaging in mathematical inquiry (Hardy, 1940; Poincaré, 1952). Sinclair (2004) identified three roles for beauty in doing mathematics: motivational, generative, and evaluative. Mathematical beauty reveals the values of mathematicians and the larger mathematical community.

Aha moments*.* An aha moment is an affective response to an unexpected idea or solution, which are cognitive events (Liljedahl, 2004). One of the most famous stories examples is of Archimedes sitting in a bath and realizing that displacement equals volume and leaping out yelling "Eureka!" For this reason, aha moments are sometimes called eureka moments as well. Aha moments are characterized by a sudden realization or insight. Mathematicians like to think of mathematical beauty as a moment of instantaneous enlightenment, like a lightbulb turning on (Rota, 1997). Hadamard (1945) talked about discovery as a flash of insight as well, also using the metaphor of light illuminating the darkness. There has been some work on aha moments (Mason, Burton, & Stacey, 1982), due to the hope that they may change attitudes and beliefs (Liljedahl, 2004).

Why create a new construct? Moments of mathematical can be relatively rare and aha moments even more so, which makes these phenomena very difficult to capture and study. In addition, based on my past work, I found that a good number of students did not respond well to the word "beauty" to describe math. It came across as an odd word to use, perhaps because of what all is brought to one's mind by the word "beauty" in everyday language. I would have needed a different term to use with them even if I had gone that route.

Satisfying moments are therefore an expanded version of mathematical beauty. It is also just a shorter way of referring to moments with intense positive emotions. The scope of this construct is kept broad intentionally, so that students may say they do indeed experience this and thus report more of them. This investigation of kinds of satisfying moments is therefore about the range of these moments which occur. Some may end up being instances of mathematical beauty or even aha moments.

Experiences with intense positive emotions provide motivation for students to continue doing mathematics and thus can be productive. Research is needed on how these experiences can provide intrinsic motivation for students to continue and value doing mathematics (McLeod, 1988).

CHAPTER 4: Method

In this chapter, I outline and justify the methods used to answer the research questions. I describe the study context of the transition to proof course, the participants, the sources of data, and the methods of data collection. I also provide a detailed description of the pilot study and how that informed the research. Data analysis will be discussed in the next chapter.

Study Context: Transition to Proof Course

The transition to proof course at this university was designed to ease the transition from calculus-based courses (e.g. Multivariable Calculus or Differential Equations, where the work was primarily computation and using formulas) to upper-level math courses that involved writing proofs. This course was required for undergraduates majoring and minoring in mathematics, unless they chose to enroll in an advanced linear algebra course, which then functioned as their transition to proof course. This course was a prerequisite for Linear Algebra, so a variety of STEM (science, technology, mathematics, and engineering) majors were enrolled in this course as well.

Content

The first half of the course focused on grammar, and the second half introduced students to basic concepts in real analysis, linear algebra, and number theory (see Table 4.1).[1] The course met for 80 minutes three days a week, for fifteen weeks.

[1] This weekly content was true at the time of data collection but has since changed.

Table 4.1: Schedule of Weekly Content for Transition to Proof Course

Week	Topics
1	Sets
2	Functions: injection, surjection, and bijection
3	Mathematical statements (negation, and, or) / Induction
4	Truth tables / Implication/ Contradiction/ Proof by contradiction
5	Converse, contrapositive / Proof by contrapositive
6	Conditional statements and quantifiers
7	Review and exam 1
8	Real analysis; open and closed / sequences and convergence
9	Linear algebra; vector space, linear functions
10	Linear algebra; vector space, linear functions
11	Number theory; division lemma, gcd
12	Number theory; modulus, equivalence relation
13	Review
14	Review and exam 2
15	Review for the final
	Final exam

Note. Description of content for each week of the Transition to Proof course. The first half was about proof grammar and techniques, and the second half of the course presented basic concepts from advanced mathematics students had not taken yet. Adapted with permission.

Course Design

This specific transition to proof course differed from a "typical" lecture mathematics course. The instructor (graduate student or faculty) lectured for roughly 120 minutes each week, with typically 1½ days devoted to lecture. For the rest of the class time, students worked on problems in groups of 3-4. A graduate or undergraduate teaching assistant also assisted the instructor with the group work portion of the class two days a week. In the past, the amount of lecture had generally decreased over the course of the semester, depending on the concepts.

Students were expected to read selected material from *How to Think Like a Mathematician* (Houston, 2009) and course-created supplementary documents before coming to class, in order to have a first exposure to the content. Online reading quizzes worth a minimal number of points provided the incentive for students to do this reading. Students also had access to the online forum *Piazza* where they could ask questions, and instructors, teaching assistants, and fellow students could answer through this system.

Homework was a central learning activity of the course. Homework was due every week, to be typed in LaTeX, a typesetting software commonly used in mathematics. Each homework typically had three types of questions: answer only, medium justification, and complete justification. The proportion of the three types of homework problems shifted over the semester, towards more complete justification full proofs. Students could also seek help on their homework from a math learning center (MLC) on campus.

Researcher Positionality: My Dual Role as Researcher and Teaching Assistant

My relationship with the transition to proof course was not that of an outside researcher. Thus, I describe my position relative to the course, because it influenced my access to the participants and the nature of the data collected.

I was a teaching assistant for the course in Fall 2016 and Spring 2017, the latter of which was the semester of data collection. I was in the classroom two out of the three 80-minute periods that the class met in order to help with group work. There were weekly course meetings for instructors and teaching assistants, contributing to my knowledge of the intentions behind course decisions. I was also a tutor at the MLC each week, where students of this course visited (including, sometimes, my own participants in this study), primarily for help with the homework for the course. Lastly, I had also observed the course periodically in the previous year, as a part of a separate research project. Thus, I had observed the nature of this course and how it had changed over time. The participants in the study were not my own students however.

This perspective influenced the study in the following positive ways. One, I had easier access to potential participants due to personally knowing all of the other instructors. Two, I was aware of what students had been taught so far in the course, which affected how I conduct my interviews with students and my interpretation of their work. Three, some students already knew me from the math learning center, so there was an added rapport; I could talk about course milestones and what was currently happening in the course with participants, e.g. commiserate over the last homework or exam.

My insider status with the course also had some limitations. There was the potential for students to see me as an "authority" regarding the class, because some students knew me first as a teaching assistant as opposed to a researcher. To prevent them from potentially asking me for help and answers, as they would a teaching assistant, I was upfront about my role in my interviews with them and told them I would have to decline helping them during the interview tasks, when it would interfere with the study. All the participants understood the different role I played when conducting the study, and it was not an issue.

Description of Instructors

Here I give brief descriptions of the two transition to proof instructors whose students I recruited for this study. Pseudonyms were chosen by the instructors.

Mr. X was an assistant professor in the mathematics department. He was the coordinator of the transition to proof course and developed much of its structure. At the time of data collection, he had taught the course for multiple semesters.

Ms. Frye was a graduate student in the mathematics department. At the time of data collection, this was her second semester teaching the transition to proof course.

Participants

The participants were N=11 undergraduate students taking a transition to proof mathematics course at a large Midwestern university (see Table 4.2). Their ages were from 18 and up. Twelve students were interviewed initially – there was one student who only completed the first of the four required interviews and so is not listed here.

Table 4.2: Background of Participants

Name	**Major(s)**	**Minor(s)**	**Year**	**Gender**	**Ethnicity**	**Instructor**
Amy	Actuarial Science	Entrepreneurship	2	F	--	Ms. Frye
Charlie	Computational Math	Computer Science	3	M	Chinese	Mr. X
Dustin	Statistics		2	M	White	Ms. Frye
Gabriella	Actuarial Science		1	F	--	Mr. X
Granger	Physics, Math		1	M	Caucasian	Mr. X
Joel	Statistics		2	M	White	Ms. Frye
Jordan	Math, Secondary Education	Chemistry	2	F	--	Ms. Frye
Leonhard	Math		1	M	White	Ms. Frye
Stephanie	Actuarial Science		2	F	White	Mr. X
Shelby	Statistics		2	F	White	Mr. X
Timothy	Math		3	M	White	Ms. Frye

Note. Pseudonyms are used for participants and instructors. Participants self-identified ethnicity using their own terms. The -- notation denotes a participant opted out of self-identifying their ethnicity.

Recruitment and Selection of Participants

I recruited participants using the following process. Recruitment was done in person. I asked the instructors of two sections of the course, Ms. Frye and Mr. X, if I could visit their class in order to talk to students about the research and ask for volunteers. Ms. Frye and Mr. X's classes had enrollments of 22 and 21, respectively. Across both classes, I selected students to vary along the following parameters: instructor, math major or not, and gender. First, I chose an equal number of participants from each section of the course, in order to account for how different instruction can influence students' proving. Second, within each instructor's students, I chose half math majors and half other. Third, I picked half the sample to be female, the other half male. The demographics of this specific course tended to be 2/3 male and 1/3 female. I chose to not mimic the gender distribution of the course, due to existing research evidence about interaction between gender and affect, especially in regard to negative emotions, so gender was an important variable. When choices still remained, I chose participants based on ethnicity, to align with representation in the course, and finally on their schedule availability. Participants self-identified their ethnicity, via a blank space on the participant form.

20 students volunteered for the study; of these, I selected an initial 12 participants according to the guidelines above. Some participants did not reply, so they were replaced by additional participants, adhering to the above rules when possible. [2]

[2] A question that may arise for the reader: Why were there so few non-white participants? The fact that most of the participants were white stood out. A little background: 7 of the 20 volunteers for the study self-identified as an ethnicity other than white. Out of these, I selected 5 for the study: a black female, black male, Asian female, Middle Eastern male, Hispanic female. When I contacted them to follow up, only 1 replied – and she could not continue the study after 1 interview. It seemed odd for a number of non-white students to sign up and then very few participate in the study itself. This sample size is small; no real

Descriptions of Participants

Here I give brief descriptions of the eleven participants who completed the entire interview series. These are portraits of the students as I came to know them over multiple interactions over time, not first impressions. As such, I include some interesting details specific to them; these profiles are not meant to be complete. I include this section in order to humanize these participants, as a reminder that these are all individuals with different backgrounds, personalities, and hopes for their future. These details do color the data, especially in examining affective issues.

Amy was a white female sophomore majoring in Actuarial Science and minoring in entrepreneurship. Amy said she has a love/hate relationship with math; she does not like math when she first starts a problem but then loves it when she is done. She especially liked the competitive and challenging aspects of mathematics, e.g., doing a hard problem that someone says cannot be done. In terms of career goals, her goal was to be an actuary.

Charlie was an Asian male (an international student from China) junior majoring in computational math and science and considering a minor in computer science. He had a penchant for problems he could do in his head and also talked about his thought process using metaphors throughout the interviews.

(cont'd) conclusions can be made from it. However, investigating whether there are structural factors that lead to non-participation by students with non-white backgrounds is worthy of future research.

Dustin was a white male sophomore statistics major, minoring in actuarial science. He wanted to do actuarial science but was majoring in statistics, to give himself more career options. He claimed he did not do well on timed tests. He expressed that he liked using examples as models for proofs and that he understood math when someone else explained it to him. He found that talking about math with other people helped him work.

Granger was a white male freshman math and physics major. He wanted to be a professor or an industrial mathematician. He explained that his class had positioned him as one of the “smart ones.” During interviews, he wrote very quickly and expressed that he felt his brain was usually way ahead of whatever he’s writing. Granger felt he was not emotional in general, let alone when doing mathematics.

Gabriella was a white female freshman majoring in actuarial science. She was good friends with Stephanie; they often worked on homework together. In the beginning, she said she would oftentimes second guess her answers, but she stopped doing this as the semester went on. She said she preferred calculus-based courses to proving; she just wanted to get through this class, as a requirement for her major.

Joel was a white male sophomore majoring in statistics and considering a minor in math. He talked about how math used to come easy to him in high school. In the beginning, he said he was terrified about the course, but as time went on, he found the material interesting. He did however talk about enjoying “grinding out” problems, where one can just do them as opposed to having to figure things out.

Jordan was a white female sophomore math major with a minor in chemistry. She wanted to be a secondary math teacher. Jordan started the semester off well but seemed to be demoralized by the class as time went on. She felt that she had put in a lot of effort and time into homework yet still received low homework scores and that she did not understand concepts as time went on.

Leonhard was a white male freshman majoring in math. He wanted to be either a high school teacher or a mathematician for aerospace engineering. He had just transferred to school this semester. Leonhard had lots of thoughts about mathematics and used metaphors to explain his thinking often.

Shelby was a white female sophomore majoring in statistics. She had taken some math classes at nearby community colleges, for the smaller class size. She liked having steps in mathematics. She also expressed that talking out loud to people helped her when she was stuck working on mathematics and that she enjoyed working with people.

Stephanie was a white female sophomore majoring in actuarial science. She was good friends with Gabriella, and they would work together on homework. She wanted to work in insurance. She came across as practical, not swayed by emotions. She said her biggest struggle was in understanding what the problem was asking for. She put stock in high performance and did get good grades but by end of semester, she was worn down.

Timothy was a white male junior majoring in mathematics. He wanted to work in informational technology (IT) afterwards. He found proving to be fun but felt he needed time to learn things, for concepts and definitions to sink in. He was especially good at talking his thoughts out loud.

Data Sources

The data were a series of four semi-structured interviews across the semester with each participant. Each interview consisted of two halves: the first half was organized around two proof construction tasks, and the second half was about satisfying moments with stimuli tasks. Figure 4.1 shows the different data sources by interview, as the second half of interviews 2-4 was different from that of interview 1.

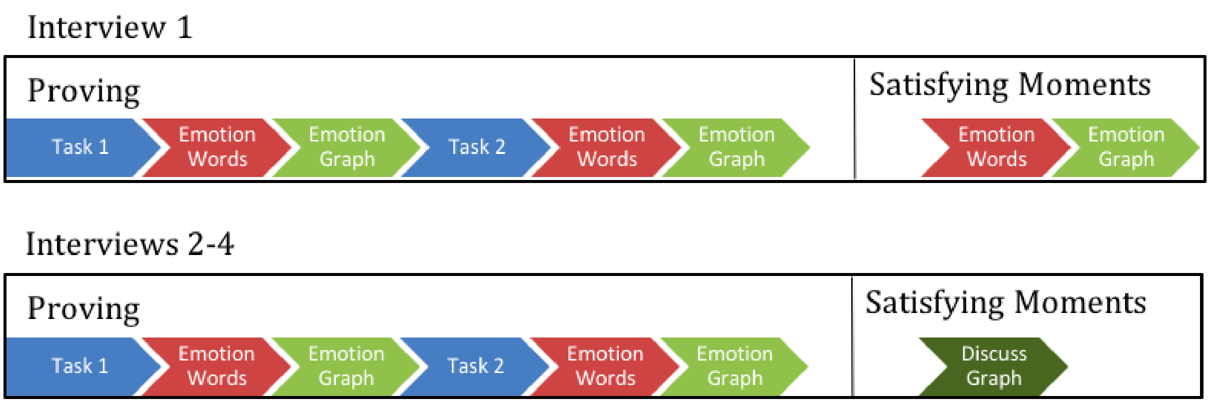


Figure 4.1. Representation of data sources (by color) within each interview

In capturing development of proving, the times at which I took "snapshots" of students' reasoning was important. As a reminder, the course was designed so that the first half is about general proof structures and the second half focused on content (real analysis, linear algebra, and number theory). I interviewed participants at these four times: middle of the proof structures section, end of the proof structures section, middle of the content section, and end of the content section. Each of the four rounds of interviews were done over a two-week span. In the following sections, I describe the design of each part of the interview in detail, including instruments and stimuli tasks.

First Half of Interview: Proving

During the first part of each interview, participants worked for no more than 15-20 minutes on each of two proof tasks. I chose to give participants two tasks, rather than only one, so that they had more than one opportunity to show their thinking at this current point of the class, in case they struggled majorly with the particulars of one task. The idea was to give students two "chances" per interview in case specifics of one tasks threw them off. I told participants they had 15 minutes but gave them a maximum of 15-20 minutes to work on each task, in order to give them enough time to showcase their thought process and attempt to overcome stuck points when any occurred. I was interested in their thought process, as opposed to analyzing their final written product. If a student was still working at the 15 minute mark, I oftentimes let them work for a minute or two until they finished their current train of thought.

Proof construction tasks: Selection. The selection of tasks for the proving section of the interview was vital. Because students' written and verbal responses to the tasks were how I chose to measure their proving at a given point in time, the nature of the statements and their possible solutions largely determined what the students did. Especially in using tasks to study students' *development*, a coherent rationale behind selection of tasks was necessary. Table 4.3 lists the proof tasks (full versions given in Appendix A).

Table 4.3: Proof Construction Tasks by Interview

Interview 1	Statement
Task 1	Suppose x and y are integers. If $x^2 - y^2$ is odd, then x and y do not have the same <u>parity</u>.
Task 2	Prove the following statement: If a and b are strictly positive real numbers, then $(a+b)^3$ never equals $a^3 + b^3$.
Interview 2	
Task 1	Prove the following statement: If x and y are <u>consecutive</u> integers, then xy is even.
Task 2	Prove the following statement: If a, b, and c are non-zero integers such that a divides b and a divides c, then a divides $(mb + nc)$, for any integers m and n.

Note. Abbreviated versions of each of the proof construction tasks. Underlined words were new definitions, which were defined for the participant; full versions of tasks given in Appendix A.

Table 4.3 (cont'd)

Interview 3	
Task 1	Prove the following statement: Suppose x, y, z are positive integers. If x, y, and z are a Pythagorean triple, then one number is even or all three numbers are even.
Task 2	Prove the following statement **without using induction**: If n is an odd natural number, then $n^2 - 1$ is divisible by 8.
Interview 4	
Task 1	Prove the following statement: If a and b are odd perfect squares, then their sum a + b is never equal to a perfect square.
Task 2	Prove the following statement: If x, y are positive real numbers and $x \neq y$, then $\frac{x}{y} + \frac{y}{x} > 2$.

The following criteria were used to select tasks. First, the goal of the task was to function as an assessment for the student at a certain point in time. Because students' progress was likely heavily influenced by instruction, it made sense to pick tasks that were similar in nature to questions they would encounter in the course, generally around the same time period (weeks) but before students actually encountered them. For these reasons, the tasks were taken from past homework assignments from previous semesters of the course, specifically Spring and Fall 2016.

Second, tasks with multiple possible solution paths, not just one, were chosen when possible. For example, statements that could only be proven easily using a proof by contradiction were excluded; however, statements that could be proven by either contrapositive or contradiction were still viable because of the choice in technique.

Third, all tasks were from one content area, basic number theory. The goal was for the tasks to not be heavily dependent on content knowledge nor a singular specific proof technique (e.g. induction). Because I hoped to make claims about the students' problem solving abilities, I wanted to minimize the effect of a lack of content understanding. For example, proofs regarding analysis concepts were excluded because a student's difficulties could be due to struggles in understanding analysis definitions or concepts and not necessarily in proving. There is a danger however in making content-free claims about students' proving (Dawkins & Karunakaran, 2016), so my claims about student proving may be specific to this content area. Even still, I would argue that basic number theory, e.g. properties of even and odd numbers, is a more broadly accessible content area than analysis, so more students can at least start the task.

The first task of each interview was designed to introduce a new definition, a novel situation with new information to deal with it. The second task was designed to elicit stuck points, where students thought they knew what to do but it would not work. I searched for tasks that looked like they would be routine but were in fact not. Interview 3 – Task 2 is an example of a task that was especially successful at what I described above. Interview 1 – Task 2 was less so, but because it was the first interview, this did not affect the analysis much. Tasks that are novel and/or have a stuck point built in are problems, as defined in the literature.

To select potential tasks, I looked through all homework sets from the previous two semesters and compiled questions that best satisfied these criteria. I used homework questions that students were likely to have not seen by the time of the interview, i.e., they would run into a homework question of that type later in the course. When I could not find suitable questions from homework, I found some using other textbooks or online resources or made my own.

Think-aloud. In order to capture their strategies and reasons for using certain strategies, I used a think-aloud protocol (Ericsson & Simon, 1980, 1981; Schoenfeld, 1985a), where participants voice their thoughts aloud about a task, either in real time or shortly after the task is complete. Previous work on how students problem-solve and prove has largely used think-aloud methods as a proxy for accessing cognition (Schoenfeld, 1992; Weber & Alcock, 2004).

Over the years, researchers have considered and examined the validity of using verbal data to infer about the thought process (Ericsson & Simon, 1980, 1993; Schoenfeld, 1985a;). In other words, to what extent does asking a subject to verbalize their thought process affect their thought process? This issue is called *reactivity* (Leighton, 2009) and affects certain kinds of experimental set-ups and questions (Schoenfeld, 1985a). Certain experimental variables can impact the data produced, such as the number of people being interviewed, the degree of interviewer intervention, and the environment under which task is being given (Schoenfeld, 1985a).

A major issue then in administering a think-aloud is the level of interviewer intervention: more vs. less and the character of it. More intervention can mean more verbalizations and thus evidence, especially of metacognitive behavior. However, asking students to reflect on their problem solving process in the moment can affect their performance (Ericsson & Simon, 1980). In addition, asking "why" questions during a task can dramatically change one's behavior (Schoenfeld, 1985a). It safer then to ask "what" questions during a performance, such as asking them to identify what they just did. Ericsson and Simon (1980) have argued that asking subjects to verbalize their thoughts but not asking for any explanation of said thoughts does not affect a person's performance. Other non-intervening moves during task performance include "I haven't heard you say much in the past couple minutes. Are you still working on the problem?" and answering specific student questions.

Based on the affordance and constraints of asking probing questions, I chose to minimize interviewer intervention during task performance. This was because my phenomenon of interest was the proving/problem solving *process* itself so keeping the process intact from a validity standpoint as much as possible was of the utmost importance. This was especially the case since my phenomenon of interest was what students do when stuck, and there was a high chance that talking would get them unstuck. In other words, I did not want students' verbalizing to affect their proving process.

I asked students right before they started a task to verbalize their thoughts out loud, as long as they felt it did not interfere with their thought process. If the student had been silent for a few minutes, I would sometimes remind them to say what they were thinking. Otherwise, I remained silent. Over time, I developed a sense for which participants were comfortable talking while working and which participants were less so, and I held back on nagging the latter group. This is one place where familiarity with the individual was helpful for minimizing the interference on each *individual's* performance, even if it meant I had to act slightly different across participants.

I then debriefed with the student immediately after they said they were done working. During this debrief, I asked all probing questions: to explain their thought process and any "why" questions, such as "Why did you do X?" Subjects can talk about their thought processes about a given task if asked immediately after task completion (Ericsson & Simon, 1981). I tried to ask probing questions in a way that did not let on if their solution attempt was correct or not, since there was more data to be collected regarding their solution after the debrief. Asking students to think-aloud but not pushing them to do so and then asking probing questions immediately after they had finished their work optimized the benefits and pitfalls of think-alouds for studying proving as a phenomenon.

Ensuring students' comfort during proof tasks. It was a major goal that students felt comfortable throughout the interview. This was especially important during the proof tasks section of the interview, where I video recorded each student while they worked on difficult tasks. It can be difficult to work with someone watching over you, let alone the fact that participants feeling pressure would affect the data. Most of my behavior throughout the interview was centered around making them feel comfortable, for instance by having students sit in my chair at the center of the desk rather than relegating them to a small chair off to the side.

I took the following steps to decrease the likelihood of students' discomfort. One, I set the video camera as far away from the participant in the room as I could while still being able to capture their written work. I also stood far away from the student while they were working on one of the proof tasks. In times when students became very frustrated, I left the room momentarily while keeping the camera rolling in the hopes that my temporary absence would decrease the pressure they felt in that moment. Students adapted to the experimental set-up very well – none of them glanced back at the camera out of self-awareness or observable self-consciousness.

Collected but not analyzed: Affective data on students' proving. In this section, I describe some of the data that was collected to capture students' emotions about problem solving. This data is not analyzed in detail in this dissertation, except for one instance in the development of students' proving results chapter.

After each proof task's debrief, I asked participants to picks emotion words they went through while working on the task and to draw a graph of their emotions. The purpose of these were to help participants describe their emotional responses to their proof tasks, as a way to capture affective data about cognition (their proof work).

Emotion words. Participants were shown cards with an affective word written on each (as shown in Figure 4.2). Five negative-positive "pairs" and one neutral emotion word were chosen with the intention of capturing the range of possible emotions while problem solving, based on literature review. Participants were free to choose other or none of these words; these given words were only meant to provide a base to start with.

annoyed	curious	
disappointed	surprised	
sad	joyful	indifferent
frustrated	satisfying	
ashamed	proud	

Figure 4.2. Physical arrangement of the 11 affective words for the Emotion Word Task

The eleven index cards with one emotion word on each were laid out on a table in front of the participant, as shown in Figure 4.2. Participants were asked to select which emotions on the cards they experienced and to say why. By putting the emotion words on tangible cards, participants could point to or handle each one physically. They oftentimes put the cards with emotions they felt in temporal order. As the interviewer, I circled the words they chose on a pre-printed piece of paper and then asked if there were any other emotion words the participant would have picked that were not present. Their emotion words expressed the *emotions* that the participant experienced, and their reasons for picking each word helped me to identify the *conditions* that led to that emotion.

Emotion graphs (adapted from McLeod, Craviotto, & Ortega (1990) and Smith, Levin, Bae, Satyam, & Voogt (2017)). The participant was given a blank graph and asked to chart their emotions during the entirety of the proof construction task. The graph allowed for a temporal look at the ups and downs in emotion over the course of their solution attempt. Graphing emotions is a technique that can be used to describe variations in students' emotional responses while solving a problem (McLeod, Craviotto, & Ortega, 1990). Participants were also asked to mark on the X-axis and/or annotate their graph with short captions at the points at which their feelings changed.

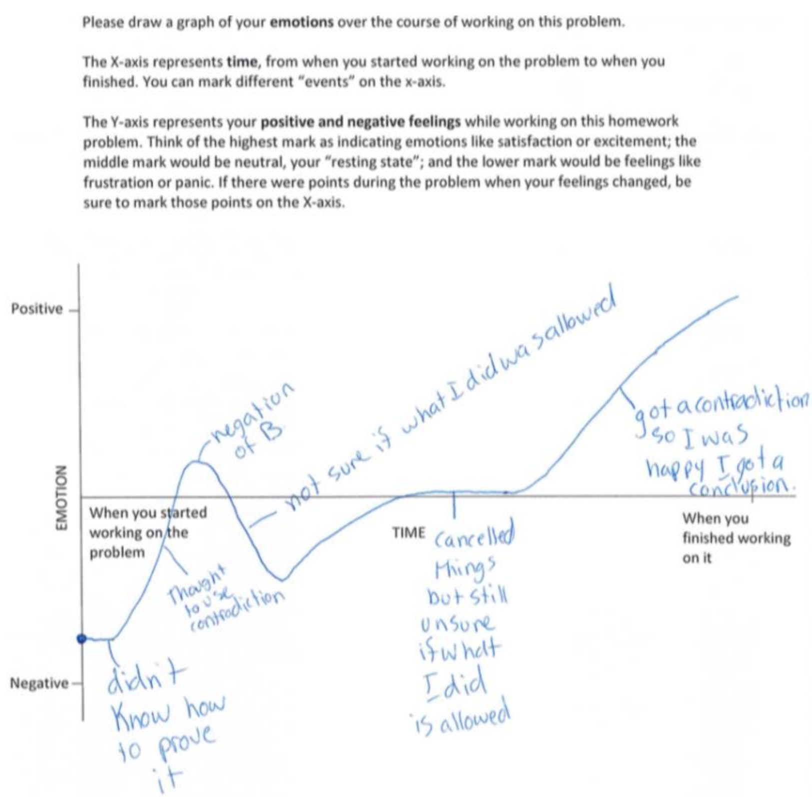


Figure 4.3. Example of an emotion graph for a proof task

Second Half of Interview: Satisfying Moments

The purpose of the second half of the interview was to capture data for my second research question, about the nature of satisfying moments related to the transition to proof. The data in this half of the interview consisted of questions about satisfying moments, discussion over emotion graph tasks and then the emotion word task (only on interview 1). The interview protocol in Appendix B lists the questions that were asked. The goal of these questions was for students to describe in full detail any satisfying moments the students had encountered in relation to the course, whether through homework problems or in class.

Self-report of satisfying moments. Self-report was an appropriate method for capturing this data because it revealed the subject's *perception* of their own satisfying moment, which was most important. For example, if a person truthfully perceived an experience as satisfying, then I could argue that this experience was satisfying to that person, even if an outside observer watching the entire experience unfold did not see it as satisfying. In other words, the label of "satisfying" is determined by the subject's emotional response, which is internal and personal.

The instructions given to students naturally then influence what they report. In designing the interview questions, I therefore introduced the idea of a satisfying moment with few constraints, so that students would report back according to however they defined it for themselves. However, in the first interview, after introducing the idea of a satisfying moment, I asked follow up questions about situations that could be satisfying, e.g., problems that feel rewarding, flashes of understanding/insight. These questions were meant to probe, to help students if they could not think of satisfying moments on their own, but also to try to capture some related affective concepts, such as aha moments. It is possible though that these follow-up questions may have influenced what students reported in the future as satisfying. As we shall see, however, the fact that students still talked about performance and that flashes of insight were still relatively rare suggests these follow up questions did not affect the data unduly.

Emotion graphs as recall. It is possible that students may not remember satisfying moments, without some kind of record. In the first interview, I asked participants in detail about satisfying moments in relation to their transition to proof course. For the most salient experience, I asked them to draw an emotion graph. This emotion graph had small variations in wording from the emotion graph for the proof task (see Figure 4.4).

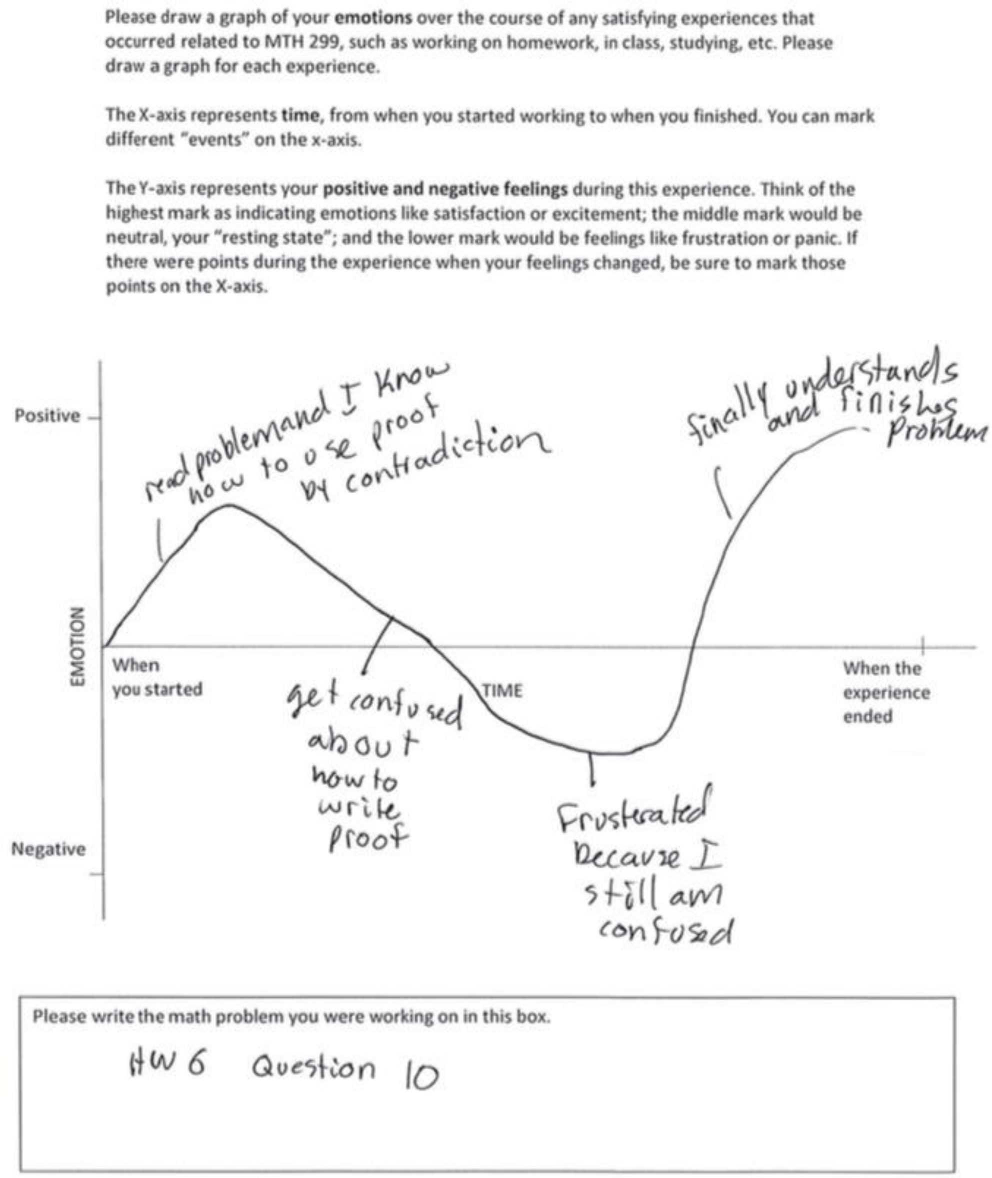


Figure 4.4. One of Stephanie's emotion graphs for a satisfying moment in Interview 2

After the first interview, I gave participants 2-4 blank emotion graphs to take home and asked them to fill it out (i.e. draw a graph) whenever they had a satisfying moment before the next interview. In interviews 2-4, participants came to the interview with already filled out emotion graphs and were ready to talk about satisfying moments they had experienced outside the interview.

The purpose of the emotion graphs was to (a) have a record of a satisfying moment presumably in real time or at least not too long after of a satisfying moment and (b) serve as a stimulus for discussing the satisfying moment during the interview. On the first interview, students were asked to pick emotion words from the index cards for their most salient satisfying moment, which I chose in real time based on the subject's responses to the previous questions about satisfying moments. The selection of the experience went as follows: if the subject discussed only one satisfying moment in the interview, I used that experience. If the subject talked about multiple satisfying moments, I picked the most intense one or the one they talked about the most. If the subject did not talk about any satisfying moments, then I didn't administer the word selection and emotion graph tasks.

On interviews 2-4, I did not ask them to pick out emotion words for their satisfying moments, in order to (a) avoid task fatigue, as this would be their third time choosing words during the interview, but also (b) students were now comfortable using emotion words when talking about their experience. The emotion word task was no longer needed then, as an artificial stimulus for talking about satisfying moments.

Interview notes. Interview notes were taken on paper during interview. They were then recorded digitally with more observations as soon as possible after the interview. I also took notes about things to ask them next time, to keep continuity across interviews. These interview notes became a source of data for some of the analyses.

Pilot Study

I conducted pilot interviews with three participants in order to test the research design and instruments the semester before the real data collection occurred. In this version of the study, the research focus was on asking students about homework problems as a central place of learning in the class and as a source for satisfying moments. The interview protocol asked students to talk about problems from the homework set they had completed that week, specifically a problem I pre-chose for its non-routine characteristics and a problem they personally found challenging.

The goal was to conduct three interviews with each of the participants, to simulate doing multiple interviews over the semester to study development. All three were recruited from one instructor's section of the transition to proof course, to eliminate possible variation due to differences in instruction. Interviews were conducted soon after they passed in their homework, to account for their memory of the proving process receding over time. Homework was due on Wednesdays, so participants were interviewed anywhere from Wednesday to the following Monday.

In the end, two of the participants conducted the set of three interviews and the other participant conducted only one. Interviews occurred in weeks 8, 10, and 14 of the class. Pilot participants were paid $20 per interview as compensation with a $15 bonus for completing the full series of 3 interviews as an incentive.

First and Second Rounds of Pilot Data Collection

Based on the first interview, I found that when asked about satisfying moments on the most recent homework, students pointed to a problem almost instantly, i.e., they could point to a specific experience. They picked a variety of words from my selection to describe the experience, sometimes adding one more of their own, and drew detailed emotion graphs of their experience. There was some confusion over what the x-axis, or "zero" emotion, represented. At my request, they added annotations for the ups and downs in the emotion graph, sometimes adding more when discussing the graph with me. All in all, the satisfaction portion of the interview went well.

A number of issues emerged, however, in the proving section of the interview. I found that it was difficult for students to discuss their thought process about problems after they had been completed, even when the interview was done on the same day as the homework had been due. Participants had a tendency to *describe* their answers quickly, with quick comments at the start about initial strategies that proved fruitless. After two sets of two interviews were completed, I compiled a full list of challenges that appeared.

Methodological issues. First, there was poor quality of participant's discussion of the process of developing their proofs, likely because I was examining the process after the fact. This lead to discussing proof as a finished product, not process. Second, students often did not finish or do some of the homework problems, leading to loss of comparison between participants on pre-chosen problems. Third, students were able to get outside help on homework (from professor, teaching assistant, tutoring center, other students, online, etc.), so their answer was not necessarily a representation of their own thinking.

Logistical issues. The tight interview window (in order to mitigate memory loss) led to a lower probability of accomplishing the full set of interviews per participant and a smaller potential number of participants due to time constraints on the interviewer's part.

These problems led to a redesign of the proving section of the interview, where I asked participants to work on a proof during the interview. Instead of completing the full set of three interviews with a design that had deep flaws, I chose to implement my revamped interview protocol for the third interview.

Third Round of Pilot Data Collection: Re-Design of Proving Section.

The purpose of this round of interviews was to test the new proving section of the interview. N=2 participants completed this interview. The priority was testing whether the verbal data produced by participants working on a proof during the interview itself, in real time, would be of better quality in answering my research questions than in the previous pilot round design. Inferring students' thought processes and strategies was much easier using this method, because (a) I could see how they approached the problem on paper, (b) they would talk aloud as they thought, and (c) I could ask clarifying questions in the moment.

Another goal of this third interview was to test out the proof tasks themselves. In selecting tasks, I chose two questions from the prior semester's (relative to the pilot data collection) course homework from around this same time in the course schedule. One task turned out to be very similar to what was done in class and thus was done relatively quickly and without impasses for the students. This task was therefore not very illuminating in terms of observing students' proving and what they did when stuck, so it was changed for the actual data collection. The other task was more of a *problem*, as that term is characterized by Schoenfeld (1992), in that there were times where participants were momentarily stuck. Both pilot participants correctly completed both proofs in the end however.

Additional Changes from Data Collection

Another major change that came out of this pilot data collection was deciding to video record participants' work. For the pilot work, I only audio recorded the interviews, but I found that participants commonly pointed to their work and homework problems while talking about them, which was lost in audio. Video of their work and hands provided another source of data validation, especially when any doubt arose from the audio.

CHAPTER 5: Data Analysis

In this chapter, I describe how I analyzed the data. Recall that my research questions were as follows:

1) How does undergraduate students' proving develop over the duration of a transition to proof class?
2) What kinds of satisfying moments do undergraduate students have during the transition to proof?

The data analysis is described here in a stand-alone chapter because a good deal of work went into deciding how to analyze this data. My two phenomena of interest were quite different, but I faced similar difficulties in addressing them. One phenomenon, satisfying moments, was a construct I conceptualized myself, so no ready-made analytical frameworks existed. The other phenomenon, proving, was backed by research especially in thinking about proving as problem solving, yet analytical frameworks that served my purpose were difficult to find. In doing this work, I consider the data analysis itself and the challenges I ran into to be a major finding in and of themselves, which is typical for qualitative work. I detail my journey through these challenges here.

I used qualitative methods, because I sought to describe and understand how the phenomena of proving and satisfying moments occurred. The Data Use Matrix (see Table 5.1) summarizes how various data from the interviews was used to answer the research questions and the analyses that were done.

Table 5.1: Data Use Matrix

Research Question	**Method**	**Data Used**	**Analysis**
1. How does undergraduate students' proving develop over the duration of a transition to proof class?			
1a. What problem solving strategies do students use in their attempted solution when stuck?	*Characterize* proving at a snapshot in time	Proof Tasks (2) -Written work -Think-aloud -Debrief Triangulation: Interview -About current approach to proofs	Look at tasks on which a student becomes stuck. Record their strategies (proof-specific intentions) in response to being stuck
1b. How do students' use of problem solving strategies when stuck change over time?	*Compare* the snapshots over time	Triangulation: Interview -Question about reflecting on change over semester (only interviews 2-4)	Look for change over a student's strategies across tasks
2. What kinds of satisfying moments do undergraduate students have during the transition to proof?	*Identify* moments *Describe* them *Categorize* them	Interview -Questions about satisfying moments Triangulation: -Emotion Words -Emotions Graphs	Bottom-up generation of codes. Add more codes from literature. Apply coding scheme & create new codes as needed (modified open coding). Note: The coding scheme itself answers this RQ.

Note. The Data Use Matrix summarizes which pieces of data were used to answer each research question (RQ) and their associated analyses.

Research Question 1: Development of Students' Proving

In this section, I describe the process by which I analyzed my data to answer my first research question: *How does undergraduate students' proving develop over the duration* *of a transition to proof class?* In general, the data collected in the proving section of the interview was used to answer my first research question.

My goal was to (a) collect "snapshots" of a student's proving at various points in time and then (b) compare how these "snapshots" changed. Because there were two stages involved in the way I viewed development, I split this research question into two sub-questions for the sake of describing the analysis:

(1a) What problem solving strategies do students use in their attempted solution when stuck?

(1b) How do students' use of problem solving strategies change over time when stuck?

The results chapter for development in students' proving addresses the original research question, not split up.

Research Question 1a

The purpose of research sub-question 1a was to characterize a student's proving at a certain point in time. All of the data generated during the first half of the interview was used here (see Appendix B): students' written work, their verbalizations during the think-aloud, and their responses to questions about their reasoning afterwards. This question was also used when possible: *How would you say you currently approach proofs right now?* The purpose of this question was to capture the student's perceptions of what their "typical" approach to proving was at that point in time, as a form of triangulation, albeit still a perception.

Research Question 1b

The purpose of research sub-question 1b was to compare the snapshots created from research sub-question 1a. This question from the interview was also used: *How do you feel your ability to write proofs has changed since the last time we met?* The goal of this question was to have students reflect on how they had developed since the last meeting and see how their sense compared to what was revealed by the tasks. Again, the intention was to capture the students' perception of how they thought their work had changed, as a complement to what was observed as the researcher.

Searching for Usable Analytic Frameworks

In studying development, I needed a way to characterize proving at a snapshot in time (RQ 1a) and compare these snapshots across time (RQ 1b). My initial plan for capturing these snapshots of proving was to use an existing problem solving framework, such as Carlson & Bloom's (2005) multidimensional problem solving framework. Because I have argued proving to be a subset of problem solving, and problems solving as a phenomenon was backed by copious research, it made sense to use existing frameworks for problem solving if they were appropriate for the data. I originally wanted to use Carlson & Bloom for getting these snapshots, because it would provide a thorough way to code all the behaviors that appear in a problem solving attempt.

Why existing problem solving frameworks proved problematic. I soon ran into two problems. One, there were frameworks to identify what *phase* of problem solving a student was in at various times in a task, but characterizing their overall problem solving process was harder. Two, I wanted to be able to identify students' strategies specific to proving, and general problem solving frameworks would not do that because of their generality naturally. Any problem solving framework would not pick up proof intricacies, and something would be lost by staying in the general problem solving analytical frame. What I really needed was a proof-specific framework. This is not a new issue; Savic (2012) has called for the need for a proof framework for conducting research on proving.

Analytical Framework: Looking at Students' Intentions When Stuck

Instead, I looked at what students did when stuck and how that changed, with close attention to strategy and monitoring and judgement components of problem solving (Schoenfeld, 1992). Stuckness is an aspect of problem solving, not all, but I argue that without being stuck, a person is not truly in "problem solving land." Thus, if we wish to tap into authentic proving, when a person is in a state of uncertainty about how to proceed, looking at what a person does when stuck is key. I operationalize what it means to be "stuck" below.

Operationalizing stuckness. Operationalizing what it meant for a person to be stuck on a problem was tricky because it required finding some observable behaviors to serve as indicators of a person's internal mental state. I conceptualized being stuck as when a person (1) realizes there is an issue that needs to be resolved and (2) are not sure what to do. These two criteria had to be present. A key insight into telling when someone was stuck was *hesitation* over what to do next. Savic (2012) differentiated between an individual facing an impasse vs. changing directions in one's proof attempt, and this difference is based on hesitation. "Stuckness" (I will often use this term despite it not being a word for the sake of simplicity) can manifest itself through silence (via audio) and through body language (via video). In my data analysis, I chose to operationalize being *stuck* as no written or verbal activity for at least 15 seconds. Body language instead became more important in telling whether a person was stuck, justifying the need to collect video. Body language behaviors which suggested a person was stuck on a proof construction task are reported later.

Analysis Process

To analyze the data, I watched the video recordings of a select number of participants' attempts on all eight tasks. I watched for points where they became stuck (no written or verbal activity for over 15 seconds). I recorded what behaviors indicated they were stuck, as judging whether someone is stuck can be difficult. When this happened, I recorded (a) my observable evidence that they were stuck, (b) why they were stuck, based off their think-aloud, the later debrief, or my own inferences, (c) actions they took (as observable on screen, on paper data, or verbally spoken during think-aloud or explained in debrief later), and (d) the intention or strategy I could infer from the action. After this, I looked over the strategies the students enacted when they were stuck and looked for patterns of change. I only used tasks where students became stuck, i.e. problems, unless noted otherwise.

The notion of actions and intentions while proving came from Karunakaran (2014), as an analytical tool. It was sometimes difficult to infer their strategy. In the best case, students stated their strategy out loud during think-aloud or talked about it in debrief. In the worst case, I had to infer their strategy myself from little to no observable data.

Difficulties in Analysis

Issues with tasks. Some of my interview tasks had unexpected pitfalls. For example, in Interview 3 - Task 1, many students took the negation of the conclusion ("one number is even or all three numbers are even") in a procedural way that led to a statement that did not make sense: "one number is odd and all three numbers are odd." This led to students having an incorrect proof, but reasoning from there on could be high- or low-quality problem solving, so it was not particularly relevant to this analysis. If anything, it provided a point of uncertainty to students which allowed for more insight into what students do when unsure.

In Interview 4 - Task 1, there was some ambiguity over what exactly was odd in the assumption: if a, b were odd or a^2, b^2 were odd. However, many students did ask for clarification – this being the last interview over a semester suggests they have been more comfortable asking me question – and it did not affect the final product. Because this research is about students' processes when stuck, tasks that accidentally cause confusion or ambiguity may in fact work in our benefit.

Same or different stuck points? One issue that had to be resolved was whether to group together multiple stuck points, if and when they really addressed the same challenge. Interview 2 - Task 1 with Timothy is an example of this: He became stuck, took a step, became stuck again, and took another step. His progress throughout this time had a stuttered nature to it, with lots of stops and starts. In cases like these, I considered all of these actions to be in response to one stuck point, as his strategies with each step were all responses to the same stuckness. I counted it therefore only as one stuck point.

Operationalizing strategy. Another difficulty was identifying the "size" of what counted as a student's strategy, whether to analyze local or more global strategies. For example, consider the different "sizes" of the following strategies:

Try to solve the problem

Try a different method

Try a different proof technique

Switch to proof by contrapositive

These are all strategies, from the most concretely actionable (local) to the most overarching (global). *Switch to proof by contrapositive* is the most concrete strategy, but their goal really is to *try to solve the problem*. However, *try to solve the problem* was not helpful for shedding any light on my research question. Looking at the in-between levels, try a different method is more general than proof technique.

My answer to the issue then was to use *the most local strategy that was proof-specific but not task-specific*. In the chain of strategies above, "try a different proof technique" is the smallest-sized intention specific to proving but not tied to the specifics of that task and thus may happen for other tasks. "Try a different proof technique" is also better than "try a different method" because the first is specific to proving whereas the latter is not. Therefore, in this example above, "try a different proof technique" would be the strategy I record.

Research Question 2: Kinds of Satisfying Moments

Here I describe how I answered my second research question: *What kinds of satisfying moments do undergraduate students have during the transition to proof work?*

Overview of Constructs and Data Analysis for Satisfying Moments

As a reminder, I operationalized the key terms in this research question as follows. By *satisfying moment*, I mean an experience characterized by significantly positive emotions, such as an aha moment. By *kinds* of experiences, I mean experiences that share some set of similar characteristics. Under this conceptualization of *kinds* and considering the limited existing research about experiences of this nature, grounded theory (Glaser &

Strauss, 1967) methods were appropriate for identifying key themes in participants' experiences. I therefore report here my method for identifying kinds, as well as the kinds themselves. In other words, the *process* by which I identified kinds of satisfying moments was as much a result as the identification of the satisfying moments themselves.

I answered this research question using three steps. First, I *identified* sections of the audio interview where students discussed satisfying moments: questions 8-19. I also looked at the Emotion Word and Emotion Graph tasks themselves and discussion around them, as needed. Second, I *described* these moments, according to participants' narratives. The word selection and emotion graph tasks helped me in describing how the situation unfolded as well, as triangulation for the audio. Lastly, I *categorized* all these different moments, as a way to create different "kinds" of satisfying moments. This categorization process was done bottom-up, using techniques from grounded theory.

Assumptions

An overarching assumption that guides this work is to stick close to participants' sense of their own experiences and what they say is satisfying. There was a choice: Do I report what participants are aware of and claim to be satisfying, or is it better to code what I, as researcher, saw as evidence of a satisfying moment *that they were not consciously aware of*? It is tempting to do the latter, to uncover things that participants themselves are not aware of in their consciousness. However, it is students' perceptions of their experience and satisfaction that matter and affect them, over any "outside" possibly more objective reading of their experience (e.g. Satyam et al, 2018). For this reason, I report on what the students identified verbally as satisfying.

Data Sources

There were two sources of instances of satisfying moments: in-interview proof construction tasks that students said were "satisfying" and out-of-interview instances. This analysis focuses on the latter. Across all four interviews with the eleven participants, there were N = 75 instances of satisfying moments; that is, 75 times participants reported some experience related to their work in the course as satisfying.

Of these 75 instances of satisfying moments, 56 had emotion graphs associated with them. This discrepancy in number comes from two main sources. One, in the first interview, I had participants draw an emotion graph for only one of the satisfying moments discussed. Two, sometimes students would talk about moments that they said were satisfying but did not draw a graph for it. A minor source is particular to one participant, who drew graphs of how he felt about each question in an entire homework assignment, not for the specific question that did feel satisfying. These graphs were not usable and therefore not counted. Regardless, this analysis does not rely on the emotion graphs.

From participants' verbal descriptions of the satisfying moments, I reduced the data to be analyzed through a careful process to preserve relevant meaning. This is described below. Specifically, I produced 1-2 sentence descriptions of what exactly felt satisfying in each experience, which were distilled representations of their experience.

Data Preparation

Distilling audio to short descriptions*.* To prepare the data for analysis, I listened to the audio of each satisfying moment in each of the four interviews for each participant. After listening to the full retelling of each satisfying moment, I wrote (a) a summary of what had occurred and (b) a short 1-2 sentence description of what was satisfying to that student, based on what they said. An example of one of these short descriptions is:

Getting a problem you've been stuck on for a while and getting it yourself (Joel-1-1).

The parenthetical identification lists participant, the interview, and a number associated with that satisfying moment. Thus, in the description above, this satisfying moment was the first one Joel talked about in the first interview. The goal of this two-step process was to carefully identify and keep *what* felt satisfying to the student. To maintain validity, I later rechecked each of my summaries and 1-2 sentence descriptions against each other, to check that no important relevant information had been lost that would influence coding.

Probing about singular satisfying moments. In many cases, I explicitly asked students whether there was a singular moment within this entire experience that felt satisfying. I did not always remember to ask this question, as it was a question that arose over the course of data collection. When I did ask it, I included their answer into my sentence description. When I did not ask it, I stuck to the summary as close as possible when writing my sentence description, trying to minimize my inferences while also trying to not lose important information about the situation. Given the emergent nature of my analysis, I did not know what information would be significant ahead of time, so minimizing inferences was a non-trivial task.

Why not code transcripts directly? In grounded theory methods, it is typical to code participants' words directly, sticking close to what was verbally uttered. I chose to depart from this tradition in my analysis: My short descriptions (the data to be coded) were by nature already interpretive; they were colored by what I noticed while listening and were written by me, not my participant. This was a purposeful decision, however, for many reasons. Firstly, the participant's descriptions of what happened were often prolonged and spread out, because I would probe with questions at different points in time. This is not surprising, as their retellings of their experiences were akin to story-telling, and story-telling does not always occur in a straightforward fashion. Secondly, as stated earlier, I sometimes would ask the participant directly what moment exactly felt satisfying within their experience, but this (a) presumed that there was a singular moment to the participant and (b) may have introduced pressure to the participant to find something to say. Thirdly, participants' tone of voice (e.g. excitement when talking about a certain point) seemed incredibly important for analyzing emotions; tone would have been lost by using transcripts that only included the spoken language.

Given these concerns, I thought it better to (a) listen to the entire event and then (b) summarize what seemed to be the satisfying moment to the student, sticking close to their interpretations of events. This meant listening to participants intently, paying special attention to aspects such as tone of voice. An outside researcher could verify this by listening to the audio as well. But admittedly with this analysis, I took into account my familiarity with each student – what I picked up about their personalities and how they communicate, much of which is not present in a transcript. Considering the nature of qualitative analysis, taking into account familiarity with participants is appropriate here.

Data cleaning: Excluded data and separating out independent instances. After listening to recorded audio, there were 75 satisfying moments. Of these, one entry was excluded as it concerned why a participant had had no satisfying moments. In another case, one instance of a satisfying moment was actually two: the participant talked about understanding equivalence classes being satisfying, and also that talking to fellow students about math was satisfying as well. An assumption that underlies this dataset is that each satisfying moment is independent from the next. I therefore separated these instances into two.

In contrast, experiences that had some element in common were kept together, even if there were multiple different aspects that were satisfying. For example, Stephanie said that doing and understanding homework feels satisfying. She also expressed that re-explaining the homework to fellow students and getting better grades than others on the homework felt satisfying. Even though there were multiple things Stephanie found satisfying, I did not separate them into different satisfying instances because both concerned the same event, homework. If I had done so, these three instances would not have been completely independent of each other. Through this process of excluding one instance and splitting one instance into two, I arrived at N = 75 experiences to be coded.

Data Analysis: Creation of Coding Scheme

Coding. I coded all N = 75 descriptions of satisfying moments using grounded theory methods to create a preliminary coding scheme. This bottom-up coding scheme was created in the following fashion:

1. Assigning raw keywords to each instance for what participants felt was satisfying
2. Aggregating all the raw keywords together
3. Consolidating keywords similar in meaning
4. Repeatedly grouping similar keywords into larger categories, and
5. Applying coding scheme and looking within each category for variation.

Steps 4 and 5 were done cyclically in many rounds, until reaching the coding scheme detailed below. The goals of this cyclical process of defining categories was to minimize overlap between top-level categories while retaining conceptually relevant categories. For example, while the *Understanding* category had subcategories that could be separated, there was variation left within the *Understanding* category. This was purposeful in that teasing apart nuanced meanings of the word “understand” would not be particularly illuminating.

Testing codes from the literature. While a framework for satisfying moments did not previously exist, there are similar ideas in the literature, including aha moments, mathematical beauty, and self-efficacy. It made sense then to build off existing theory and connect to the literature, in order to grow our knowledge collectively. I derived codes from the following sources:

- Mathematical beauty: Sinclair (2006), Hardy (1940), Inglis & Aberdein (2014), Blåsjö (2012)
- Aha moments: Liljedahl (2004)
- Self-efficacy: Bandura (1977)

These sources were chosen based on their thoroughness or uniqueness in examining the topic.

To test out the emergent codes against the research, I coded a “representative” subset of the entire data: N = 30 of the satisfying moments. I chose instances that exemplified typical instances or were unique. With the combination of these two, the goal was to have relatively high theoretical saturation of all the experiences within my dataset. After this test, I dropped some of the codes because there were no recorded instances in the representative dataset. A few codes from this round were retained in the final set.

Difficulties in coding scheme creation. A number of issues arose while in the process of creating the coding scheme using bottom-up methods. First, it was confusing to keep straight coding for *what* was satisfying versus *why* something felt satisfying. For example, doing well on homework is *what* is satisfying but because it is an indicator of my understanding is *why* that event was satisfying. Second, instances had multiple codes, which meant any instance could fall into multiple groups, making constant comparison difficult. Nevertheless, I focused on the criteria and what to include and exclude with each additional instance. Coding along multiple dimensions – e.g. code for type of success, difficulty, and people involved – did not work because this started to capture contextual elements and not main elements; this was too much information that it was obscuring the main themes. Lastly, there was a large amount interrelatedness between codes, which made refining the coding scheme difficult.

In the following chapters, I discuss the results to both research questions. Chapter 6 concerns development in students' proving, and Chapter 7 is about kinds of satisfying moments that students encountered in relation to the transition to proof.

CHAPTER 6: Development in Students' Proving

In this chapter, I answer my first research question, *How does undergraduate students' proving develop over the duration of a transition to proof class?* I discuss a selection of some of the major kinds of productive changes that occurred, illustrating each using participant(s) as examples. I focus on four important, prevalent developments: (1) sophistication in how students chose proof techniques and their rationales for their choices, (2) awareness about how a solution attempt was going and harnessing that awareness for subsequent strategies, and (3) using examples to notice patterns and kickstart insight when stuck, and (4) becoming comfortable with exploring and monitoring. Next, I present a longitudinal profile of one individual, to highlight how growth in reasoning and performance do not necessarily happen together. I then discuss some less prevalent developments and end with a cursory analysis for developments across the sample.

I interviewed each of the 11 participants four times over the semester and in each interview, they worked on two proof construction tasks. This chapter uses the data from the eight proof tasks for each of the 11 participants (see Table 6.1, a repeat of Table 4.3).

Table 6.1: Proof Construction Tasks By Interview

Interview 1	Statement
Task 1	Suppose x and y are integers. If $x^2 - y^2$ is odd, then x and y do not have the same <u>parity</u>.
Task 2	Prove the following statement: If a and b are strictly positive real numbers, then $(a+b)^3$ never equals $a^3 + b^3$.
Interview 2	
Task 1	Prove the following statement: If x and y are <u>consecutive</u> integers, then xy is even.
Task 2	Prove the following statement: If a, b, and c are non-zero integers such that a divides b and a divides c, then a divides $(mb + nc)$, for any integers m and n.

Table 6.1 (cont'd)

Interview 3	
Task 1	Prove the following statement: Suppose *x*, *y*, *z* are positive integers. If *x*, *y*, and *z* are a Pythagorean triple, then one number is even or all three numbers are even.
Task 2	Prove the following statement without using induction: If n is an odd natural number, then $n^2 - 1$ is divisible by 8.
Interview 4	
Task 1	Prove the following statement: If a and b are odd perfect squares, then their sum a + b is never equal to a perfect square.
Task 2	Prove the following statement: If x, y are positive real numbers and $x \neq y$, then $\frac{x}{y} + \frac{y}{x} > 2$.

Indicators of Being Stuck

As I watched videos of students' proof tasks, I made notes about what observable behaviors contributed to my judgment that students were stuck. A list of these are included below, across the participants' video data I watched:

- Silent
- No writing
- Stares at paper
  - Holds paper closer
  - Sits back from paper, to look at it from a distance
- Taps/plays with pen
- Touch face with hand or pen

These behaviors were not exhaustive and individuals exhibited different behaviors specific to themselves, but I believe these behaviors cover much of what we see when a person is stuck.

Students' Performance on Proof Tasks

While the focus of this chapter is on student's development in problem solving, I provide some attention first to the quality of the written arguments they produced in the four interviews. One common way to think about development in proving is in terms of performance – are students more successful at proving as the course goes on?

For this reason, I coded students' written work on each of the proof tasks for correctness. I assigned their attempts to one of three categories: *Correct, Partially Correct,* and *Incorrect.* The idea behind each category was to match standards set in the course, i.e., what students would receive as a score for their written work if they passed it in for homework. One reason for this choice was that students' own standards of whether a proof was correct or not would be influenced by the course's standards. *Correct* proofs were those that would receive full credit on homework, *Partially Correct* proofs would likely get at least half credit on homework, and *Incorrect* would get less than half credit. In assigning these "grades" I drew on my experience as a teaching assistant in the course. Table 6.2 below provides more clarity on the criteria for each correctness category, in terms of content of the written work as well. Note this is a rubric for proof as a product, whereas my analysis regarding problem solving is about process. Therefore, conceptual, logical and expression issues were all considered errors.

Table 6.2: Rubric for Scoring Performance on Proof Tasks

	Correct	**Partially Correct**	**Incorrect**
Criteria in terms of the course standards	Would receive full credit in course	Would receive at least half credit in course	Would receive less than half credit in course
Criteria in terms of content	Correct proof, with no conceptual or major logical errors. May contain trivial mistakes (e.g. using same variables, minus sign, etc.) that do not affect validity of proof.	Overall idea of proof is clear and correct. May contain 1-2 conceptual or expression errors, depending on severity.	Anything less than Partially Correct or at least 2 severe errors
Common errors	-Proved something more general than given statement -Minor expression issues that do not affect validity or logic of proof	-Wrote negation incorrectly -Stated contrapositive incorrectly -Did not justify a step (unless this is the point of the proof, in which case Incorrect) -Incomplete Cases: Forgot a case -Minor expression issue that do affect validity or logic of proof -Informally written in words (but idea is correct) -Started with the goal rather than proving it	-Proved the converse

Table 6.3 below shows students' performance on each task across the interviews. In order to easily look across students' success over the entire interview series, correctness is quantified here in the following way: 1 denotes Correct, ½ denotes Partially Correct, 0 denotes Incorrect. The last column shows their correctness score across interviews (out of a possible total of 8).

Table 6.3: Performance across Proof Tasks by Participant

Participant	1-1	1-2	2-1	2- 2	3-1	3-2	4-1	4-2	Total
Amy	0	1	1	½	1	½	1	1	**6**
Charlie	½	1	1	1	½	0	0	½	**4** ½
Dustin	0	0	½	½	0	0	0	0	**1**
Granger	1	1	1	1	1	0	½	1	**6** ½
Gabriella	½	1	1	0	½	0	0	0	**3**
Joel	1	½	1	½	0	0	0	1	**4**
Jordan	0	½	1	0	0	0	0	1	**2** ½
Leonhard	½	0	1	½	0	0	0	0	**2**
Shelby	0	½	0	½	0	0	0	0	**1**
Stephanie	1	1	1	½	0	0	0	0	**3** ½
Timothy	1	1	1	1	½	1	½	0	**6**

Note. The headings denote Interview-Task (e.g. 1-1 means Interview 1-Task 1). Cells: 1 denotes Correct, ½ denotes Partially Correct, 0 denotes Incorrect.

Looking across the interviews, Granger had the most success across the eight tasks (6 correct, 1 partially correct), followed by Amy (5 correct, 2 partially correct) and Timothy (5 correct, 2 partially correct). Dustin and Shelby got the least correct across all interviews (only 2 partially correct each). It is important to note that the difficulty of the items was equated across the interviews; in my estimation, difficulty generally increased over time. Later proof tasks were more reliant on increased content knowledge of basic number theory and had more complicated solution paths. If item difficulty had stayed the same throughout, we would expect that performance would improve as interviews progressed. Because difficulty cannot be assumed to be constant but in my view increased, participants' poor scores on later tasks should not necessarily be taken as an indicator that they had not improved. In the following analyses, I only discuss tasks where students were at least partially or completely correct.

Cross-Individual Developments

I present the most pervasive developments and practices that appeared, i.e. the developments that occurred across the largest shares of my eleven participants. The goal is to describe the developments that have the most grounding in data. For these developments, I take a cross-individual analysis: I discuss one to three student examples to (a) illustrate what that development looked like as it unfolded but also to (b) highlight any variation in how that development occurred. I provide a summary of the changes seen for each development. I would like to note that my choices of participants are not meant to say that other participants did not show these developments nor even that these are the *best* examples across participants. The students discussed as examples of each development are merely to be illustrative of the development, serving the reader.

Development A: Changes in Choosing a Proof Technique

One common development that occurred across participants were changes in how they chose what proof technique to pursue, when approaching constructing a proof. By *proof technique*, I mean tools such as direct proof, proof by contradiction, proof by contrapositive, cases, and proof by induction—the techniques that were taught in the course. For the sake of redundancy, I will often refer to proof by contradiction as just contradiction and proof by contrapositive as just contrapositive. The first half of the course was about learning proof techniques, so naturally many students generally thought about what proof technique to use as a major way of approaching constructing a proof. In addition, homework tasks were often written in such a way that one of the proof techniques led to an easier proof, over using other proof techniques. Eight of eleven participants showed signs of this development, based on data from the interview notes and across all tasks (not only the ones where they became stuck). I discuss two participants here, as illustrative examples of this development.

Example A1: Favoring one technique. From the beginning (Interviews 1 and 2), Stephanie favored proof by contradiction over all other techniques when constructing a proof. In Interview 1 – Task 1, she immediately jumped to trying proof by contradiction because the statement was an implication, having an "if-then" structure: "When I see the if-then statement, I immediately think I can do this by contradiction." She explained she felt comfortable using this technique. Figure 6.1 shows her work, where she immediately identified the assumption as "A" and conclusion of the statement as "B" and wrote the negation.

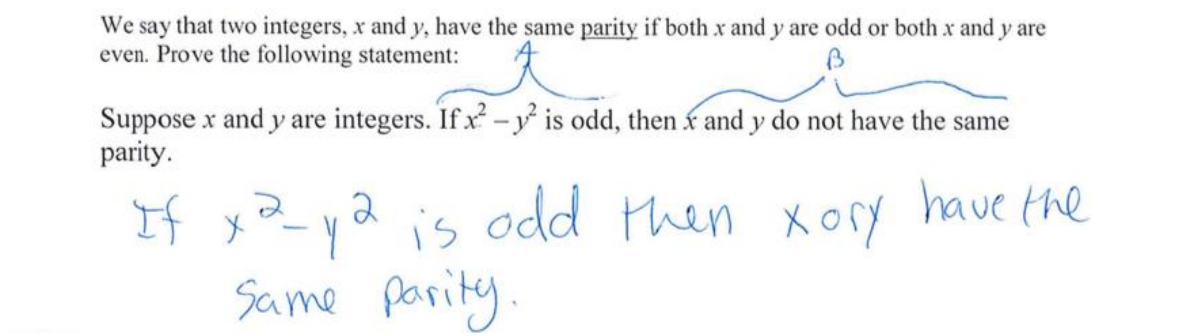


Figure 6.1. Beginning of Stephanie's work on Interview 1 – Task 1

Note that Stephanie technically wrote the negation incorrectly; the correct negation is "A and not B" i.e. "$x^2 - y^2$ is odd and x or y have the same parity." Instead, she because she wrote the negation as an implication, a common error. However, this error did not affect the rest of her proof and her reasoning for picking proof by contradiction is unaffected by her execution. It is interesting that already by the first interview Stephanie felt most comfortable with proof by contradiction, considering that this was new knowledge they had recently learned in class, not something they came to the course already knowing.

In the next interview, Stephanie go-to method was still proof by contradiction. Upon starting Interview 2 – Task 1, she said "I can see that this is an if-then statement, so automatically I'm going to try to use contradiction, but I don't know if it will work or not." She explained during the debrief that “When I read an if-then statement, I'm most comfortable using negation or a contradiction. So then I just try that, even though I know it doesn't always work, but I just try it." Note how the use contradiction is automatic for her, and she herself said outright she does not always know if proof by contradiction will lead to a correct solution. The general structure – that the statement has “if” and “then” clauses – is enough to determine that she can use her favored technique, but she did not make use the statement in any further way to guide her choice of technique.

Stephanie did indeed get stuck on her proof by contradiction, so she switched to proof by contrapositive (see Figure 6.2).

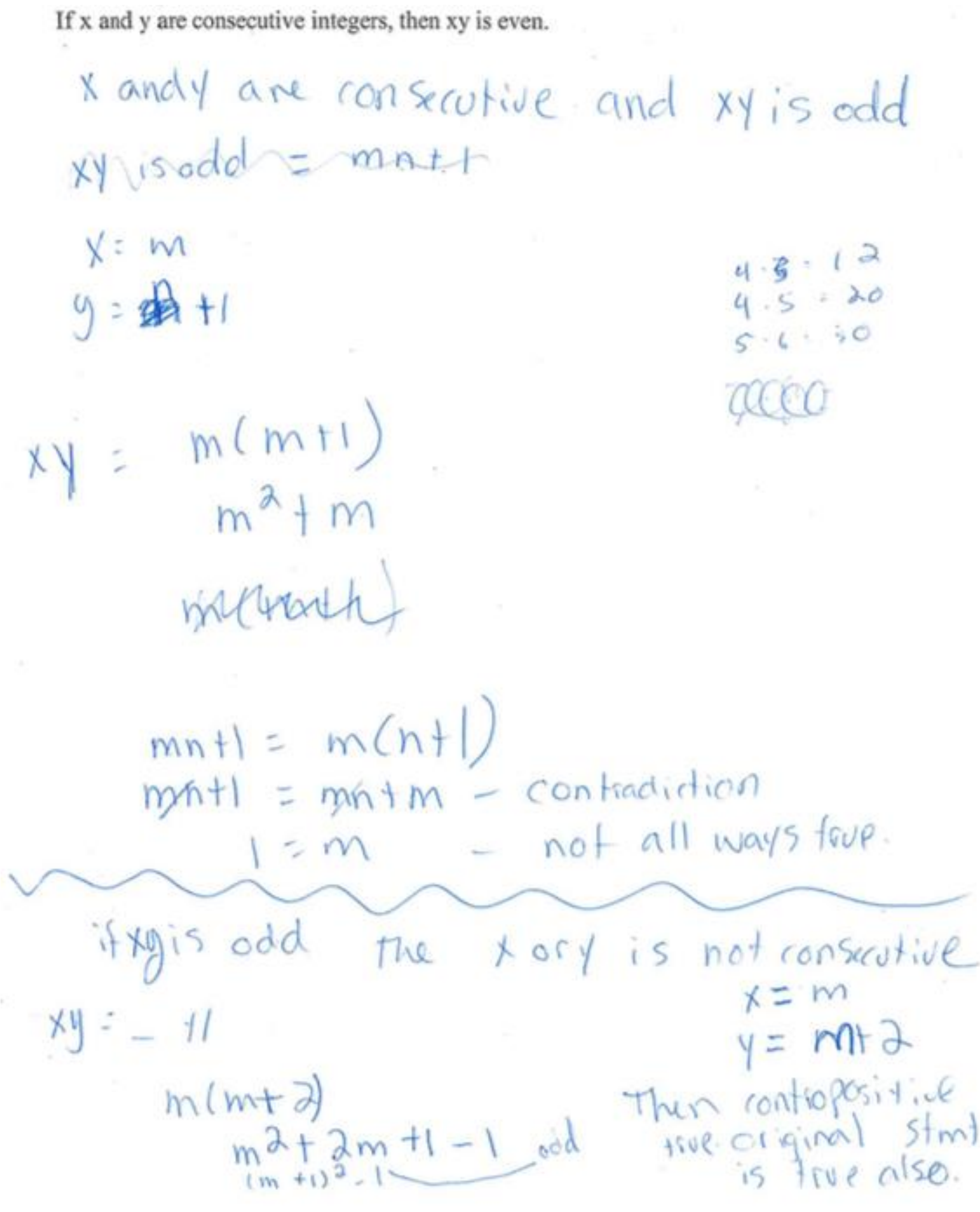


Figure 6.2. Stephanie's move to contrapositive on Interview 2 – Task 1

She explained during the debrief, "I'll try contrapositive and then I felt a little better after I tried contrapositive just because I thought [out of] both of them, probably one of them was gonna be right." Stephanie did not give a rationale for why specifically proof by contrapositive, just that it was another technique.

Summary. These snapshots of Stephanie's thoughts during Interviews 1 and 2 showcase how a student can "latch on" to a proof technique and use it whenever they can. Stephanie did have a condition for when to use proof by contradiction, when she sees an if-then statement. However, this applies to nearly all statements to be proven in the course that we can safely say this is her general technique. Stephanie becomes less dependent on proof by contradiction and her rationales do become more sophisticated over time, but her work was unfortunately incorrect on all four tasks on Interviews 3 and 4. We turn then to a different student in order to better see how choice of proof technique and rationale changed over time.

Example A2: Recognizing advantages of a technique, independent of statement. I now present the case of Timothy, to show development that extends what we saw through Stephanie. Timothy was similar to Stephanie in having favored proof techniques in the beginning, but his rationales became more sophisticated and based on the statement itself as his interviews progressed, in addition to producing correct or partially correct proofs.

Figure 6.3 shows Timothy's attempt in Interview 1 - Task 1 to construct a proof for the statement, "If x^2 – y^2 is odd, then x and y do not have the same parity." When stuck in the beginning, he re-read the question and wrote what was known. At this point he switched from his direct proof attempt to proof by contrapositive.

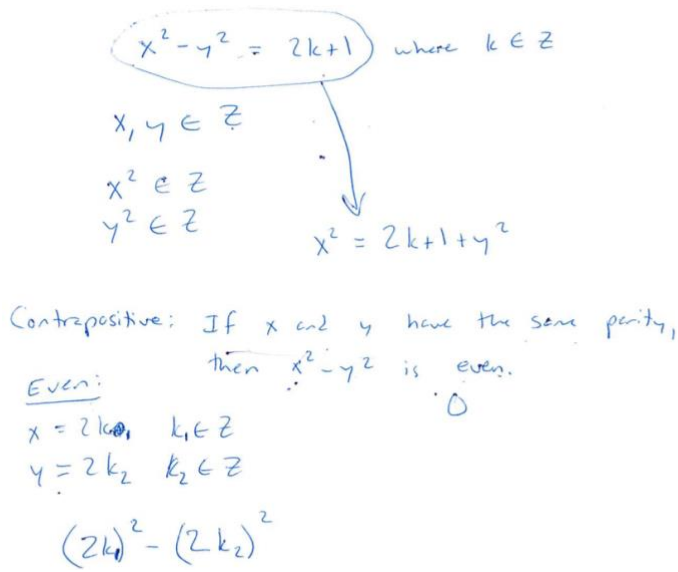


Figure 6.3. Beginning of Timothy's work on Interview 1 - Task 1

When asked why he selected contrapositive, he explained it was a method from class but also that it was a logically equivalent tool to direct proof that he could use:

> *Timothy:* It was confusing me when I'd try to think of it the normal way so I knew the contrapositive is true, it's basically the equivalent, logical equivalent.
> …
> *Interviewer:* So actually, so how did you come up with contrapositive?
> *Timothy:* Looking at it straightforward didn't…it wasn't working for me so I know we learned in class that the contrapositive is basically not B implies not A. I knew we said that was logically equivalent, so if I could prove the contrapositive was true, then I could prove the original statement was true was kinda my thinking with that.

He explained that a direct proof method was not helpful in generating a proof, but he gave no specific rationale for choosing contrapositive over other proof techniques. His explanation implied that contrapositive was a legitimate tool from class, so why not use it? While it is possible he may have had some internal reason for using contrapositive, he neither mentioned this on his own nor articulated any further reasons when questioned.

Later in this interview, he talked more about contradiction being one of his "go-to" methods and why:

Timothy: I always go about it with either contradiction or induction or straight up so I kinda knew that I might be able to contradict this never equaling that, so I wrote out the contradiction...I guess contradiction is a little easier for me to think about. You just say the first part of the implication is true and the second part is false. So it's just easier in my head, I guess, just to think about rather than switching around the implication, negating both parts.
Interviewer: Okay
Timothy: So I guess that's why I go to that first.

Timothy expressed here that contradiction was easier for him than contrapositive, which involves negating the assumption and conclusion. His insight about the work involved in setting up the two different proof techniques – contradiction vs. contrapositive – was true. It is important to note that he had some rationale for why he might use contradiction, but it was couched in terms of ease of use, first and foremost.

The idea of ease of use as determining choice of proof techniques showed up in latter interviews. In his work for Interview 2 - Task 1 (see Figure 6.4), Timothy started by defining x and y using the definition of consecutive numbers and in calculating xy, became stuck over what to do.

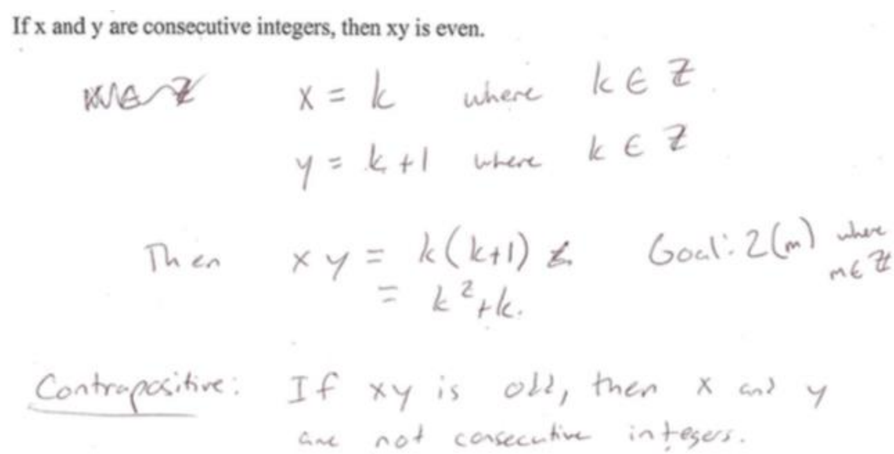
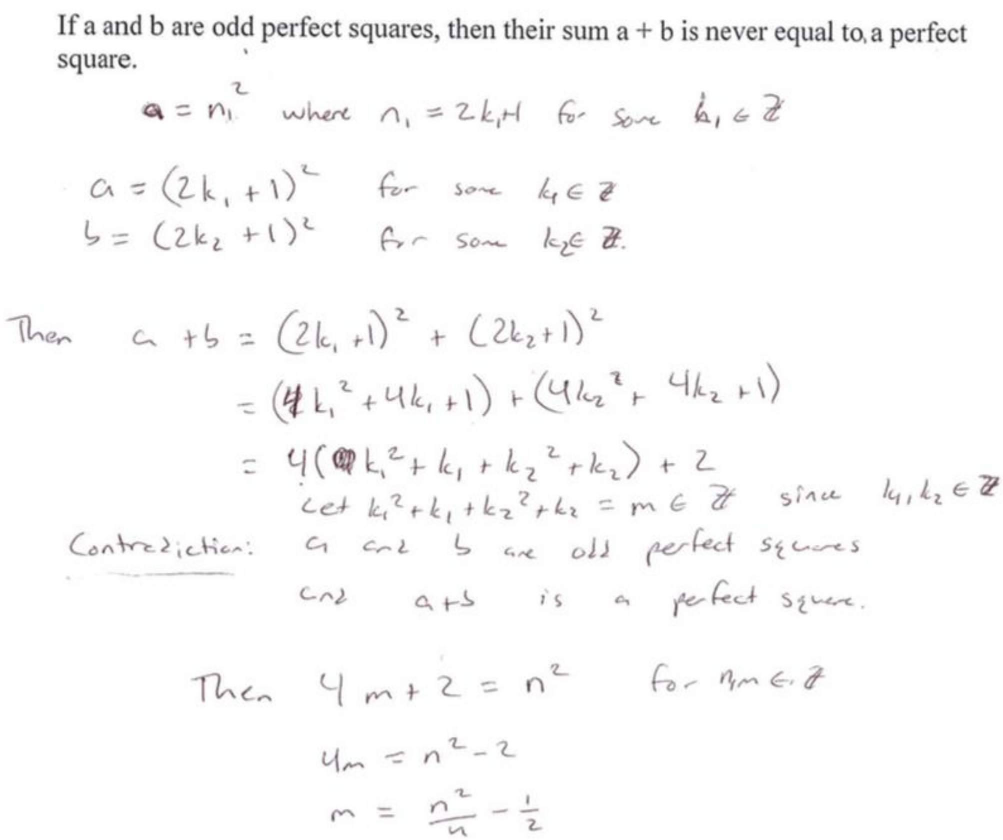


Figure 6.4. Timothy's switch to contrapositive on Interview 2 – Task 1

He then switched to contrapositive when stuck because "sometimes that's an easier way for me to look at it." Similar to Leonhard's reasoning, he knew that contrapositive was easier on some level for him but not for any reasons specific to the statement and did not further articulate why. What exactly made this method easier remained unknown to him or at least was not clear enough to him to easily articulate when asked. (In the end, his contrapositive proof was not to his liking and also not correct).

But by the end of the interviews, Timothy showed sophisticated thinking in considering which proof techniques to use. In Interview 4 - Task 1 (see Figure 6.5), Timothy became stuck after computing the goal (a+b) directly.


If a and b are odd perfect squares, then their sum a + b is never equal to a perfect square.

$a = n_1^2$ where $n_1 = 2k_1+1$ for some $k_1 \in \mathbb{Z}$

$a = (2k_1+1)^2$ for some $k_1 \in \mathbb{Z}$

$b = (2k_2+1)^2$ for some $k_2 \in \mathbb{Z}$.

Then $a+b = (2k_1+1)^2 + (2k_2+1)^2$

$= (4k_1^2+4k_1+1) + (4k_2^2+4k_2+1)$

$= 4(k_1^2+k_1+k_2^2+k_2) + 2$

Let $k_1^2+k_1+k_2^2+k_2 = m \in \mathbb{Z}$ since $k_1, k_2 \in \mathbb{Z}$

Contradiction: a and b are odd perfect squares and a+b is a perfect square.

Then $4m+2 = n^2$ for $n, m \in \mathbb{Z}$

$4m = n^2 - 2$

$m = \frac{n^2}{4} - \frac{1}{2}$

Figure 6.5. Timothy's work on Interview 4 – Task 1

He explained that he used contradiction because “it’s easier when I know something like is equal to something or is something.” His rationale was similar then to Leonhard’s for this same task. He then gave this further rationale for why contradiction:

> I was trying to prove that it’s not equal to a perfect square and I know from past experiences, **it’s easier when I know something is equal to something or is something.** So I tried to use contradiction because **I knew I could say then it is a perfect square.**

His argument was that he wanted to be able to work with an equality, much like Leonhard.

Timothy also gave a rationale for not using another method, contrapositive:

> I thought about contrapositive, too, but then it would say that A and B are not perfect squares and that’s again, like something’s *not* so I mean, it’s easier for me to work when I know like a straight definition of something. So if I could keep this, I knew if I could keep this, like they are perfect squares and say this is a perfect square, then it’d be easier to work with.

His explanation was similar to his prior one about equality of objects being easier, i.e. knowing things are not equal is not as helpful. His subgoal then was to find a proof technique that would give him a+b *is* a perfect square.

This task is notable however for drawing out Timothy’s observations on contradiction:

> I never really thought about it this way but I realized when you use the contradiction, you don’t really have the assumption and conclusion anymore…**you can actually pick any part of that statement you want and work with it. Rather than with an if/then statement, you start with the assumption and try to work to the conclusion**. So you’re not as limited, I guess.

Timothy gave a high-level explanation of the nature of proof by contradiction. He found proof by contradiction to be freer than other techniques, due to being able to work with all parts of the statement. This stood in contrast to starting with the assumption and trying to prove the conclusion as is done in direct proof but also proof by contrapositive. It is of separate note that this revelation came about during this interview context, based on the "I never really thought about it this way but..." clause. The interview served as a vehicle for reflection on proof techniques for Timothy.

Summary. Timothy's went from picking a proof technique (1) because it existed as a tool to (2) having a fuzzy sense that it would be easier to (3) explaining how the content of the statement drives the problem solving approach to (4) articulating understanding at the meta-level of how a technique functions as logical tools. His later interviews revealed insights for when to use contradiction that did not depend on statement content but instead meta-level structure.

Comparing developments in choice of proof techniques. Both Stephanie and Timothy showed similar growth in how they chose proof techniques to pursue through most of their interviews. Both discussed liking and being drawn to certain techniques, as their "go-to" method. Timothy's latter interviews showed some level of weighing the utility of different techniques, to think about which would be *better*, whether it be a cleaner proof or just easier. He noticed that being able to set things equal provided the prover with more to work with; contradiction was therefore the most useful technique, based on the content of the statement.

The difference between the two lies in where they ended: Timothy came up with a general insight for when contradiction was useful. By looking across these two students, we can see this general trajectory in how students grew in how they chose techniques to use. If we conceptualize this specific development as a series of stages, Figure 6.6 illustrates the stages students tended to step through.

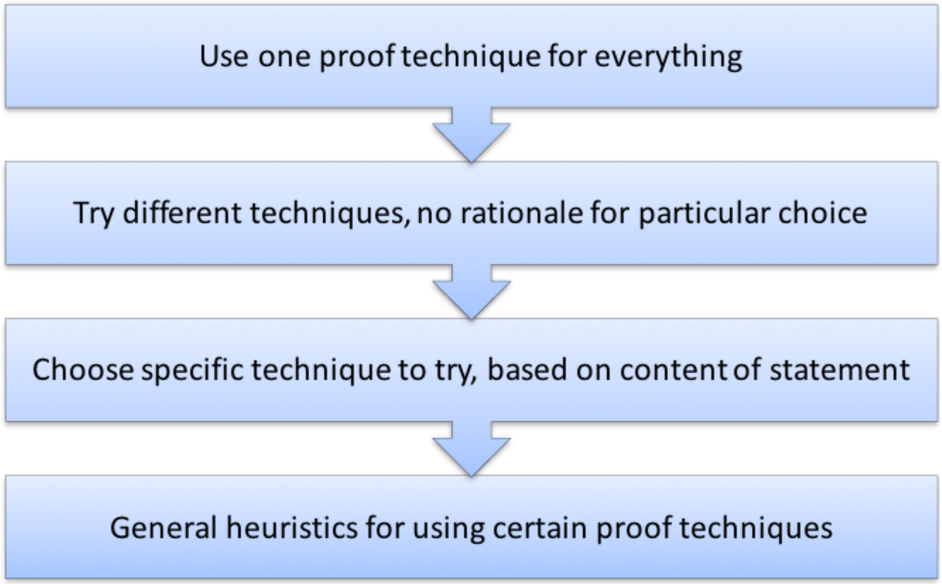


Figure 6.6. Stages of development in how students choose proof techniques to pursue

To use an analogy, let us think of proof techniques as hammers. In the beginning, students have a certain hammer they like for reasons that tend to be personal and not mathematical, and they use this hammer for all tasks, regardless of the nature of the task at hand. After some time, they start using different hammers other than their favorite but have no clear rationale for why one over another; they just pick up a different one when the need arises. They then start using specific hammers for specific tasks (attending to content of the statement to be proven), but without yet explaining why they are doing so. Finally, some students see when to use certain hammers over others, understand the advantages of each, and can explain why. The same way different hammers work better in different situations, different proof techniques can lead to more straightforward proofs.

Development B: Assessing How the Solution Attempt Is Going and Harnessing It

Another common development among participants was a growing metacognitive awareness of how their solution attempt was going, usually when they felt they were on the wrong track. Four of the eleven participants showed development of this kind. Being aware of how one' s solution is going is normal and to be expected; what is important is examining what students did in response to their awareness (albeit involuntarily) and how that guided them to better solutions. For these reasons, I highlight individual tasks where students showed they had this awareness, with the implication that this did not appear in earlier interviews. I discuss three students to show variation in how students harnessed metacognitive awareness: Granger, Timothy, and Jordan.

Example B1: Intuitive awareness lead to restart. Granger was another student who was aware when things were going wrong, even if the reason why was not clear. In Interview 2 – Task 2, he said from the start "this is going to throw me for a loop" – and it was indeed a difficult task for him. He became stuck at some point and took multiple attempts, as can be seen by all the cross-outs in his scratch work in Figure 6.7.

Interview 2 Task 2	Prove the following statement: If *a*, *b*, and *c* are non-zero integers such that *a* divides *b* and *a* divides *c*, then *a* divides (*mb* + *nc*), for any integers *m* and *n*.

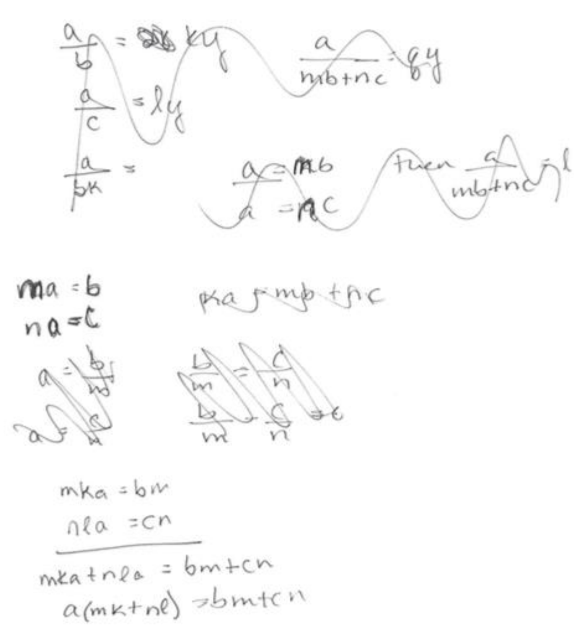


Figure 6.7. Granger's scratch work on Interview 2 – Task 2 (statement provided)

The following exchange happened during the debrief, in which he showed awareness that things were off:

> *Granger:* So I was like, "What am I doing? This isn't right. Something's not right here."
> *Interviewer:* But it sounds like you had a sense that... You knew that like, "I am not doing this the right way."
> *Granger:* Yeah, I definitely did. I don't know. I just know... **I don't know how to explain that. You just know when something isn't right.**
> *Interviewer:* Is it like when it's [this attempt is] not helping you get anywhere or it's not clarifying things? Or is it really just like an intuition?
> *Granger:* Yeah, **just like an intuition, like, "That does not... This statement absolutely doesn't make any sense with this," and I was like, "It can't be right." ...But you know it's just like, "This does not agree with the definition at all, so what am I doing?" And then I just reassess the situation and I'm like, "Okay, let's start fresh."**

Granger knew something was wrong, intuitively. He could not pinpoint what exactly was wrong but had an awareness that this could not be a correct way to go about it. He also explained his strategy of starting over:

> *Granger:* Usually, on homework I would pick a page and start going and, I don't know, it's a weird thing, I'd be writing or something, and if it's wrong, I'd cross it out and I'd try again if it's wrong... Eventually, if I get to this much space where I've gotten**... I just flip to a whole new page and it's like a refresher like, "Okay, you start a whole new... What's going on."**
> *Interviewer:* So that kind of helps, it sounds like.
> *Granger:* Yeah, yeah, definitely. I don't know**. It's intimidating when you see a whole bunch of crossed out marks and it's just like your brain is focusing on what you got wrong and... Yeah.**
> *Interviewer:* As opposed to fresh ideas or trying new things.
> *Granger:* And we learned in... Ironic, I learned in psychology, when you're trying to figure out a problem, your unconscious mind is also thinking about it but you don't realize, it's unconscious, but...So as I'm flipping the paper over and just like resetting myself, also, my unconscious is thinking about what I did wrong already, so it doesn't matter, I already know what not to do.
> *Interviewer:* So you don't need to look at it to...
> *Granger:* Yeah, exactly. And looking at it, it messes up consciously what I'm doing unconsciously.

His subsequent strategy was to abandon his past attempt and ways of thinking completely and start afresh. It worked on this task: Near the bottom of his scratch work, he started working in a more helpful direction and was able to get to a correct proof.

Summary. Granger knew intuitively that his work was off, expressing that “you just know when something isn’t right.” This mathematical sense for when things were off was helpful here, in that it led him to let go of what he had done and start something afresh, leading to a correct solution.

Example B2: Awareness lead to finding new strategy. I now highlight one task where Timothy’s awareness drove his solution attempt. In Interview 2 - Task 1, he became stuck multiple times over the course of proving: “If x, y are consecutive numbers, then xy is even” (see Figure 6.8).

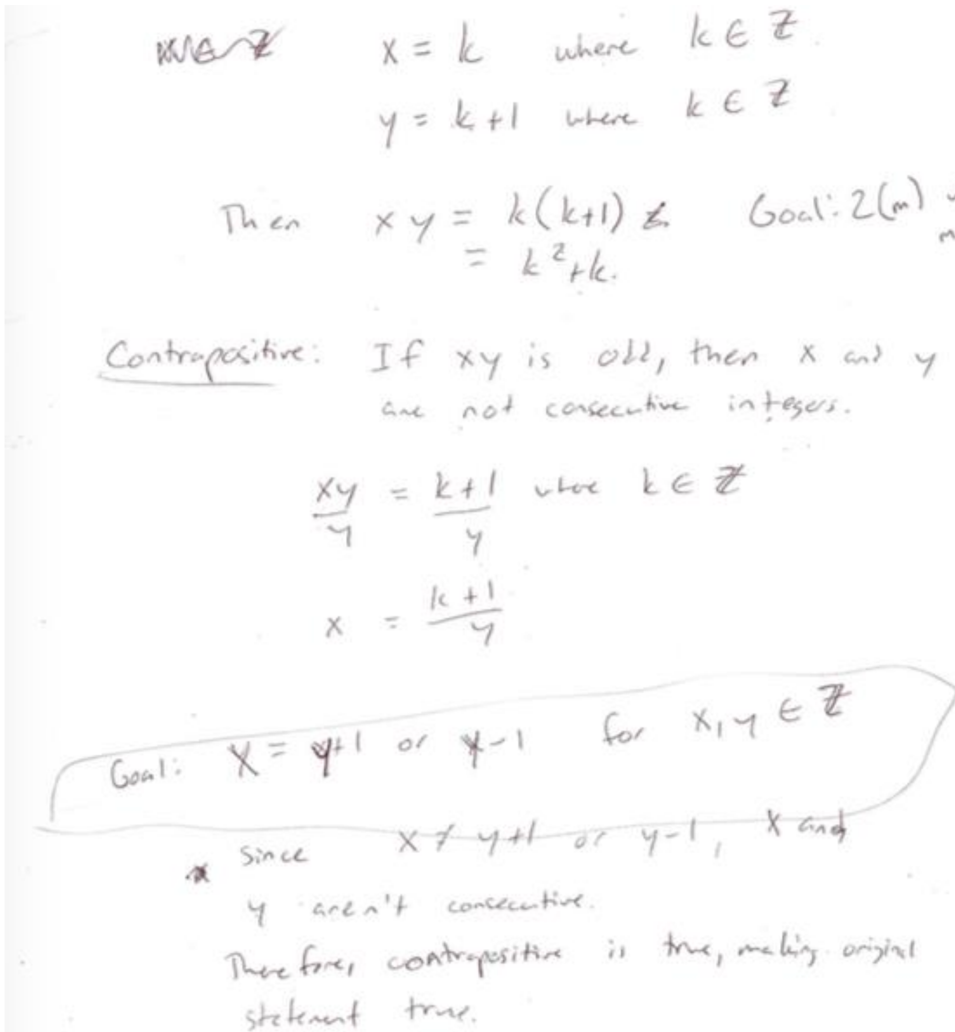


Figure 6.8. Timothy's first attempt at Interview 2 – Task 1

In response, he started reasoning out loud about the mathematical relationships ("If [x,y are] not consecutive, they wouldn't have this relationship...what does this tell me?") and explained that he would continue to try it this way but didn't know if it would work or not and "can't think of any other way" right now. He was assessing his attempt while working: "I finished it out because I just wanted to get something down but I didn't really like that one." After getting stuck twice more, he ended up with proof but he was unconvinced about it; he did not feel good about it.

In the debrief, he explained that "I didn't really like that one [proof by contrapositive]. And then I went back because I really wanted to do something with this directly. I liked that better." (The latter proof will be discussed in a later section). His use of the word "like" indicated a judgment part cognitive and part affective, of a sense of "fit" being off rather than fully about the correctness of this proof. His first proof did not sit well with him then, enough to lead him to look for a different solution. This sense served him well, as indeed his first attempt was not correct – but the one he came up with later was.

This affective metacognitive sense playing a role in his work was further indicated by his emotion graph and words for this task.

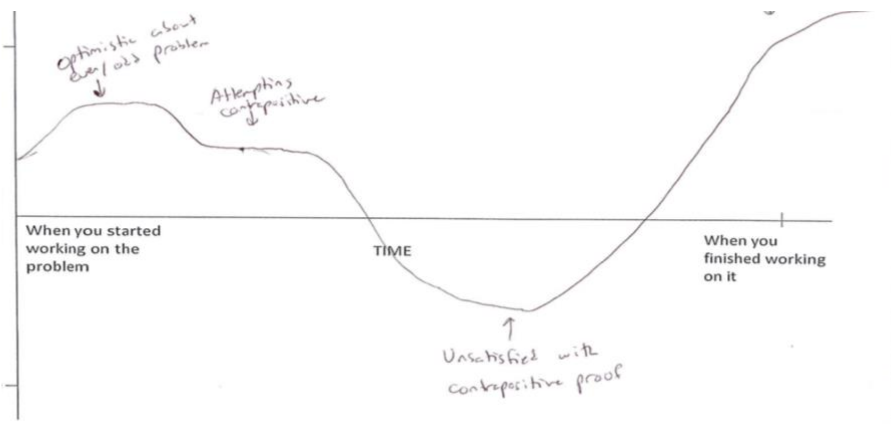


Figure 6.9. Timothy's emotion graph for Interview 2, Task 1. Note that the dip occurs when he was unsatisfied with his contrapositive proof.

His emotion words were "annoyed" when he was stuck and "frustrated" and "disappointed" when unsatisfied with his proof. In fact, his dip in emotion in the graph came from dissatisfaction about his contrapositive proof specifically. It is possible that he was dissatisfied because he thought his proof was not correct and that manifested itself through his emotions. However, even if this is true, it is interesting (in light of other research questions in this dissertation), that he spoke about the acceptability of his proof affectively.

Summary. In summary, Timothy showed a metacognitive, affective awareness and monitoring of his proof attempt. It drove him to keep thinking and look for another way, even though he had reached an end in his work. In a later section, I examine how Timothy was able to act on his awareness and find a better proof, more to this liking.

Example B3: Aware but stayed on same solution path. I now present a contrasting example, of a student who was aware when something was wrong but continued her strategies, not changing direction. Jordan became stuck and was aware that something was not working, but she would move past it and continue with her current strategy.

In Interview 3 – Task 1 (see Figure 6.10), Jordan was stuck in the beginning, stating she felt like she did not understand what she was proving.

Interview 3 Task 1	Prove the following statement: Suppose *x*, *y*, *z* are positive integers. If *x*, *y*, and *z* are a Pythagorean triple, then one number is even or all three numbers are even.

Figure 6.10. Statement of Interview 3 – Task 1

She had an idea about using two cases, where one case would be setting one of x,y, or z to be even and the other case would be setting all three of these variables to be even. She was confused though because she felt she was starting with what she normally would show. Regardless, she forged ahead with using cases on x,y, and z for the equation $x^2 + y^2 = z^2$. Figure 6.11 shows her work on this first case:

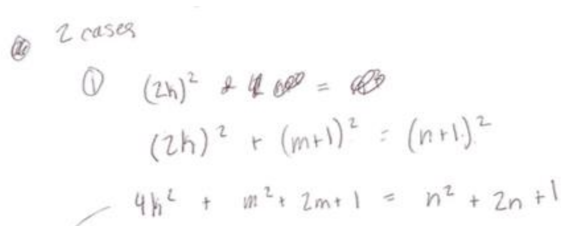


Figure 6.11. Jordan's beginning work on Interview 3 – Task 1. She assumed one variable was even and the other two were odd but it lead to a statement that did not help her.

She became stuck again after she expanded her terms, because she did not know what to do with the last line in Figure 6.11. She ignored it and skipped ahead to the second case, where x, y, and z were all even. Afterwards during the debrief, she said “I feel like maybe I got the first part, but then, I don't know. I just feel like there's a better way to properly do it.” Jordan was unsure about her outcome; she was aware that something was wrong and that there must be a better solution but that never changed her strategy of using these particular cases on x, y, and z throughout her attempt. Indeed, Jordan’s solution was incorrect in the end.

On Interview 4 - Task 1, she began by checking some examples and did some algebra but then became stuck with how to show that the expression was not a perfect square. From then on, her work was a stuttered series of stops and starts (see Figure 6.12): factoring out a 2, getting stuck, taking the square root, getting stuck, factoring out $\sqrt{2}$, and getting stuck. Jordan’s solution here was also incorrect in the end.

Interview 4 Task 1	Prove the following statement: If a and b are odd perfect squares, then their sum a + b is never equal to a perfect square.

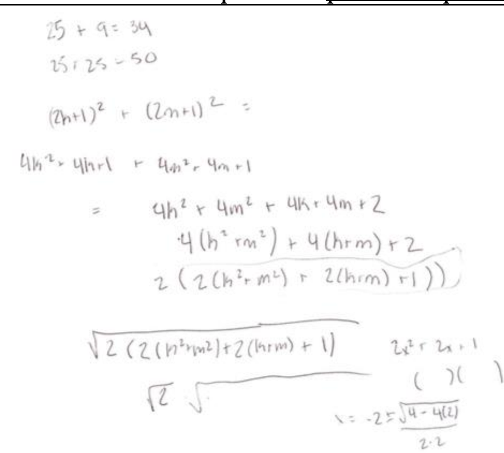


Figure 6.12. Jordan’s work on Interview 4 – Task 1 (statement provided). The last few lines are where she experienced multiple stops and starts.

During the debrief, she was honest:

> *Interviewer:* How are you feeling about it overall?
> *Jordan:* I just don't think I'm allowed to do that. I don't think I did it right.

Jordan was aware that her solution attempt was off (and in fact believed she had taken invalid mathematical moves). When she was stuck during the proving process, she ignored that something was off and kept pushing forward via algebraic manipulation.

Summary. These tasks from the latter two interviews showed that by the end of the course, Jordan knew when her attempt was off and had some idea of why (e.g. not knowing how to formally show something), but she would ignore it and move past it and/or not alter her current path.

Comparing developments in awareness and using it. All three of these students showed awareness when things were not going well. Granger's was more while he was working, feeling intuitively something was wrong in this process, while Jordan and Timothy's attention were focused more on not liking the outcome.

However, these three reacted differently to feeling something was wrong: Jordan would continue on with her current plan of attack, Timothy would re-assess what he was doing by reasoning out loud about the relationships, and Granger would start on a fresh page in order to not be influenced by this past thinking. Another way to examine this is to look at the conceptual "level" at which they worked: Jordan stayed grounded at the level of the algebra to try to make her way of thinking work, whereas both Granger and Timothy went back to the top level of the problem. Timothy in fact "zoomed out" of the problem (he physically would lean away from the paper) and muse about the task as though with a bird's eye view. Both Granger and Timothy stumbled upon correct proofs for their tasks here, whereas both of Jordan's proofs discussed here were incorrect. In summary, all three were aware of how their attempt was going, but Granger and Timothy used that to guide themselves to better solutions successfully.

These three students serve as variations of what students' awareness of how their proof attempt is going looks like and their subsequent metacognitive strategies: (1) continuing with the plan (Dustin falls into this category too), (2) abandoning the current path completely, and (3) playing around with what one is drawn to in order to find a new path. This last variation in particular may be an example of the inquiry-driving role of mathematical aesthetics in leading the mathematician to investigate certain avenues of solution attempts over others (Sinclair, 2004).

This discussion may make it seem like Jordan did not experience development. It is important to note that having awareness that one may have used invalid mathematical moves (as she worried about on Interview 4 – Task 2) is far better than assuming one's solution is always correct. Jordan may have continued on her current approaches when stuck because she thought them the most likely path to success or did not know what else to do, in the same way that Timothy and Granger thought changing their approach would lead to a correct proof and/or did not what else to do. The key difference was in how Jordan did not know what to do but stayed in that confused state, whereas Timothy in particular engaged in practices that helped him go from not knowing what to do (same state as Jordan) to figuring out what to do. Awareness that one's attempt is not going so well is the first step; using that effectively to get oneself unstuck is the next.

Development C: Exploring and Monitoring

Working without already knowing how a solution would go was another development seen across participants. Students were used to tasks in their past mathematical courses, from K-12 through calculus in college, that lend themselves to clear methods and procedures upon reading the task. As will be discussed in a latter chapter, students also found it satisfying being able to see the entire solution path ahead of time.

But during the transition to proof course, students became stronger at careful, intentional "winging it" – working and exploring without knowing what will happen in advance and noticing when a key piece of information for constructing the proof arose. Rather than remain stuck and wait for the solution path to materialize in one's head, it can be better to start working and see what comes up. This practice is about effectively managing oneself when there no clear strategy is apparent.

Four of the eleven students showed growth along these lines. I only discuss one student – Amy – for this development because she served as a representative for the changes seen in the participants analyzed. But more so, Amy is a case of a student who was high achieving from the start and did not change much throughout the interviews. Recall from Table 6.2 that she got 5 tasks correct and 2 partially correct, out of 8. Her performance therefore already had little room to grow, but moreover, her approach when stuck did not undergo serious changes – except for one singular change described below.

Amy considered herself as a planner, always thinking ahead. Over time, she became more comfortable with working without a specific strategy in mind. She was a strong performer and confident in the class from the start, often finishing tasks quickly. Amy was outwardly confident in her mathematical ability and oftentimes saw how to do tasks right away. For example, she wrote her proof for Interview 2 – Task 1 in under four minutes. While discussing other things at the end of the second interview, Amy said this about herself:

Amy: I just plan, I don't know, I plan everything super far in advance.
Interviewer: Oh, okay. So when you go in...
Amy: I just feel like for everything, I just look ahead. Even when I'm doing math problems. I just like, in my brain, I think about what I'm gonna do before I start doing it.

Amy specifically noted that this was how she did mathematics, always planning out her mathematical actions in advance and thinking ahead in the problem.

But even with her disposition towards planning, Amy became comfortable with working on her feet as interviews progressed. In Interview 4 – Task 2, she decided to use proof by contradiction so that she could work with the "≤ 2" part of the conclusion but became stuck briefly after that because she did not know what to do now. She said out loud that she did not have a plan while working but that she would figure something out (see Figure 6.13).

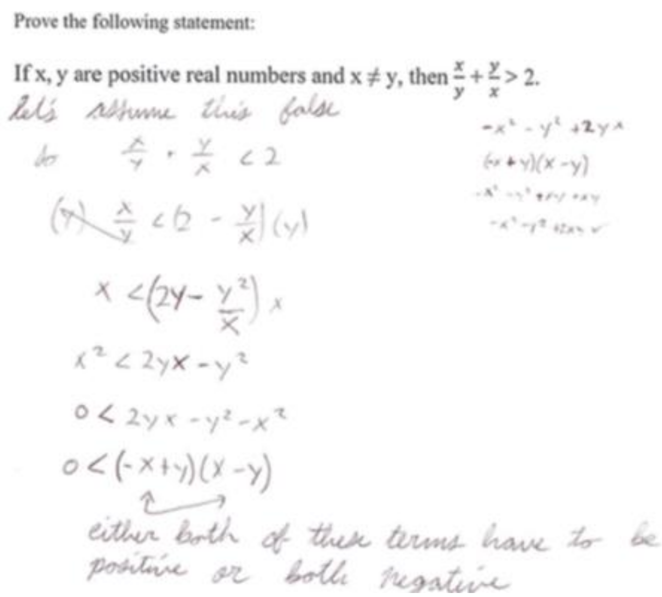


Figure 6.13. First half of Amy's proof for Interview 4 – Task 2

Amy intentionally chose to explore the mathematical situation, manipulating the equations algebraically, with no clear purpose. This proved fruitful, as she noticed the contradictory nature of 0 < (-x+y)(x-y). After this, the rest of her work argued why this was an impossible situation for (-x+y) and (x-y). Thus, she had found a contradiction, as shown in Figure 6.14.

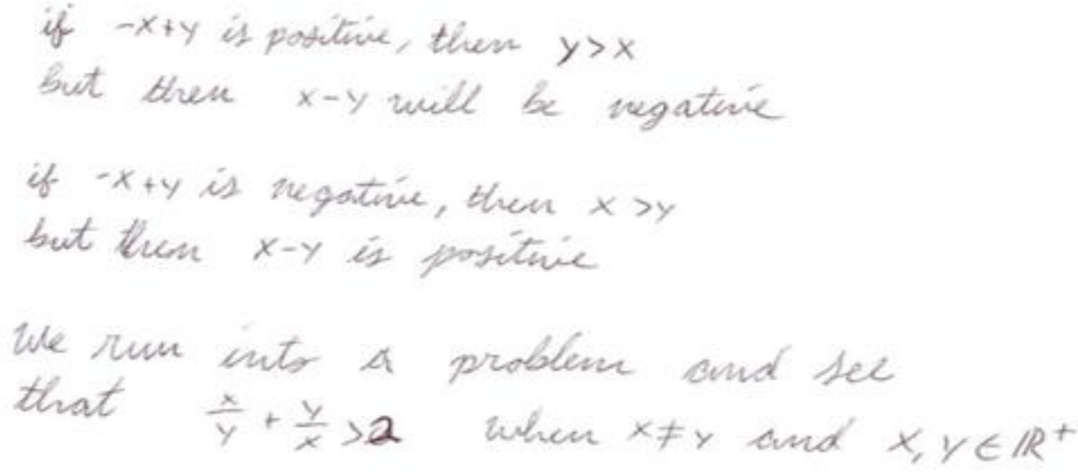


Figure 6.14. Second half of Amy's proof on Interview 4 – Task 2, sans the final lines.

During the debrief, she talked about what was going on when she was stuck early on:

> *Interviewer:* Okay, are there any points in this problem where you feel like you got stuck? That you'd call stuck?
> *Amy:* I feel like this whole portion, I was kind of stuck, but I was just like, "Just check through the algebra until you can get to something." I was like, "I don't see this going anywhere, but I'm sure it will. Just keep going."

Her proving process on this task showed how she did not know at the beginning what she was going to do but was able to roll with the punches. She worked without a specific goal in mind and when a potential avenue appeared, she pursued it and found the contradiction.

Summary. With Amy, she moved from planning out steps ahead (based on her own words) to being comfortable exploring and monitoring her work when unsure what to do. The important thing was her noticing an insight when it arrived. Some of this may have been due to the difficulty of proof tasks; these tasks were no longer so easy that their solutions could be seen right from the beginning (compared to traditional K-12 math), so some level of working without knowing what will happen is part and parcel of a true problem in proving.

Development D: Using Examples to Get Unstuck

As this analysis has focused on what students do when stuck, it is sensible to pose the question: Do students develop effective ways of becoming unstuck? One could say this is a hidden goal of a transition to proof or any problem solving course. A productive practice specific to mathematics emerged during some of the latter interviews, where student would check examples as strategy located within a temporal string of strategies. Three participants - Charlie, Granger, and Timothy - showed this behavior during interviews, through a cursory analysis. I present Timothy here as a representative, to showcase how example checking was used to become unstuck.

Timothy developed a robust practice of using examples when stuck, over the course of interviews. In an earlier section, I indicated Timothy was not happy with his first solution to Interview 2 – Task 1. As a result, he was silent for some period of time, which I interpreted as some version of being stuck. He looked back at his work, reasoned out loud "What if we assumed k is odd…and odd squared is going to be an odd" and then imagined what would happen. While imagining, he thought of "plenty of examples like this [from class] where you give a generic odd and even value in this case and then solve it out." He then said, "I guess I could look at it a different way" and had an insight about taking even and odd cases on k (see Figure 6.15).

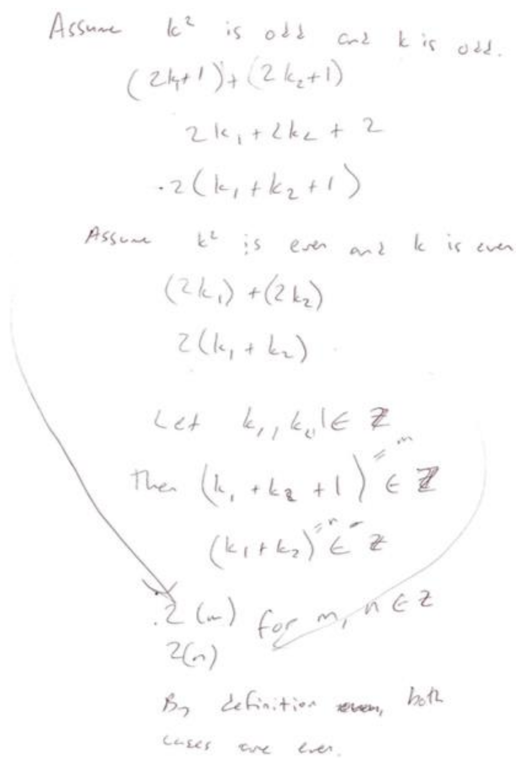
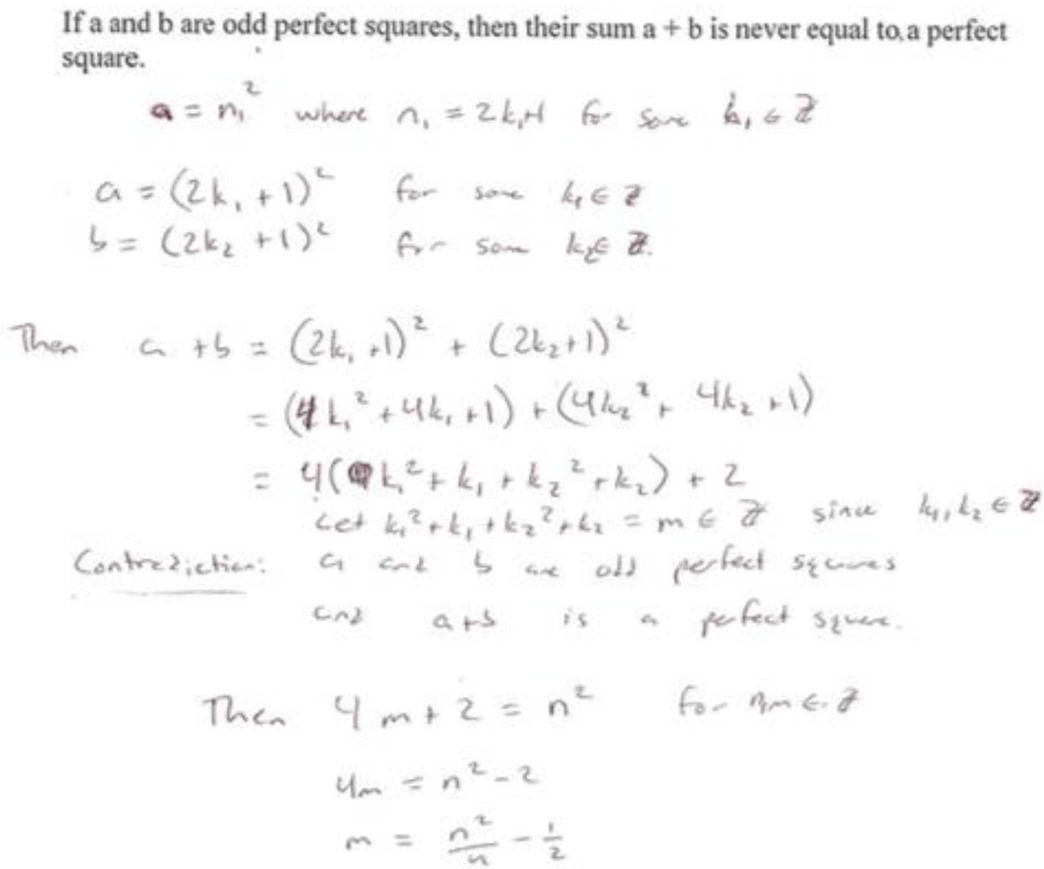


Figure 6.15. Timothy's new solution on Interview 2 – Task 1, after thinking of example exercises from class.

His proof was indeed correct. In this task, Timothy's examples were from example exercises from class; he drew on past mathematical situations he had seen.

The last interview provided a view of Timothy's more typical way of using example checking, however, to get out of tough situations. The last line of his work in Figure 6.16. for Interview 4 – Task 1 shows he got to a formula for m but then became stuck again. This was in fact his fourth stuck point on this task.


If a and b are odd perfect squares, then their sum a + b is never equal to a perfect square.

$a = n_1^2$ where $n_1 = 2k_1+1$ for some $k_1 \in \mathbb{Z}$

$a = (2k_1+1)^2$ for some $k_1 \in \mathbb{Z}$
$b = (2k_2+1)^2$ for some $k_2 \in \mathbb{Z}$.

Then $a+b = (2k_1+1)^2 + (2k_2+1)^2$
$= (4k_1^2+4k_1+1) + (4k_2^2+4k_2+1)$
$= 4(k_1^2+k_1+k_2^2+k_2) + 2$
Let $k_1^2+k_1+k_2^2+k_2 = m \in \mathbb{Z}$ since $k_1, k_2 \in \mathbb{Z}$

Contradiction: a and b are odd perfect squares and a+b is a perfect square.

Then $4m+2 = n^2$ for $m \in \mathbb{Z}$
$4m = n^2 - 2$
$m = \frac{n^2}{4} - \frac{1}{2}$

Figure 6.16. Timothy's work on Interview 4 – Task 1. He reasoned out loud about the implications of the last line.

He then started reasoning out loud: If m has to be an integer, what does n have to be? He mused out loud, let's say n^2 is an integer. He stopped for a moment and "zoomed" out from his work, thinking about what he needed conceptually. He checked an example out loud and noticed $n^2/4$ would have to have a ½ in it, to cancel out the -1/2. He had some realization about why the claim was true, based on his examples, but did not know how to officially show m was not an integer within the allotted time. In the end he gave up, but his work was partially correct and his strategy got him quite close to noticing what the contradiction was in a way that other students did not, that m could never be an integer in this situation.

On the final task of the fourth interview (Task 2), his example checking came to full fruition (see Figure 6.17).

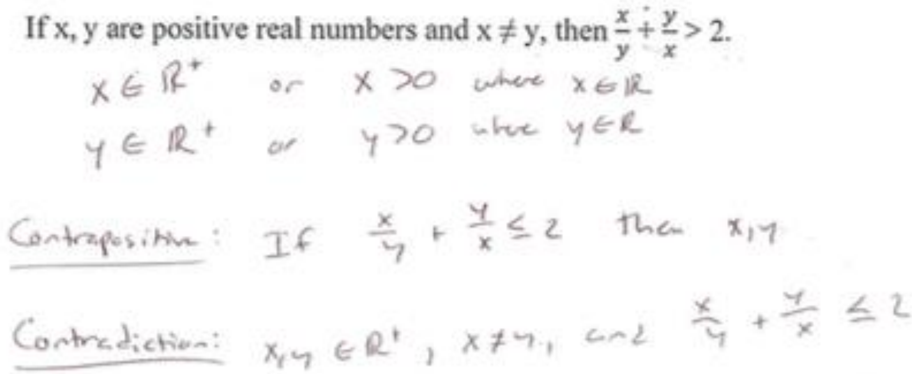


Figure 6.17. Timothy weighing proof techniques on Interview 4 – Task 2

He became stuck on the first item because he was not sure how to negate the statement. He then reasoned out loud what his issue was and possible decisions he could take (see the contrapositive and contradiction set-ups in his work.) This is akin to parallel processing in assessing which of many solution paths is a good idea. He then checked some examples: "I'm just thinking of examples in my head now so like going at it straight, so let's say we chose 1 and 2, so ½ + 2 is greater than 2." He then stopped and switched to contradiction. What he had done in this instance was reason out loud about the issue -> imagine multiple paths -> check examples -> try a different proof method. Ultimately, Timothy's work was incorrect due to multiple algebraic errors, but his approach of using examples in conjunction with other strategies is unaffected by this.

Summary. Timothy developed a practice of what to do when stuck, whether knowingly or not, as interviews progressed. He did some combination of these strategies in this relative order: Look back over work -> reason out loud -> imagine what would happen if certain things were true -> check examples -> have an insight that establishes a direction. It is important to note that Timothy put reasoning out loud to good use here, based on how frequently he became unstuck after doing so. His use of talking out loud stood out across the sample, even when considering that participants were explicitly asked to think out loud.

Timothy used examples as a way of instigating insights, whether intentional or not. He reasoned out loud about what was known, musing about the content at hand, in a sense looking for something to work with. The important thing, however, was that he was attentive enough to notice something when it came up. An example is not a proof, but it can provide an idea for a proof, and he used examples in this nuanced way.

It is of note that this practice may have originated from his instructors. Timothy's instructor, Ms. Frye, reported[3] that she spoke to her class about using examples to get an intuition about why a statement was true but that examples did not count as a proof. Another participant (Granger) also said that Mr. X suggested checking examples as well.

Longitudinal Case: Leonhard

In contrast to the cross-individual discussion of developments, I now present a profile of development by following one individual across the interviews. The purpose of this section is to illuminate what can be gleaned from paying attention to an individual's development. Here I follow the changes seen in Leonhard because over the interview series he showed growth in certain areas – his affect and his proving process – but his performance declined.

[3] (Ms. Frye [pseudonym], personal communication, May 19, 2018)

Following Leonhard's Process for Choosing Proof Techniques

In the beginning, Leonhard's baseline was to choose proof techniques based on what he knew and was familiar with. In Interview 1 - Task 1, Leonhard chose to use proof by contradiction to approach this problem, despite being a little stuck because he was not being sure how to negate the conclusion (see Figure 6.18).

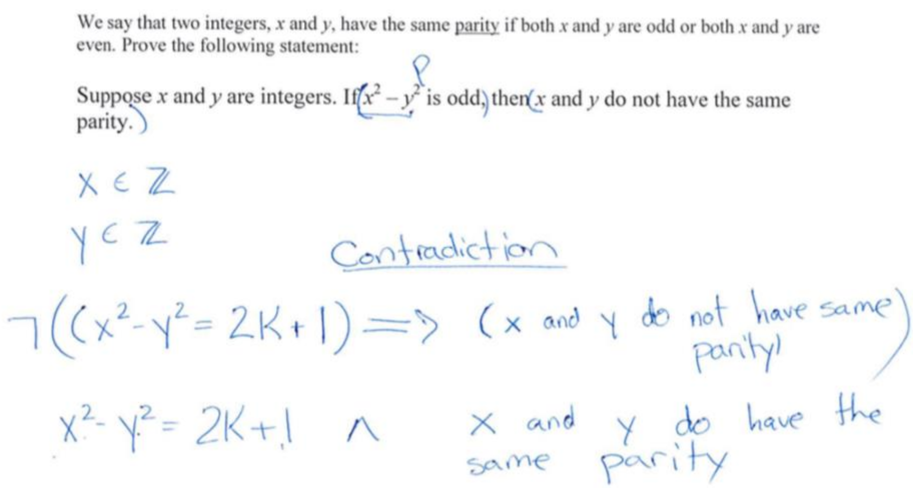


Figure 6.18. The beginning of Leonhard's work on Interview 1 – Task 1

His rationale for that choice was that "A lot of time in class whenever we're proving an implication, we use contradiction I guess so that's why it's my first thought." He used contradiction because that is what they used in class and he was used to it.

In approaching Task 2 of that same interview (see Figure 6.19), he used proof by

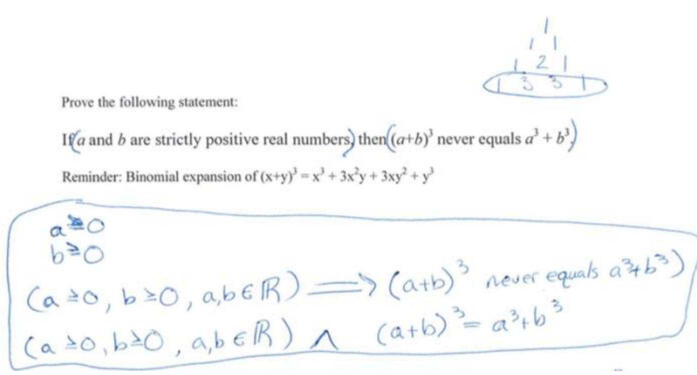


Figure 6.19. Writing out contradiction and not finding it helpful; beginning of Leonhard's work on Interview 1 – Task 2.

contrapositive this time when stuck. At first he wrote the contradiction statement but his explanation was that “I didn’t really know how the contradiction would end up looking and how nice looking it would be to use a contradiction[4] so I used contrapositive.”

He had a sense then that contradiction would not be so “nice looking,” so better to avoid it and use the contrapositive. This could mean that proof by contradiction would not be so clean or would require more work. In fact, Leonhard wanted to use contradiction, as established on the last task as his “go-to” method. Only because he was worried about it did he switch to contrapositive. His move to contrapositive specifically was motivated then but only because it was another technique; his rationale used general terms and he did not articulate it in more detail.

In Interview 2, Task 1, he wanted to do direct proof but became stuck because he was unsure whether what he wanted to do would work. He applied the definitions to x and y and then was stuck again over what method to use, direct proof vs. proof by contradiction. He became stuck again in choosing whether to do direct or contradiction. Ultimately, he chose contradiction and the reason he went with it was: “I decided to do contradiction because I know how to do it.” Leonhard chose what method to use based off what he felt he could do at that point in time, his own sense of fluency with methods and .

---

[4] It should be noted that Leonhard made some errors here: a and b strictly positive means they cannot be 0, so his written work should state that a > 0, b > 0, not “greater than or equal to.” In addition, the negation of “never equals” is not “always equals,” which the equal sign implies.

As time progressed, there was clear growth in his reasoning for his choices – even though his solutions were overall incorrect. Interview 3 – Task 1 is an example where Leonhard cycled through a few options for proof techniques, as seen in his written work (see Figure 6.20).

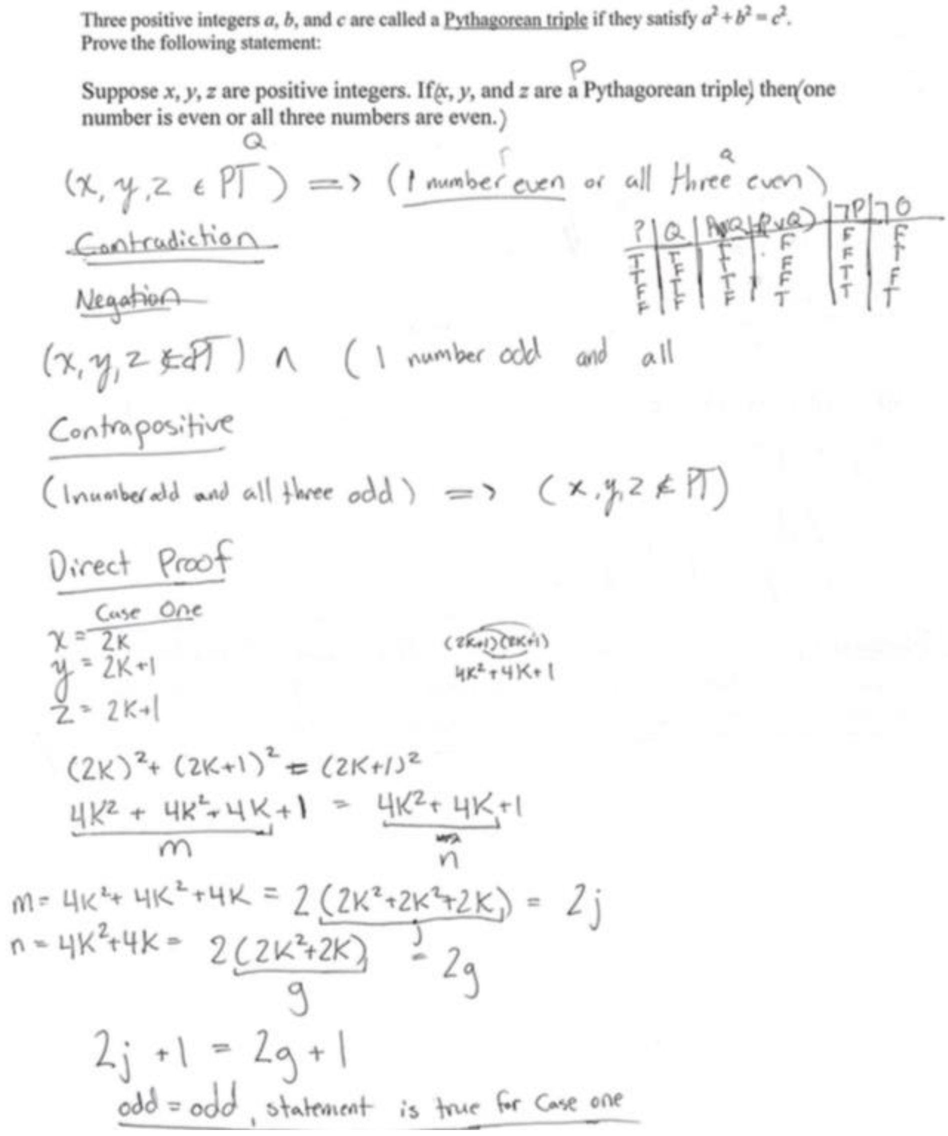


Figure 6.20. Leonhard’s work on Interview 3 – Task 1

He used proof by contradiction but then became stuck in writing the negation, because his negation of the conclusion did not make sense: "One number is odd and all three numbers are odd" did not seem possible to him and so he stops writing after the "all." He had negated the "or" when in actuality it was not a logical "or"; the correct negation is "two or none of the numbers are odd." Then he switched to proof by contrapositive but then realized he had the same issue with how to negate the conclusion, as before. So, he switched again to direct proof. His rationale for why contradiction in the first place was as follows:

> I'm biased towards contradiction so I usually like to do that...my mind goes straight there [to contradiction]. I like it the most because...at some point you usually run into something that just comes out sounding weird. So then you have to be right I guess.

Leonhard admitted that contradiction was his favorite, so he tended to use it whenever he could. He liked it, because of its unique nature in producing something nonsensical. He later added, "I don't know what possessed me to write this [contrapositive]," because he ran into the same issue. Leonhard knew he liked certain methods over others and had some rationale - in how proof by contradiction results in a nonsensical claim and that he should have known to use contrapositive. His rationale was still general, however, in that contradiction was a technique he liked and that his fondness for it drove his usage of it.

Interestingly, he mused out loud about how his underlying idea may have been to check which proof techniques did not work well here and see what is leftover: "I guess this was a good way of crossing out the things that you can't do so you can find the things that you can do." However, the qualification of "I guess" at the beginning of his words suggests we should not put too much stock into this claim about his thinking.

By the fourth interview, Leonhard showed growth in the precision and detail given in his rationales for his choice of proof technique. In Interview 4 - Task 2, he was stuck in the beginning and his subsequent actions were to identify the assumption and conclusion, test a couple examples for x and y, and then try proof by contrapositive (see Figure 6.21).

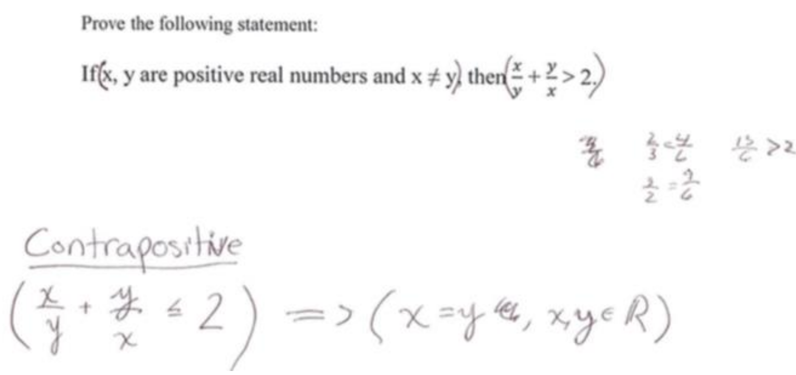


Figure 6.21. Beginning of Leonhard's work on Interview 4 - Task 2

His rationale for contrapositive was, "You can't really do much with x not equal to y. But you can do a whole lot with x = y," and "The contradiction wouldn't give me anything to work with." Leonhard wanted to start with x = y because he saw how equality was more useful than not equal to in proving, and neither direct proof nor proof by contradiction provided an equality. *He decided what proof technique to use based on specifics of the statement to be proven.* In addition, his rationale also explicitly explained why another proof technique (contradiction) would be less useful here. In the end, Leonhard had a rationale for why his chosen proof technique was a helpful approach and why other techniques would be less helpful. In the end, his proof was incorrect, as reaching a true statement (2 ≤ 2) is not the same as showing the conclusion, but his rationale for why use contrapositive was coherent.

Making Sense of Leonhard's Growth

Over the course of these interviews, the rationales Leonhard gave for why he chose the proof techniques that he did became more sophisticated. He moved from choosing certain methods (1) for little to no reason to (2) having some rationale, with a general sense of one technique being better than others to (3) based on the statement itself. Leonhard showed clear growth, yet if we "de-couple" growth from performance, we see that Leonhard's work was oftentimes incorrect (see the excerpt from Table 6.1 below).

Participant	1-1	1-2	2-1	2- 2	3-1	3-2	4-1	4-2	Total
Leonhard	½	0	1	½	0	0	0	0	**2**

Across the interviews, he got 1 task correct and 2 partially correct. Moreover, his work for the last two interviews (four tasks) was all incorrect according to the scoring rubric, due to making substantial errors and/or missing crucial pieces of the proof. Interestingly, Leonhard's perception was that his work was correct on three of these four tasks; he showed great confidence, as can be seen in his emotion graphs for these tasks in Figure 6.22.

Over the interviews, even though his success on tasks stagnated, Leonhard showed progress in terms of affect, of having confidence in his work. There are some good things to this, in how he had a positive orientation towards his work, but it is also worrying when a student does not notice major flaws in their work. Leonhard is an example then of where a student's confidence is high and their reasoning and rationale for their decisions is high, but these do not necessarily lead to correct work.

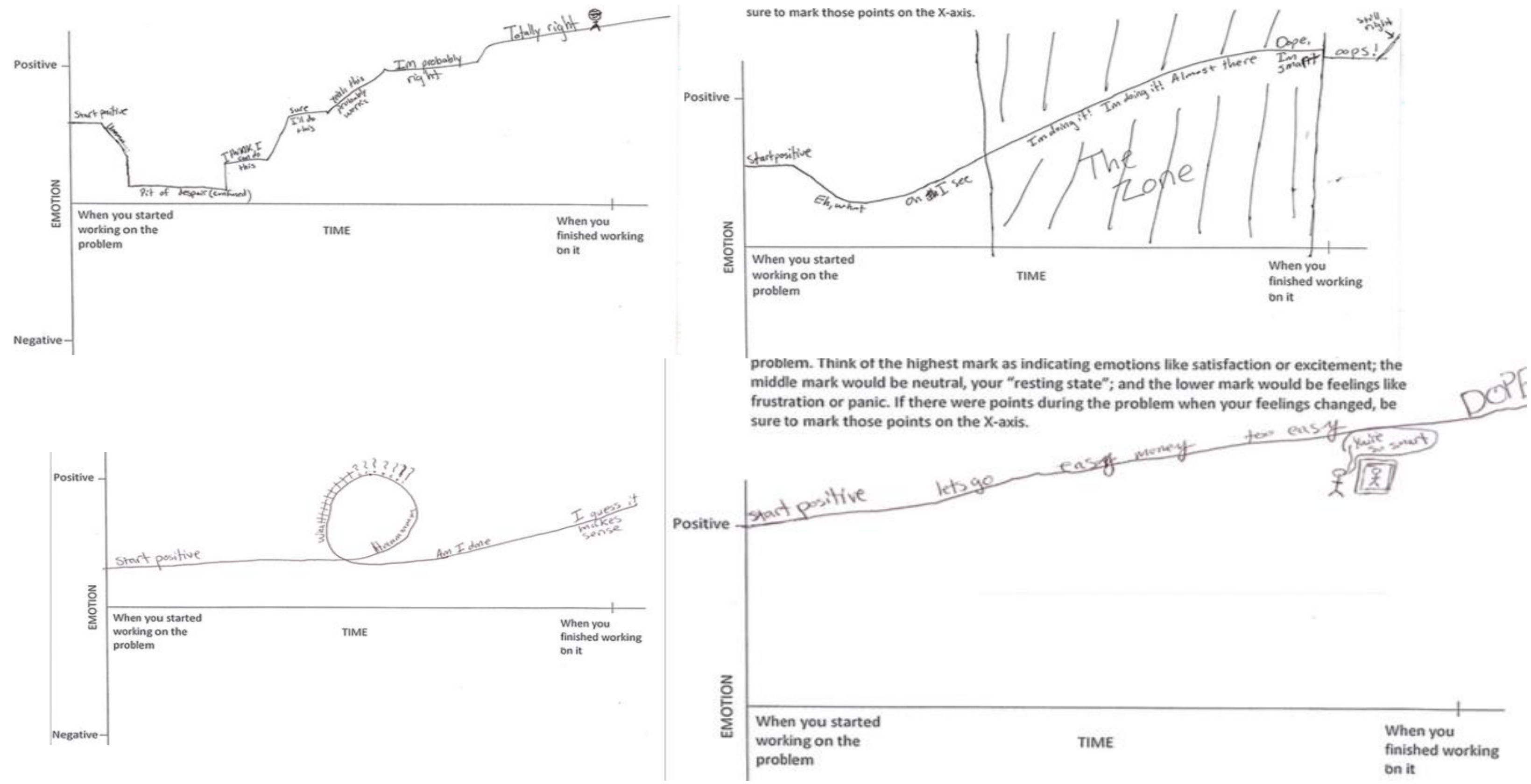


Figure 6.22. Leonhard's emotion graphs for Interview 3 – Task 1 (top left), Interview 3 – Task 2 (top right), Interview 4 – Task 1 (bottom left), Interview 4 – Task 2 (bottom right). His graphs indicated high positive emotions about his work on Interview 3 and Interview 4 – Task 2 but his solutions were incorrect.

There is a difference then between reasoning and execution: Leonhard reasoned well but his execution was flawed. Can we say Leonhard understands contrapositive? Another interpretation of this profile is that progress in terms of process does not always manifest itself in terms of performance, as measured by objective correctness. Judging a student based on solely their written work does not necessarily capture the thinking and reasoning behind their choices that was valid, which alone is valuable growth in proving.

Developments with Limited Data

Here, I talk briefly about some other developments that occurred but were less pervasive across participants. These developments are not particular responses to being stuck but are approaches to proving in general.

Development E: Attending to the Goal

In contrast to working without a plan, some students were more attentive to the goal while proving, as opposed to just working. Charlie only showed signs of being stuck on two tasks of the eight, but one of the tasks serves as a great example of getting stuck and then unstuck. On Interview 2 - Task 2 (see Figure 6.23), Charlie experienced multiple wrong directions and multiple stuck points, due to interpreting the definition of divides incorrectly.

Interview 2 Task 2	Prove the following statement: If a, b, and c are non-zero integers such that a divides b and a divides c, then a divides $(mb + nc)$, for any integers m and n.

Figure 6.23. Statement for Interview 2 – Task 2

During his second “wrong way” (in his words), Charlie set his equations for a equal to each other ($\frac{b}{m} = a = \frac{c}{n}$) but realized what he was doing was not helpful for the goal:

"That is not my aim." In response to this stuck point, his subsequent actions were to start with his goal and work backwards and this lead him to a correct proof.

Charlie attended to the goal even when not stuck. During Interview 3 – Task 2, he thought about whether what he was doing was helpful for what he had to show: "I think this is not a good idea for prov[ing] this [statement]." Indeed, thinking about the goal is something his instructor, Ms. Frye, recommended in class at some point, likely explaining this development.

In contrast, Leonhard had a habit of working on a proof and reaching a true statement, thinking that meant his work was correct. This is an example of how not attending to the goal can lead to incorrect proofs. An example of this occurred on Interview 4 – Task 2 (see Figure 6.24).

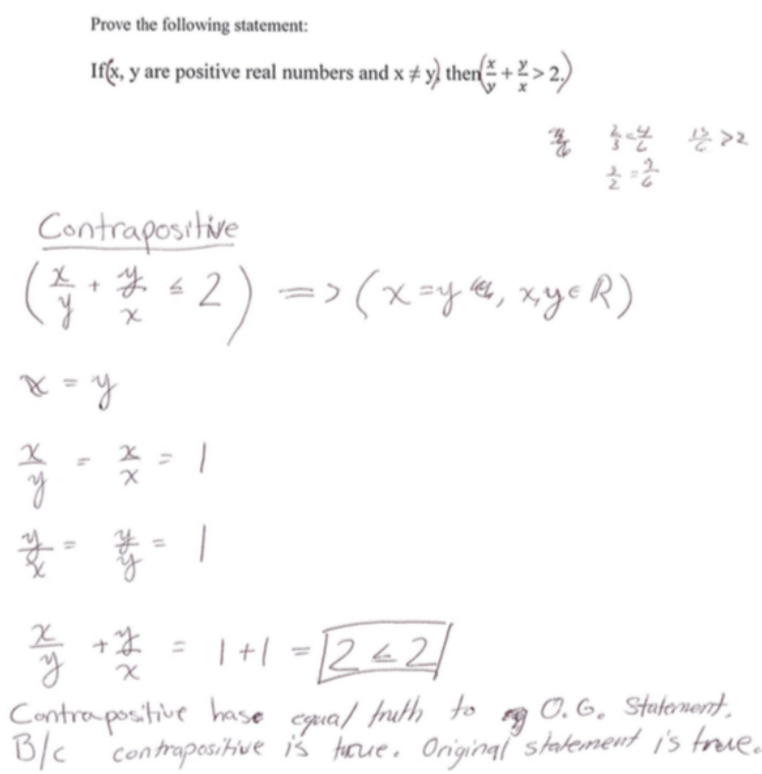


Figure 6.24. Leonhard's work on Interview 4 – Task 2. He reached a true statement and thought he had shown the claim.

He worked until he happened upon a true statement at the end of his work, that $2 \leq 2$, and believed confidently that he had written a correct proof. Reaching a true statement does not mean one has proven the claim, however. Paying close attention to what needs to be shown is important.

Development F: More Systematic Ways of Approaching the Statement

The last observed development I discuss was in how students systematically broke down problems. Here, I provide two cases: Leonard who did this from the start vs. Timothy who did this over time and reported it as an area in which he felt he had grown.

From the beginning until the end, Leonhard had his own process whenever he read a task: Identify the assumption and conclusion, oftentimes assigning them P and Q (as is standard nomenclature). Figure 6.25 shows this consistent practice across interviews.

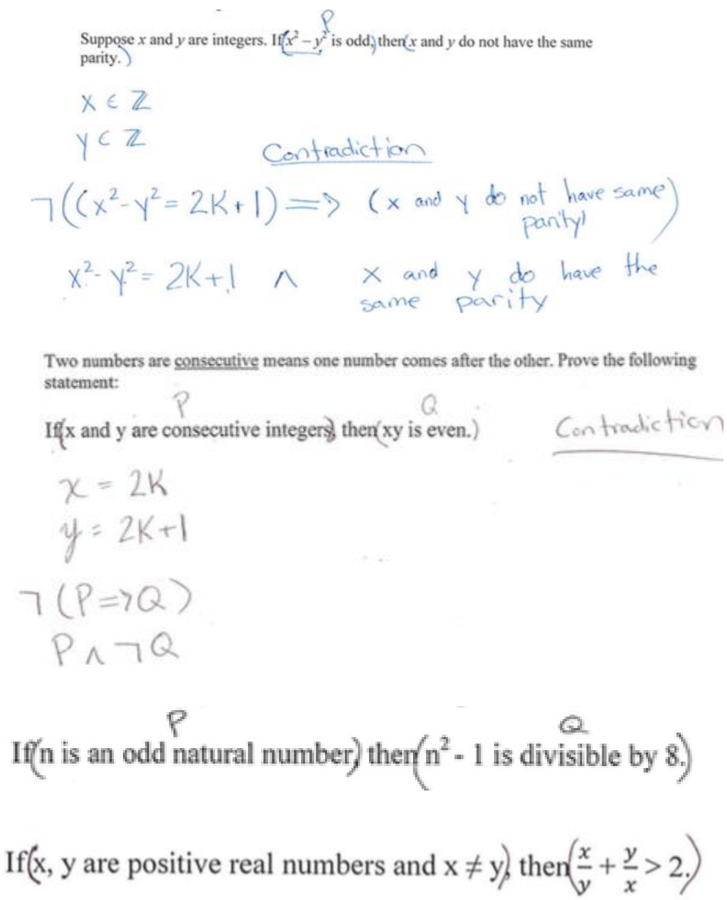


Figure 6.25. Example of Leonhard's systematic approach to start problems, from Interviews 1 through Interview 4. He identified the assumption and conclusion of the statement, using parentheses.

In contrast, Timothy became more systematic in his approach over the course of the class, specifically in how he broke problems down. In Interview 2, he revealed how he was just now "getting" the latest definitions (convergence, open/closed):

> *Timothy:* I felt like a lot of kinda new definitions were kinda thrown at us quickly
> *Interviewer:* Yeah
> *Timothy:* So it was kinda like sorting through and learning each definition kinda one at a time.

He expressed that he was taking his time, because it was a lot of new definitions and information to sort through at once.

However by Interview 3, Timothy said he had noticed changes in himself. He explained that he now knew how to break definitions down, thinking about each part separately. He now looked at definitions and proved them according to the order of the quantifiers that appeared:

> *Timothy:* I look at the definition now and actually try to go quantifier by quantifier like we've been talking about and try to like, so it helps a lot because now I understand why I'm doing the steps in the proofs rather than just like following the rules or whatever. So I think that's definitely like the biggest thing that's changed and it's definitely helped out, just so I can understand...like understanding why you're doing something...helps you do it"

He said that quantifiers made more sense now, whereas before he would follow steps in his notes from class and not really be sure why he was doing what.

> *Interviewer:* Yeah, so before, it sounds like maybe in class, you guys would do like a convergence proof
> *Timothy:* Uh huh
> *Interviewer:* And like you could do it but it sounds like maybe you didn't always know why you're doing it?
> *Timothy:* Yeah, **I'd just try to follow like what we did in class and try to do the same thing and it didn't, especially like convergence, it doesn't apply to every problem the same**
> *Interviewer:* Yeah
> *Timothy:* **So when you know what you're trying to do, it helps out a lot more.**

Systematically breaking down definitions and approaching proving that way helped him understand what he was proving better. Based on how often Timothy appears as a case of development in this chapter, it is clear that he reaped the benefits from this.

Developments Across All Participants

I focused on the four common developments, a longitudinal profile, and two less pervasive developments across my sample. Table 6.4 shows the developments that occurred across all eleven participants, based on interview notes. It is important to note that of all the participants, Timothy seemed to grow the most over the interviews. Some

Table 6.4: Developments in Proving, By Participant

	Changes in how one chooses a proof technique	Harness awareness of how solution attempt is going	Check examples in conjunction with other strategies	Work without a plan	Imagine multiple paths	Approach becomes more systematic	Check work using other methods	Draw on familiar examples	Try multiple methods	Attend to goal
Amy				X						
Charlie	X		X						X	X
Dustin	X									
Granger		X	X	X						
Gabriella	X	X								
Joel		X			X					
Jordan	X									
Leonhard	X				X	X	X			
Stephanie	X							X		
Shelby	X								X	
Timothy	X	X	X	X		X		X		X

participants – e.g. Amy and Granger – started the interviews high performing and so did not show much development. Others, e.g. Dustin and Jordan, became more dejected as the course went on and showed the same one development, in how they chose proof techniques. Future analyses will delve into the other developments listed here and other participants.

Conclusions

The developments shown here can be grouped into three broad categories (see Figure 6.26). *Fluency* refers to students' skill level, e.g. using multiple proof techniques

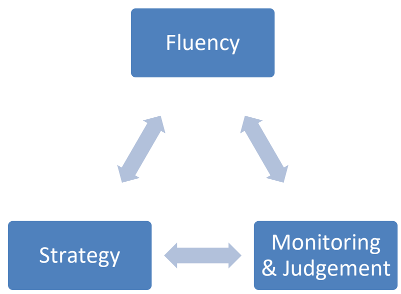


Figure 6.26. Three Categories of Proving Development

and/or wielding them quickly, without struggle. *Strategy* refers to students' intentions in trying to solve a problem. *Monitoring and Judgment* refers to students' ability to pay attention and collect information about how the solution attempt is going (monitoring) and making decisions on what to change (judgment).

It is theoretically possible to have monitoring and not judgment or vice versa. Monitoring without judgment would be being aware that your work is not going well but not knowing why or what to do next. This in fact describes Jordan, who monitored her work but did not use that information to change course when stuck. Judgment without monitoring would be making arbitrary decisions, not based on any of the information from the attempt. This latter idea is difficult to imagine and was not seen in this sample but may be possible.

The fact that these major categories mapped back onto Strategy and Monitoring & Judgment of Schoenfeld's (1985b) components of problem solving was a validity check of my conceptual framing. In addition, even though fluency with proof techniques and logic was not a development I had set out to look for, it is sensible that it showed up here. It is difficult to imagine students showing strategy and monitoring & judgment without some proficiency in proof techniques and logic.

CHAPTER 7: On the Nature of Satisfying Moments

In this chapter, I answer my second research question, *What kinds of satisfying moments do undergraduate students have during the transition to proof?* Informally, this question led to identifying what events felt satisfying to participants, (i.e. led to significant positive emotions), and then categorizing them. First, I discuss the kinds of satisfying moments and how often each kind occurred. Then, I present more focused analyses: combinations of codes that co-occurred together and student profiles of satisfaction.

Identification and Description of Codes

The coding scheme is described in Table 7.1. Each code is a kind of satisfying moment, as emerged from the data or derived from literature. I discuss characteristics of each of the kinds, providing prototypical example(s) from the dataset as needed for the purpose of illustrating what each kind is conceptually. Results of applying the coding scheme will be discussed after.

Table 7.1: Coding Scheme for Kinds of Satisfying Moments

Code	**Sub-code**	**Keywords in Data**	**Criteria for Code**	**Example**
External			Satisfaction is about some external element to the individual, such as a task or situation.	
Completing Task(s)		Figure it out No stuck Know how to do	Figuring out how to do something, typically a mathematical task. Emphasis is on the <u>accomplishment</u> of solving a task. Excludes struggle.	
Overcoming Challenge(s)	Present	Struggle Stuck Can't see/do at first Hard	<u>Struggling</u> on a task and overcoming it. This includes problems that are perceived as hard to the participant.	Anything where you struggle first and then figure it out (Jordan-3-1)
	Comparison to Past	Something I struggle with Not good at X	Present day struggle and overcoming it is set against the backdrop of a <u>previous struggle on a similar kind of task or situation</u>. Participant compares two time points: the present to existing history.	Getting one side of induction to look like another, something she struggles with (Stephanie-1-2)
Partial Progress		Better Improvement Best I can	<u>Incremental or partial mastery</u>. Includes improvement and doing better than I did before or to the best my present capability.	Understanding a problem better (Gabriella-3-1)
External Validation	Grades	Self and Authority Points/Full credit	Receiving good scores, grades, or other outcomes as the source of satisfaction.	Getting good grades (Jordan-1-1)
	Assessments	Self and Authority Exam/Mini-exam	Doing well on a significant assessment, specifically an exam. Excludes homework.	Didn't get stuck on mini-exam (Jordan-1-2)
	Authority Figures	Self and Authority Praise	Authority figure (often instructor) giving praise to the person specifically, e.g. saying work looks good.	TA saying her work was "perfect" on a hard problem she worked on by herself (Amy-3-1)
Internal			Satisfaction is about some internal state.	
Understanding	General	Making sense Understanding (I) get this	Understanding <u>how</u> or <u>why</u> something works, usually a concept, task, or method; a sense of things falling into place or order.	Understanding real analysis because it's understanding a concept (Jordan-4-1)

Table 7.1 (cont'd)

	Aha Moment	Realize Turning point Enlightenment/Revelation "Clicked"	A singular <u>moment</u> of mathematical understanding. Often characterized more intensely as realization or insight.	Instructor's other explanation clicked for him, a revelation (Granger-1-1)
	Seeing the Solution	See	Using the word "see" as a visual metaphor for knowing the solution path or what to do.	Felt good about this proof, could see it (Dustin-2-1)
Internal conviction		Know it's right	Expressing personal conviction in the veracity of one's work, i.e. that they have the <u>right</u> answer or what they found was true	Knowing he'd gotten it right... before getting the grade (Timothy-4-1)
On my own		By myself On my own No help	Doing something in present time on their own, <u>without any help</u> (people, notes, etc.). This idea has to be explicitly expressed by participant.	Getting a homework problem right all by myself (Jordan-1-3)
Properties of Math			Satisfaction is located within the <u>mathematics</u> itself by the participant.	
Useful		Applies Universal	This technique or way of thinking is useful for other problems, e.g. applies to another problem.	Learning the method & applying it to another problem (Granger-3-1)
Simple		Simple Easy Familiar	Task marked by a sense of <u>ease</u> and effortlessness. This can be throughout the entire time or a task becoming easy after an event.	Questions that are easy, simple (Charlie-4-1)
Interactions with People			Satisfaction comes from an <u>interaction</u> with other people specifically.	
Social Comparison		Self VS Others Only one/me Doing better at X than others Proving people wrong Compared	An interaction of an adversarial nature among peers, e.g. involving competition. This code includes situations such as: • being the first or only one to do/know X • being/doing better at X than others • proving other people (classmates, authority figures, etc.) wrong	TA said no one would get it…She was the only one in class to get it (Amy-4-2)
Friendly Interactions		Self AND Others Helping Explaining to others Contributing "if they can do it, so can I"	An interaction of a non-adversarial nature with peers, often helping or working together. Examples include: • helping, teaching, or explaining to others • contributing or debating ideas • vicarious experiences	Being able to explain a problem to someone else such that it makes sense to them (Jordan-1-4)

External Codes

External codes were situations in which satisfaction was about some external element to the individual, such as a task or situation.

Completing Task(s). An instance was coded as *Completing Task(s)* if it referred to figuring out how to do something, typically a mathematical task. The emphasis in these cases was on the accomplishment of solving a task. This category included situations like "not getting stuck on a problem." This category excluded reference to struggle, so the codes *Completing Task(s)* and *Overcoming Challenge(s)* were mutually exclusive.

Overcoming Challenge(s): Present and Comparison to Past. In contrast to *Completing Task(s)*, this category contained all instances that described a challenge, in that there was direct reference to an obstacle or struggle. This category is essentially a more problematic version of *Completing Task(s)*. Tasks that were talked about as "hard" fell under this category.

It is important to note that the sense of challenge was specific to participant and their relationship to the task at hand; the same task was a challenge to one student and not to another. In some cases, participants used terms like "difficult looking problems," which implied that perhaps the task was not personally challenging to them but appeared challenging. In these cases, I looked at the associated emotion graphs, and the graphs started with negative levels of emotion. Hence, data triangulation with the emotion graphs in these instances showed that "difficult looking problems" were perceived as challenging when students experienced negative emotions at the start.

Within this category of *Overcoming Challenge(s),* two clear subcategories emerged. Some instances referred to occasions where participants discussed facing a challenge in the present time (*Present* challenge), whereas other instances referred to a history of challenge or struggle on a similar type of problem (*Comparison to Past*). A common example of the latter subcategory was working on a kind of problem that has been a struggle in the past. If the instance referred to a past struggle, I coded it as *Comparison to Past*; if not, I coded it as *Present*. Therefore, *Present* and *Comparison to Past* are mutually exclusive within the *Overcoming Challenge(s)* category.

Partial Progress. The idea of incremental growth or experiencing progress as good is a common idea (Dweck, 2006). The criteria for Partial Progress is incremental or partial mastery, including references to improving or doing one's best. Common keywords include: improvement, progress, and "best I can."

External Validation: Grades, Assessments, and Authority Figures. In contrast to satisfaction coming from an internal sense of accomplishment of a task, External Validation is about outside sources determining one's success. External validation is essentially a form of extrinsic motivation, where the motivation to do something comes from outside rewards (Middleton & Spanias, 1999). I conceptualize External Validation as taking place between the participant and some authority, whether that authority be a person or an assessment.

Because external validation as a concept can be quite broad, I separated three types of external validation into individual sub-codes: Grades, Assessments, and Authority Figures. These codes were not mutually exclusive, in that an instance could fall under multiple of these sub-codes. An instance was coded as Grades if satisfaction came from external performance, a common one being receiving good grades. An instance was coded as Assessment if the instance referenced a significant assessment, constrained here to an exam or mini-exam. Even though homework was also an important assessment in the course, I excluded homework from this category because homework was frequent and therefore a normal occurrence to the student, whereas the mini-exam and exam in this course happened less frequently. In addition, I wanted to separate out significant and rare events, such as exams.

While seeming quite similar, Assessment was different from Grades in that Assessment concerned the background in which a satisfying moment took place whereas Grades focuses on the outcome as satisfying. Another difference is that knowing how to do something on an assessment served as an indicator of mastery to participants, irrelevant of the actual grade assigned. This did mark a slight departure from my principle to code only what the students themselves verbally mark as contributing to satisfaction. Instead, if the instance took place during an assessment, I coded the instance as Assessment, regardless of whether the participant explicitly referred to the exam context being a factor. The reason for my departure from sticking close to the participant's own interpretations was that I thought this was a situation where participants may not explicitly say "this was satisfying because it was on an exam."

A third sub-code was Authority Figures, was assigned when a person of authority offered praise or other forms of validation to the participant. People of authority tended to be instructors or teaching assistants for the course. An example of this was, "answering instructor's question in the way he was looking for because your correct response means you understand the topic and validation from him (Stephanie-1-2)". Instances that included talking with an authority figure but where that person was left in the background and not the foreground of the experience were excluded from this category.

Internal Codes

Internal codes were situations which did not refer to some objective external element, like a problem, but instead were more internally located to the participant.

Understanding: General, Aha Moment, and Seeing the Solution. Understanding as a whole can be a source for mathematical beauty (Sinclair, 2006). One can think of understanding resulting from "things falling into place." However, the difficulty here is that the term understanding on its own can take on multiple meanings. Just the words "Understanding a question" can vary in meaning: making basic sense of what a question is asking, knowing how to do the question, or grasping how the concepts in the question relate to each other. What students mean when they use the term "understanding" did not necessarily match what mathematics education researchers mean by the term. It was difficult to tease apart these different meaning, especially knowing vs. understanding. For example, in "understanding what he did wrong (Granger-2-3)," did understanding actually mean knowing? Because of these difficulties, I decided to not tease apart these different meanings into separate sub-codes. I felt very little would be gained conceptually by separating *Understanding* into multiple sub-categories.

Two ideas were however worthy of being separated into sub-codes: *Aha Moment* and *Seeing the Solution*. *Aha Moment* captures those experiences of understanding with a short temporal duration, often characterized as a realization, a revelation, or something "clicking." I decided that this idea was worthy of its own sub-code because (a) temporal duration can be inferred from how participants discussed their experience and (b) sudden rushes of understanding as satisfying was an idea present in mathematical beauty literature (Sinclair, 2006). *Seeing the Solution* was generated by noticing that several instances referred to situations where participants were working on a task and then could "see" clearly the line of reasoning or what to do next in order to solve the task. *Seeing the Solution* is a metaphor for things coming into focus, borrowing from visual language. Participants specifically used the word "see," as shown in some examples below:

(1) Felt good about this proof, could see it (Dustin-2-1)

(2) Saw how to do induction problem on exam, which he initially thought he could do it but had gotten stuck using his usual ways (Joel-4-2)

(3) When someone says the problem in a way that makes it click for her, making it easier to visualize the situation and see if it's true (Shelby-1-3)

I did not code instances as *Seeing the Solution* if (a) the word "see" could be substituted by "realized" or "know how to do" without loss of meaning and (b) there were no other indications by the participant that the word "see" was central to their experience. The second instance above, from Joel, is an example of where "saw" could be replaced by "realized," but he spoke emphatically about the seeing the path forward by working backwards, so the word seemed central to his experience.

Figure 7.1 shows the relationship among the subcategories of *Understanding*.

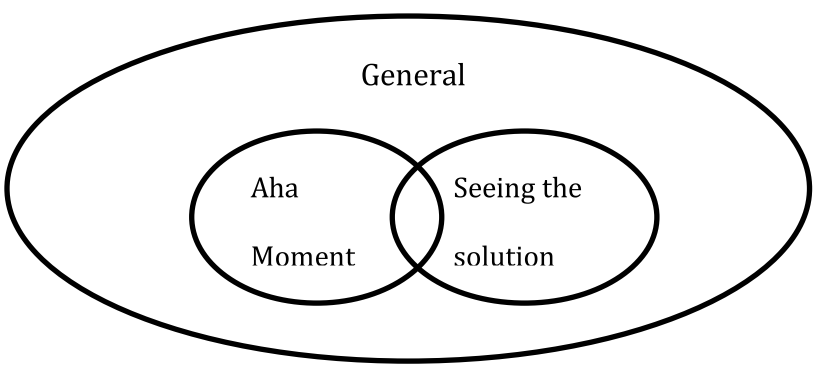


Figure 7.1. Representation of how the *Understanding* sub-codes are related. *Aha Moment* and *Seeing the Solution* are different constructs but can overlap. *General* accounts for all other kinds of understanding.

Aha Moment and *Seeing the Solution* are not mutually exclusive, as there can be overlap. *Understanding: General* was used if *Understanding: Aha Moment* and *Understanding: Seeing the Solution* did not apply (*Understanding: General* is mutually exclusive with *Aha Moment* and *Seeing the Solution*) except in rare circumstances when the satisfying moment involves multiple types of understanding. One such instance was "Figuring out specific moves on a proof he couldn't see how to do at first: cases and factoring. Made sense (Dustin-4-2)." Here the satisfaction came from "seeing" the mathematical moves that complete the proof but also that they make sense in a more general way, so it fell under *Understanding: General* and *Understanding: Seeing the Solution*. Another instance of this was "When he finished a convergence question (which he's generally not comfortable with, unsure what to do), looking back on the proof as a whole, seeing it made sense. Had a realization that was a turning point in the problem (Timothy-2-1)." This instance fell under *Understanding: General* and *Understanding: Aha Moment* because there was a realization while proving but also general sense-making from looking over his proof as a whole after finishing it.

Internal Conviction. *Internal Conviction* refers to the participant's own sense that what they have done is correct or true. The instance that best illustrates this concept is "Knowing he'd gotten it [a convergence problem] right away during the exam before grade came back (Timothy-4-1)." The satisfaction occurred during the situation itself, immediately knowing that he had found the answer; he did not need a grade in order to know it. Another example of this was "didn't get stuck anywhere on problem, 'I know this is true' (Charlie-3-1)." The phrase "I know it's true" spoke to a certain immediate conviction and content with his work. The key word for this category was therefore "know it's right/true."

This code originated from the data itself, but the notion of conviction has ties to mathematical beauty. Conviction in finding the correct answer may be related to a confirmation of one's own ability, that a person was indeed able to find the right answer – a sense of "I did it." Conviction in the truth of one's mathematical work, however, may be more related to mathematical beauty.

On my own. This code refers to a participant doing something on their own or without help. Common keywords include: by myself, himself/herself/ourselves, on my own, no help. Instances had to refer to events that had already happened in order to be included in this code; future expectations were excluded (e.g. believing they can do this problem on their own next time).

This code was difficult to apply because students were very often working alone in many of the experiences they discussed. I decided to only code an instance as "on my own" if the participant explicitly used those words or some of the other keywords, rather than code all instances where the participant happened to work alone. This decision aligns with my principle of coding the participant's interpretation of what was satisfying, not mine. My assumption was that if accomplishment happening on their own was vital to what felt satisfying, the participant would use those words. For example, "didn't get stuck on mini-exam (Jordan-1-2)." technically involved the participant doing something on their own, because it was an exam and exams are solitary efforts in this course. However, Jordan emphasized that the heart of the satisfaction was not getting stuck, not that she was able to do this on her own., so this was not coded as *On my own.*

Properties of Mathematics Codes

This cluster of codes refers to times when participants spoke to techniques, methods, or the mathematics itself as being satisfying. This is in contrast to prior categories which involved relationships between the mathematics and the participant. Codes about what properties of mathematics are commonly seen as beautiful were added from the existing literature at one point, but after a test coding, most of them were removed because they had no instances. The only two concepts that are properties of mathematics originated from the data itself and are described below.

Useful. This code refers to a mathematical technique or way of thinking as being useful for doing other tasks in the future. An example of an instance that spoke to the idea of utility was "Learning the method and applying it to another problem (Granger-3-1)." This category had a large overlap with the notion of mathematical utility or universality, which is a common characteristic of mathematical beauty (Sinclair, 2006). Common keywords for this code were "applies" and "universal." This category could have been called Applicability or Universality, but the term *Useful* captures the utilitarian and practical nuances as well.

Simple. Simple captured instances when a participant emphasized satisfaction coming from the simplicity of a task or when a task becomes simple. Common keywords include "simple," "easy," and "familiar." This code is characterized by a sense of ease in doing the task such that it is effortless. This is a code that comes from mathematical beauty, the long-held idea that simple mathematics is beautiful (Hardy, 1940; Wells, 1990).

It could be argued that talking about simple tasks actually refer to the relationship between the solver and the task, as what one person finds simple another person may not.

However, many of the participants talk about the mathematics itself being simple, so I am sticking close to their interpretation in considering Simple a property of the mathematics. At first, one may think that this category should fall under *Completing Task(s)*, because simple tasks tended to mean the participant did not get stuck. While this was true, this category also includes experiences where a task became simple, i.e. the task was challenging and then something happened where it became simple and was from there on easy or clear to do. Because of this variation, Simple exists as a category that can factor across Completing Task(s) and Overcoming Challenge(s).

Interactions with People Codes

These codes referred to experiences where interactions with other people were described as satisfying. In order to fall into this general cluster of codes, the participant had to refer to the interaction itself as being satisfying. For example, a clause like "the rest of the group didn't get it" references people but was used just to set up that the mathematical task was a challenge. An instance with this clause would therefore not necessarily count as an interaction then.

Social Comparison. This category refers to satisfaction that comes from comparing oneself to others – and coming out ahead. Here, the comparison is between the self vs. peers. Social comparison also has links to self-efficacy, as a type of vicarious experience (Bandura, 1977). Typically, social comparison has an adversarial or competitive nature and involves at least one other person. The three main situations that occur are:

- being the only or first one (compared to others) or to do/know X
- being or doing better at X than others
- proving other people (especially classmates or authority figures) wrong

While many of these situations had challenges inherent in them, instances were coded *Social Comparison* when it was the interaction or the comparison with others that was satisfying. If the people in the experience had disappeared, would the experience still be satisfying? If no, then the experience was not coded as *Social Comparison*.

Friendly Interactions. This category concerned non-adversarial interactions, as different from *Social Comparison*. Typical situations that counted as friendly interactions were working with, talking to, or helping fellow students in the course. The core idea of this category was of a person working with their peers, rather than working versus peers. Instances where the satisfaction came from interactions with instructors or TAs did not fall in this category because they were authority figures, so they would be categorized as *External Validation: Authority Figures.*

Applying the Coding Scheme to the Data

After finalizing the coding scheme, I then went back and uniformly applied the coding scheme to the dataset. The results of the coding process are given in Table 7.2, from most to least frequently occurring kinds of satisfying moments. Instances contained multiple kinds, so each instance could and frequently did take on multiple codes. The average number of codes per instance was 2.3 codes, so each instance on average had 2-3 codes assigned.

Table 7.2: Frequency and Percentage of Satisfying Moments by Kind

Kind	**#**	**% (of N = 75)**
Overcoming Challenge(s)	37	49%
Present	26	35%
Comparison to Past	11	15%
Understanding	34	45%
General	23	31%
Aha Moment	8	11%
See the solution	7	10%
Completing Task(s)	21	28%
External Validation	18	24%
Grades	8	11%
Assessments	10	13%
Authority Figures	3	4%
Interactions	16	21%
Friendly Interactions	12	16%
Social Comparison	6	8%
On my own	13	17%
Simple	11	15%
Internal Conviction	5	7%
Partial Progress	4	4%
Useful	2	3%
Total Codes Applied	169	-

Note. This table lists the percentage of each code out of N=75 satisfying moments. The codes are listed from most to least frequently occurring. Data were frequently assigned multiple codes, hence percentages do not add up to 100%. The total number of codes applied is included in the last row.

Based on percentages across the full dataset, the most common types of satisfying moments were the following: *Overcoming Challenge(s), Understanding, Completing Task(s)*, and *External Validation.* Each of these accounted for more than a 20% share of the data. A second tier of codes captured at least 10% of the dataset: *On my own, Friendly Interactions*, and *Simple*. *Social Comparison, Internal Conviction, Useful, and Partial Progress* each applied to less than 10% of the data.

Overcoming Challenge and Completing Task(s) Account for a Large Portion of Data

Not surprisingly, many satisfying moments (49%) involved overcoming challenges. This finding is sensible, because the course was designed to be challenging for students and thus accomplishing a challenge can make an experience out of the norm, against one's expectations of what would happen. A little over a third of the data – 35% - concerned present challenges, making this the larger of the two sub-codes. Nevertheless, 15% of the data involved the comparison of the present to past challenges. This result suggests the importance of a person's history, that past experiences can influence the satisfaction of an experience, especially overcoming long-standing struggles.

The fact that experiences with a lack of challenge were satisfying too (*Completing Task(s)*) is also not surprising and confirms informal observations that as educators we grapple with: Students often enjoy tasks on which they do not have to struggle. In other words, students enjoy exercises. Accomplishments absent of struggle also confirms the importance of mastery experiences, from self-efficacy (Bandura, 1977).

Experiences that involved both overcoming challenges and completing task(s) accounted for 77% of the entire dataset. This high percentage suggests that satisfying moments tended to be about mastery, regardless of whether or not there was struggle.

External Validation vs. Understanding: Unexpected Results

External Validation was a code I expected to account for a large part of the dataset because of the emphasis on grades and performance across society. Indeed, 18 instances - about a quarter of the dataset - fell under *External Validation*. Of those, eight instances were about grades, nine took place on assessments, and two were about authority figures. There were only two instances that were both about grades and assessments:

- She’s not good at showing convergence but did it on the exam, felt it was the best convergence proof she'd ever done, and got full credit for it (Amy-4-1)
- Scoring well on the mini-exam (Leonhard-1-1)

The overlap between *Grades* and *Assessments* within *External Validation* was therefore minimal, considering how tightly grades and assessments are intertwined.

While *External Validation* did account for a little over a fifth of the data, it was not the most common kind of satisfying moment. *Understanding* took a larger share of the data, a little less than double that of *External Validation*. There was indeed a large amount of within-code variation for this category, from basic sense-making to knowing how to do something to understanding concepts.

History and expectations play a role when it comes to what level of understanding is satisfying, explaining some of the variation seen within the *Understanding* code. Basic sense-making can be satisfying when even that is difficult to come by. For example, Jordan discussed in the fourth interview how her instructor's explanation made sense, to the point that she felt she could do similar problems on her own next time. At this point in time, Jordan was worn down by the course and felt like she was not understanding very much. It is natural then that even just following along with what an instructor said could be satisfying, although this example did have a link to then being able to successfully complete a task.

What are the implications then of such a large portion of the dataset falling under *Understanding*? First, students do find understanding satisfying – they want to understand and when they do, it feels good. This is important for two reasons. One, this corroborates mathematicians writing about understanding as a quality of mathematical beauty and that students experience and appreciate at least some version of this. Even if aha moments occur less frequently (10% of instances here), it is heartening that students find regular understanding to be satisfying. Two, reform efforts in U.S. mathematics education have often pushed for more understanding in mathematics classrooms (NCTM Standards, Common Core State Standards), while counter-efforts have often called for a return to basic skills and facts with the belief that understanding will come later. That students report understanding as satisfying – and are aware themselves that understanding feels good – provides support at the student-level for efforts pushing for more understanding in the mathematics classroom.

Interactions with People: Friendly Interactions

I chose to group both *Interactions with People* codes together in the table, to show that they accounted for 21% of the data. *Friendly Interactions* accounted for 16% of the data, more than *Social Comparison*. The most common *Friendly Interaction* reported was in explaining or helping others with a question (6 of the 10 instances). There were some interesting other cases that fell in this category. For example, Shelby voiced two instances of talking and working with others in the MLC:

- Working with others (not just for getting answers); when student says something that makes problem click, to the point that you can tackle those problems on your own later (Shelby-1-1)
- Talking about math in the MLC with people and writing on the board if they're people who she can bounce ideas off of. (Shelby-4-2)

Other types of friendly interaction came from Gabriella, about debating with others and also this experience from the beginning of the course:

> She struggled the first week, feeling like doesn't understand, can't contribute, and thinking she's the only one lost. But she's ‘getting’ the class now, so that she can contribute in groups and get correct answers. This came from someone else telling her she can get through this course and others are struggling too. (Gabriella-1-1)

A student with the same major as Gabriella who had already taken the course in a previous semester assured her that it was a difficult course but that she could get through it. This is an example of a vicarious experience (Bandura, 1977). Gabriella noticed that someone else similar to her (with the same major) could also get through this helped her.

It is important to note that while these interactions concerned helping one another, personal mastery was still present. When a student was able to explain a homework question to another student and can see evidence that they understand, that served as confirmation of one’s own level of understanding. The frequency of non-adversarial peer-to-peer interactions that appear in the data suggest that mathematics classrooms, even upper-level undergraduate mathematics classrooms, can benefit from facilitating more interactions between students. This course encouraged the use of structures outside the classroom, where students could meet in a common space to work and talk together about homework. Even if take the most cynical interpretation, that students do this for their own sense of mastery and understanding, the conclusion from the data is straightforward: Some students enjoyed helping each other.

On My Own & Simple

On my own and *Simple* were the most common codes among the second tier, capturing 17% and 15% of the data respectively. There was variation within *On my own* in regard to what “no help” constituted. Instances ranged from the basic “getting a homework problem right all by myself (Jordan-1-3)” to not getting help from friends: “Figuring it out by himself before others because he wants to know why it works and it's more satisfying than having someone tell him how it works (Leonhard-1-3).” There was one instance that referred to not using resources: “Did convergence question without notes, after worrying because they hadn't done one [a question] like this in a few weeks (Stepanie-4-2).”

There was variation within what was meant by *Simple* as well. Two instances referred to a task becoming simpler, e.g. “got stuck on a problem he expected to be easy, so thought about it differently and it clicked in a way that problem became simple. Now he knows what to do with these problems (Joel-2-1).” It was the task becoming simple when the previous state was one of confusion, that is important to notice here. In addition, two instances referred to thinking:

- Gets excited about it [induction] now because easier and doesn't have to think (Timothy-1-1)
- Questions that are easy, simple, and/or that he can do just by thinking about them (Charlie-4-1)

The first instance referred to not “hav[ing] to think” as satisfying. It is not surprising that induction is the content mentioned here, as induction has an algorithmic and procedural nature. At face value, this is similar to students finding exercises pleasing. At a base level, the sentiment makes sense; we want to conserve the amount of resources needed to do a task, so when we are able to do a task with little effort, that manifests itself in the form of an aesthetic feeling.

Data that Did Not Fit into the Coding Scheme

One instance did not fall into any category. It concerned writing homework in the typesetting language LaTeX; the participant said it felt satisfying because it was like programming. This suggests something about the writing of a proof, specifically the product, was satisfying. There was an element of this in one instance of Amy's satisfying moments: "proud of writing proof for a lengthy and hard-looking problem (1-1)." Timothy also talked about looking back at his proof as a whole and feeling good about it making sense. There was not enough evidence in the text itself to warrant a code, but there may be a possible code for *creating a product* or *proof as a creation*. This may appeal to a more aesthetic take, of evaluating a product for its beauty. This would need to be explored with more data.

Partial Progress as Rare

There were only two instances that spoke to the idea of partial mastery as satisfying. Both of these came from Gabriella:

- Doing the best she could on the test, knowing how to do most of the hard ones, after instructor says he doesn't expect people to finish, making her emotions drop (2-1).
- Understanding a problem she didn't understand better by talking to instructor, given that she had no idea at first and a TA's explanation didn't help much (3-1).

Typically, when refining a coding scheme, a code containing only two instances in a dataset of this size would be a likely candidate for elimination. I did consider removing the *Partial Progress* code, but this idea of progress could not be subsumed by my other categories easily. Additionally, I think it is telling that incremental growth, which is supposed to be a good thing, *is generally not reported as satisfying*. The two instances reported here came from the same student, which raises questions about its generality. Full mastery and accomplishment, with or without struggle, make up a large share of satisfying moments. This suggests the question – is there something about mathematics as a domain that makes partial mastery not as satisfying as in other domains, such as learning how to play an instrument or running? Even taking into account that satisfying moments are highly personal, perhaps mathematics educators need to underscore to students that partial progress in mathematics is something to be proud of, in and of itself.

Combinations of Kinds

As mentioned previously, getting the codes to a point where the interrelations could be minimized was difficult. It took many rounds of refining the coding scheme to do this. One way is to look at what codes co-occur with each other. In other words, when I label an instance code A, am I likely to also label it code B? This would suggest that perhaps code A and B should be collapsed. However – just because two codes co-occur do not mean code A and B are the same construct (Bakeman & Gottman, 1997). In this section, I discuss codes that seemed to co-occur together but seem to be separate constructs.

My judgments are given further backing from a co-occurrence matrix I constructed. Each cell of this co-occurrence matrix corresponds to a row X and column Y, and each cell represents the co-occurrence of Y with X. Co-occurrence was calculated as the % = all instances labeled code X and code Y / all instances labeled code X. In other words, each cell is a conditional probability: Of all instances of code X, what percent were also labeled code Y?

Completing Task(s) + Simple

One frequent co-occurring pair was *Completing Task(s)* and *Simple*. A third (33%) of *Completing Task(s)* instances were also coded as *Simple* (see Table 7.3) and in turn, 64% of *Simple* instances were also coded *Completing Task(s)* (see Table 7.4).

Table 7.3: Co-occurrence of Codes with Completing Task(s) & Overcoming Challenge(s)

Code	#	External Validation	Internal conviction	Progress	On my own	Understanding	Useful	Simple
Completing Task(s)	21	29%	5%	0%	24%	29%	0%	33%*
Challenge: Aggregate	37	24%	8%	5%	22%	43%*	0%	11%

Note. Only selected codes are shown here in columns. External Validation and Understanding are aggregates across sub-codes.

Table 7.4: Co-occurrence of Codes with Completing Task(s) and Challenge(s)

Code	#	Completing Task(s)	Challenge: Aggregate
External Validation: Aggregate	18	33%	50%*
Internal Conviction	5	20%	60%*
Progress	3	0%	67%*
On my own	13	38%	62%*
Understanding: Aggregate	34	18%	47%*
Simple	11	64%*	36%

Note. Only selected codes are shown here in columns.

This combination makes sense, in that experiences that lack struggle are likely to also be simple or feel effortless to the person. There could be an argument that these constructs are so inter-related that they are the same, but *Simple* also includes instances where challenging problems became simple. So *Simple* tasks as I have defined in this study are not exactly the same as *Completing Task(s)*, but they do tend to occur together.

Overcoming Challenge + Understanding

Another set of co-occurring codes were *Overcoming Challenge(s)* and *Understanding*. Table 7.3 shows that 43% of *Overcoming Challenge* instances involved *Understanding* and Table 7.4 shows that 47% of *Understanding* instances involved *Overcoming Challenges*.

Both of these codes have a large number of instances themselves (n=37 and n=34 respectively).

There were also a number of codes that were unidirectional in relation to challenge, as in many of these codes' instances were also coded as *Overcoming Challenge(s)* but not vice versa. Table 7.4 shows the codes for which many of their instances were also coded as challenges: *External Validation* (n=18), *Internal Conviction* (n=5), *Partial Progress* (n=3), *On my own* (n=13), and *Social Comparisons* (n=6). The frequencies of some these codes are quite small, which may explain why *Overcoming Challenge(s)* accounts for a large share of each. *External Validation* and *On my Own* have double-digit frequencies, so I will discuss them, as an example of this unidirectional relation.

Table 7.4 shows that 62% of *On my own* instances were also coded as *Overcoming Challenge(s).* It may seem at first that doing something on one's own has an equal effect on accomplishment with or without challenge, looking at the similar conditional probabilities (24%, 22% respectively) in Table 7.3. But when looking at the whole of *On my Own* instances (see Table 7.4), 62% were also challenges, whereas only 38% were accomplishments without challenge. *On my own* is therefore not necessary to feel good about doing a challenge, but when someone is proud of doing something on their own, it tended to be something challenging.

Friendly Interactions + Understanding

Friendly Interactions tended to involve *Understanding* and to a lesser extent occurred on non-challenging tasks. In Table 7.5, a third (33%) of *Friendly Interactions* involved tasks without challenges.

Table 7.5: Co-occurrence of Interactions with People with a Selection of Codes

Code	# of items	Completing a task	Challenge: Aggregate	External Validation: Aggregate	Internal Conviction	On my own	Understanding: Aggregate
Social Comparison	6	17%	50%	17%	17%	17%	50%
Friendly Interactions	12	33%	8%	8%	8%	17%	67%

More interestingly, 67% of *Friendly Interactions* also fell under one of the *Understanding* codes. In turn, 24% of *Understanding* instances were about *Friendly Interactions*. This suggests that satisfying *Friendly Interactions* tended to have an understanding component, but satisfying *Understanding* instances did not require peer interaction.

This result makes sense given that the most common *Friendly Interactions* were variations of teaching fellow students and watching them understand. In addition, helping others can produce confirmation of one's own understanding. Working and talking with others can deepen one's own personal understanding. This suggests that one important kind of peer-peer interactions which are satisfying are the ones that provide deeper understanding of the content.

Half (50%) of *Social Comparisons* were also coded as *Overcoming Challenge(s)*, but there were only n=6 instances of *Social Comparison* in the entire dataset. This makes it difficult to make any strong inferences about commonly co-occurring codes for *Social Comparison*. However, if we consider this co-occurrence to be a claim with limited data, it makes sense because social comparison is often adversarial or competitive. Besting others at something difficult can be more indicative of mastery than besting others at an easy task.

Clustering of Satisfying Moments by Individuals

These results also raise the question from an individual-centered standpoint: Do individuals tend towards certain kinds of satisfying moments? In other words, what is the variation of kinds within one person's satisfying moments? In this cursory analysis, I present satisfying moments by participant. Table 7.6 shows the results for each participant, i.e. which codes each of their satisfying moments fell under and how many.

Table 7.6: Kinds of Satisfying Moments by Participant

Participant/ Code	Completing Task(s)	Overcoming Challenge(s)	External Validation	Internal Conviction	Progress	On my own	Understanding	Useful	Simple	Social Comparison	Friendly Interactions	**Total Codes**	**Total Moments**
Amy	1	5*	2			2	1			3		**14**	**6**
Charlie	7*	2	3	1		1			5*		1	**20**	**8**
Dustin	7*	2	2				4		2			**17**	**9**
Gabriella		5*	1	1	2	1	4*				2	**16**	**6**
Granger					1		4*	1			1	**7**	**5**
Joel		4	1			1	3*		1			**7**	**4**
Jordan	4*	1	2			1	4*				2	**14**	**8**
Leonhard	2	4	3	2		2	2		1	2	2	**20**	**9**
Shelby		3	1				6*	1	1		3	**15**	**8**
Stephanie		6*	2			4*	1			1	1	**15**	**7**
Timothy		5*	1	1		1	4*		1			**13**	**5**

Note. Each cell shows the frequency of satisfying moments per participant by code. The * denotes codes that made up at least half of that participant's total satisfying moments. Total Moments is the number of satisfying moments, where Total Codes is the sum of codes applied across all their instances. Empty cells are zeros, which have been omitted for clarity. Overcoming Challenge(s), External Validation, and Understanding are aggregates.

By looking at codes that accounted for at a majority (at least half) of the participants' satisfying moments, four major profiles are revealed, which I discuss below.

Case A: Students Who Enjoy Completing Task(s)

One profile is that of the student whose satisfying moments come mainly from accomplishing tasks that are not challenges, i.e. exercises. Three of the eleven participants fit this profile: Charlie, Dustin, and Jordan. Charlie extended this profile in that simplicity was embedded in many of his satisfying moments: "Truth table elicited positive emotions because it was simple and easy (Charlie-1-1)." and "It's a topic, convergence, that's his strength: simple and didn't get stuck (Charlie-3-2)." Charlie also talked about tasks "that he can do just by thinking about them (Charlie-4-1)." Being able to solve a question in one's head implied not needing to expend effort and also spoke to clarity: An answer that came naturally just from thinking is satisfying. There is an effortless, almost comforting, feeling to the satisfying moments described by not just Charlie, but all three of these participants.

Case B: Students Who Enjoy Overcoming Challenges(s)

The next profile is that of the student who really enjoys accomplishing challenging tasks. Three of the eleven participants fit this profile: Amy, Stephanie, and Leonhard. In Leonhard's case, only four of his nine satisfying moments were about challenge; while that is technically less than half of his instances, he is honorarily included in this category because the greatest share of his instances were this kind.

There are other codes that can go with *Overcoming Challenges* too. Amy focused on *Social Comparison*. During the interviews, it was clear Amy was competitive and especially loved proving people wrong, e.g. "Getting a problem that TA said no one would get (Amy-4-2)." It makes sense that a person who likes challenges would also be motivated by social comparison, as social comparison often has a competitive nature. In Amy's case, sometimes challenge and social comparison occurred in separate instances, but there were instances where they occurred in tandem, so the combination of these two kinds makes sense.

Another kind that co-occurred with enjoying challenges was doing them on one's own. Stephanie liked doing questions on her own, which again was sensible because doing something without the help of others can be thought of as a more general form of challenge. When a student solves difficult math problem by themselves, this can be interpreted as being competitive with oneself, in that they have exceeded their own expectations of themselves.

Case C: Students Who Enjoy Understanding

A surprising profile may be that of students for whom understanding is everything. Two of the eleven participants fit this profile: Granger and Shelby. Both of these students talked quite a bit about understanding the mathematics in their satisfying moments. This is corroborated by how they were also the only participants to talk about usefulness and applicability of certain techniques or methods as satisfying: "Learning the method and applying it to another problem (Granger-3-1)" and "Likes this [table method] because it's...universal, in that she used it on 3 problems this past week (Shelby-3-1)." While there is an element of being happy about having a procedure, these students' instances are fundamentally about liking a certain piece of mathematics for its power to do more.

Case D: Students Who Enjoy Overcoming Challenges(s) and Understanding

As a combination of the previous two profiles, there were also students for whom it both challenges and understanding constituted most of their satisfying moments. Three of the eleven participants fit here: Gabriella, Joel, and Timothy.

For Gabriella, making sense of the question was a big struggle she talked about throughout the interviews. Once she understood what the question was asking, she generally knew what to do. Of her four instances of *Understanding*, one concerned basic sense-making, but the other three were indeed about a deeper level of understanding.

For Joel and Timothy, all but one of their instances concerned understanding, but they also reported a relatively low total number of satisfying moments across the four interviews (4 and 5 respectively). For Joel, two of his four instances were about “seeing” the solution in problems where he was stuck and one was about knowing what to do in the future. While again there is a procedural flavor to Gabriella and Joel’s instances, the sense that the mathematics fell into place for them was apparent. Timothy in fact reported aha moments (not shown in Table 7.6) in three of his five satisfying moments. Understanding can therefore come in different ways: basic sense making, to knowing what to do, to instantaneous realizations that illuminate the path forward. Taking satisfaction from both challenging problems and understanding, *especially* when the understanding comes from working on challenging problems, may serve students well for their mathematical future.

Conclusions

In this chapter, I answered the research question, *What kinds of satisfying moments do undergraduate students have during the transition to proof?* Through grounded theory techniques, I developed a system of kinds of satisfying moments. The most commonly occurring ones in this data were *Completing Task(s)*, *Overcoming Challenges*, *Understanding*, and *External Validation*. The aggregate of interactions with people, both *Social Comparison* and *Friendly Interactions*, also applied to a large share of the dataset.

Additional analyses showed how codes related to each other and which ones stood out. Common combinations of co-occurring codes revealed the following pairings: *Completing Task(s)* & *Simple*, *Overcoming Challenge(s)* & *Understanding*, and *Friendly Interactions* & *Understanding*. Four student profiles of what students most often found satisfying were revealed: (a) *Completing Task(s)*, (b) *Overcoming Challenge(s)*, (c) *Understanding*, and (d) a combination of *Overcoming Challenge(s)* and *Understanding*. Based on all these analyses, accomplishment both with and without challenge, understanding, and working with and/or helping fellow students seem to be major kinds of satisfying moments.

CHAPTER 8: Discussion

In this chapter, I consider the methods and results of this study in relation to research on proof and problem solving and design of introduction to proof courses. First, I summarize the findings to both of my research questions. Then I discuss the implications of this work, to contextualize my results. I also identify the limitations and future research and development suggested by this study, with respect to studying proving and satisfying moments. I end with some concluding remarks with respect to task design and affect in mathematics. As a reminder, the research questions were:

(1) How does undergraduate students' proving develop over the duration of a transition to proof course?

(2) What kinds of satisfying experiences do undergraduate students have during the transition to proof?

Findings Related to the Development of Students' Proving

My first research question focused on the nature of productive changes in students' proving work over the course of the study. Four developments were observed over multiple students in the sample: (1) increased sophistication in how they chose proof techniques to use and their rationales for why, (2) awareness about how a solution attempt was going and harnessing that to change their strategies, (3) becoming comfortable exploring and monitoring when which strategy to pursue is unclear, and (4) checking examples in conjunction with other strategies as a way to trigger new insights when stuck. Some of these developments were specific to the context of proof, such as choice of proof techniques; others, such as awareness of how one's attempt is going, were more general problem solving and thus less proof-specific.

Results indicated that students showed growth in fluency, strategy, and monitoring and judgement in how they reacted when they were stuck. This was evidenced in the approaches they chose to try next, the rationale for their choices, and how they monitored their progress. For example, early on, students tended to use a favorite method (proof by contradiction, contrapositive, etc.) for all problems, indiscriminately. But it was not always the case that these developments led to improvements in the students' proof performance. The imperfect correlation between growth and performance could be seen in Leonhard's individual case. Leonhard's reasoning changed and improved over the course of the interviews, yet his solutions were incorrect for the last two interviews. What does it mean then to have positive growth but stagnated performance? Some may see this as evidence that a student did not in fact improve, but I claim that a de-coupling of performance and growth is appropriate here. This is an age-old case question in educational research and remains for the future.

Although I only discussed a few of the developments in detail, multiple types of developments could be seen in the students. Beyond types or categories of development, there were also multiple ways a development could emerge. This is sensible, that different students would grow in different ways and that that growth would look a little different. The phenomenon of multiple and relatively simultaneous developments can be conceptualized metaphorically as many ropes, each made up of many strands, representing a different way of getting to the development. This is important to acknowledge because oftentimes there is an unspoken assumption that there is one path for learning mathematics and the goal of instruction is to move students along that path. Instead, there are many productive paths of proving development.

Findings Related to Satisfying Moments

The second research question was about identifying the kinds of satisfying moments students experienced in relation to the course. These kinds were identified using grounded theory techniques and grouped broadly into external and internal situations, properties of the mathematics, and interactions with others. The most common satisfying moments among participants concerned completing task(s), overcoming challenges, understanding (as an aggregate of its various forms), external validation, and interactions with people. Some codes directly indicated the nature of a satisfying experience, whereas others, like *On my Own*, appeared to function as a sort of modifier, where its presence seemed to strengthen a satisfying moment. For example, "solving a difficult problem all by myself" is likely more satisfying than "solving a difficult problem but with help."

Certain kinds of satisfying moments stood out in the analysis, and certain aspects of experiences tended to co-occur: (a) *Completing Task(s)* with *Simple* tasks and (b) *Overcoming Challenge(s)* with *Understanding*. It is important to note that *Understanding* was not necessary for feeling satisfaction at *Overcoming a Challenge*, but when a person did feel good about *Understanding*, the situation was typically challenging. This nuance in how challenges and understanding give rise to satisfaction was shown by the four profiles that cover this sample of students: those who enjoy (a) *Completing Task(s)*, (b) *Overcoming Challenge(s)*, (c) *Understanding*, and (d) a combination of the two, *Overcoming Challenge(s)* and *Understanding*. Last, interactions with people were frequently connected other aspects, based on other co-occurring codes: *Friendly Interactions* with *Understanding*. These interactions point to overarching characteristics behind the kinds of satisfying moments.

In thinking about what lies at the heart of satisfying moments as a phenomenon, two ideas emerged. First, mastery seemed to be an overarching characteristic. Partial progress, confirmations or reassurances of present mastery, and expectations of future mastery may explain *why* the situations discussed above were satisfying. However, there were a few instances where mastery did not fully explain the satisfaction; understanding and interactions with others did. Understanding can serve as confirmation of present mastery, but there was something about sense-making that intrinsically seemed to feel satisfying to many of the participants. Understanding involves things falling into place, a sense of "fit," (Sinclair, 2006) which does not fall squarely under the umbrella of mastery. Mastery and understanding overlap then, but there is an aesthetic component to understanding that mastery on its own does not seem to capture. The same applies to interactions with people, especially *Friendly Interactions*; working together with people and helping others to understand has elements that are satisfying which fall outside the purview of pure mastery.

Second, expectations likely played a large role in what was reported as satisfying moments. The results share much in common with the idea of self-efficacy (Bandura, 1977), the expectation of success. Students' expectations seemed to mediate whether an experience was perceived as satisfying. When a student was successful in a situation that was expected to be unsuccessful, this positive discrepancy between expected and actual outcome may have been linked to satisfaction. There is an element of surprise in the expected outcome, which corroborates past work on the importance of surprise in aesthetic responses to mathematics (Satyam, 2016).

This difference between expected and actual outcomes may explain then why students remembered these events, elevating events up and out from the milieu of everyday proving that constituted their normal experience. An experience that followed expectations would be considered "normal" and therefore may not stand out in memory, including situations where failure is expected and then indeed felt. In other words, memorable events are more likely later be reported as satisfying. The observation does beg the question: Are there satisfying moments that are not memorable events? What happens to those? I argue that if an event is not in a person's awareness, then it lacks the power to be a satisfying moment. However, I speculate that we have experiences that we do not place importance on in memory but are still felt at a subconscious level, affecting us later. This is likely beyond the test of empirical data with our current methods, so it remains a philosophical musing.

Based on these results, I proposed a theory of the phenomenon of satisfying moments. This theory came out of my observations that some of my codes were of different "types." Many were situations (e.g. *Completing Task(s), Interactions with People*), but a code like *On my own* acted more like a moderating variable, appeared only in conjunction with other codes and so likely moderated the strength of the main relationship. In addition, codes like *Understanding* and *Partial Progress* were more abstract than other situations. Figure 8.1 illustrates how the different codes may relate to each other, to explain how certain situations give rise to the feeling of satisfaction (satisfying moment).

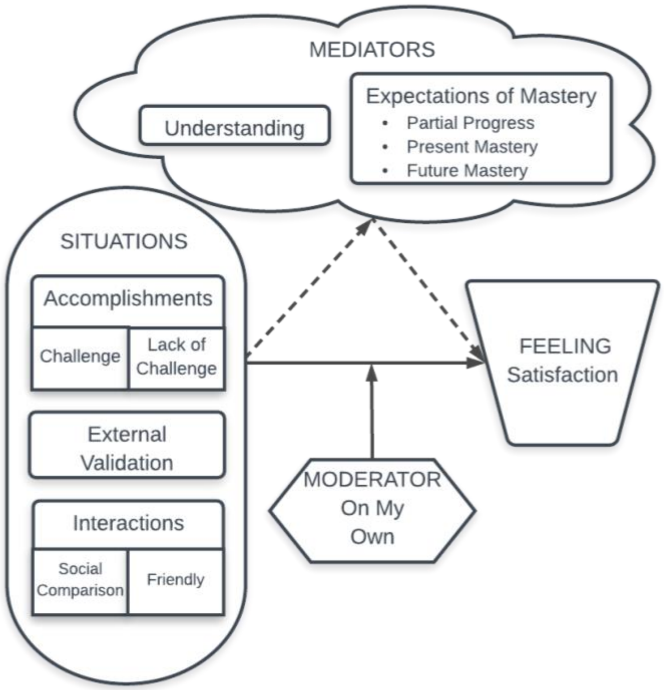


Figure 8.1. Possible model for how satisfying moments occur as a phenomenon. Situations (independent variables) give rise to the feeling of satisfaction (dependent variable). *On my own* may act as a moderating variable, in that it strengthens the elicitation of satisfaction. Understanding and expectations may act as mediating variables, explaining why those situations elicit satisfaction. Note the variable-paradigm is used for illustrative purposes here.

Accomplishments with and without challenge, understanding, external validation, and social interactions with people covered the range of the majority of satisfying moments in this dataset. One can think of them as situations that elicit the emotional response of satisfaction. Working on challenging problems by yourself and/or without needing help (*On my own*) may strengthen the feeling of satisfaction, thereby acting as a moderating variable, which moderates the relationship between the independent and dependent variables. But above all, understanding and expectations of mastery may be what mediate (explain) how certain situations give rise to satisfaction as an emotional response. This means, without understanding or an expectation of mastery, the situations on the left in Figure 8.1 do not give rise to satisfaction.

The depiction of this model was influenced by the independent-dependent variable paradigm, with mediating and moderating variables. This is my speculation; my research methods do not support making any causal arguments. In fact, the situations that give rise to satisfaction as a feeling are not manipulatable; they only provide opportunities for situations to happen. Grounded theory is useful for revealing the categories, but not necessarily for unpacking how the categories relate to each other. However, I offer this up as a speculative theory, based on the varying conceptual types of my codes.

Findings Related to Connections Between Proving and Emotion

Throughout the analysis of proving, connections between affective and cognitive aspects of students' activity presented themselves (though I did not pose a research question to address them), particularly in how their emotions interacted with their awareness of their solution attempts. This corroborates the idea that what students value mathematically (e.g., efficiency, straightforwardness, cleanliness, etc.) may draw and guide them to what (Sinclair 2004). These values manifest themselves though emotion. For example, Granger and Timothy showed strong negative emotions towards solutions they thought were wrong and demonstrated how their emotions influenced their future attempts. The important result here is in how they harnessed strong negative emotions to search for alternate solutions.

A cursory examination at the emotion graphs students drew and emotion words they picked while proving also provided preliminary findings about the relationship between cognition and affect. Analyzing this in full is beyond the scope of this study, entailing different research questions. More robust frameworks for looking at this need to be developed.

Implications

Now I discuss some implications of this work, relating my findings to those of existing studies when possible. I separate these issues into theoretical, methodological, and pedagogical foci.

Theoretical Issues

This work contributes to existing literature on proving, specifically in its focus on students' developments over the course of a single class. This study had a longer duration (a semester) than most studies examining students' proving that are non-interventions and with repeated interactions with multiple students, not just one or two. This work was longitudinal in the short-term sense, examining students proving work across one semester.

Formal-rhetorical aspects of proving may actually be problem-centered. Although this study focused on the problem solving aspects of proving, the developments discussed earlier revealed the amount of decision making that goes into even writing the first line of a proof. Students took the content of the statement to be proven into account when deciding how to begin a proof. This suggests a revisiting of the distinction between formal-rhetorical and problem-centered aspects of proving (Selden & Selden, 2007). In the formal-rhetorical phases of proof construction, the first and last statements are seen as following logically from the statement to be proven and can be stated without a great deal of thought. Acts that we expect to be formal-rhetorical, such as writing the first line of a proof, may actually be more complicated and dependent on content. Selden & Selden

(2007) do not treat these aspects as dichotomous but teaching formal-rhetorical and problem-centered aspects of proving separately may not give us the desired results in students if the interplay of these two aspects is lost.

Noticing as crucial to proving. The act of "noticing" appeared multiple times in students' development, in exploring and monitoring until students noticed something helpful and in using examples before noticing patterns. Teacher noticing of students' mathematical thinking has been a topic of research (Sherin, Jacobs, & Philipp, 2011; Jacobs, Lamb, & Philipp, 2010), but students' own noticing seems to be especially key here. How does a student know where to pay attention and to notice certain things? In solving mathematical problems, there can be many mathematical objects and relationships to attend to; noticing as a phenomenon can be quite complicated (Lobato, Hohensee, & Rhodehamel, 2013). One could speculate that successful provers notice important relationships when they appear, since the solution to problems that are truly problems is not clear from the beginning. Many of us have had the experience of noticing some important piece that makes everything fall into place. In these moments, the solution can seem so obvious after that point, hence the overuse of the word "trivial" in mathematical circles. How do we teach students to notice when something important arises in their work? As a focus for future research, how can we study the development of student noticing?

Role of confidence in proving and its implications. To segue from the discussion of noticing, what role does confidence play in noticing – and proving in general? Amy was my example student for exploring and monitoring; she was comfortable just working without a strategy and noticed when something useful appeared. I noted separately that

Amy had high confidence in her mathematical work, from the beginning of the interviews. Did the fact that she had confidence in her work play a contributing factor in her productive noticing, in trusting herself that she would notice important insights when they came up and that she would act on it when she saw it? Others with high levels of confidence in their work, e.g. Granger and Leonhard, also had moments like Amy's. Students with low levels of confidence in their work, e.g. Dustin and Jordan, struggled.

This connection between confidence may have implications then for the importance of confidence in proving, a process so ridden with failure when held in comparison to most students' prior mathematical work of computation and exercises. Learning mathematics is difficult; students experience repeated failures and that failure is often taken as an indicator of a lack of (a fixed) ability. Perhaps confidence acts as an insulator of self against these failures? A sane person experiencing the regular failures in learning how to prove would likely quit, finding it not pleasing. This would be a natural reaction, all things considered.

Similarly, if we think then about which demographic groups in the United States are culturally associated with confidence, could confidence partly explain why we see mostly white males and a dearth of women in higher levels of mathematics? Sociocultural factors may be at play, perhaps making feeling confident about one's abilities – especially in mathematics – harder for some than others. We should not be so quick, however, to claim that individuals from underrepresented backgrounds should just be more confident. How mathematical classrooms treat students showing confidence as indicators of correctness and/or intelligence (when not always warranted) should be examined.

Algorithms as satisfying. Students often spoke about mathematical tasks that had a clear set of specific steps as satisfying. For example, a number of participants spoke about enjoying induction in their discussion of satisfying moments. Why are steps and algorithms pleasing? This counters the idea that understanding and other forms of explanation are intrinsically pleasing (Sinclair, 2006). Is it that the reduced cognitive load translates to the affective domain as feelings of contentedness? One answer to this may be in what makes a mathematical situation feel like a puzzle and other situations not.

Methodological Issues

This work involved a relatively large number of novel constructs and data analysis techniques, adapted from existing literature. I describe some of the ways in which what I did may help others.

Studying impasses without intruding? People tend to get quiet when they are stuck. Going into this study, I had expected that some participants would have a difficult time talking when they were stuck. I did not anticipate that this would be true across all the participants, but they all fell silent when stuck. This was true even for my one participant, Shelby, who said she found it useful to talk when stuck and would in fact turn to me to say her thoughts out loud when stuck on the proof tasks, after I offered to be her listener. But even she would fall silent at times when she was stuck.

Why is this important? For any research that focuses on what students do when stuck, *it is important to not force them to talk out loud when stuck*. Talking out loud could easily change the nature of the students' experience. Students are deep in thought, expending much cognitive effort, and queries like "What are you thinking" may be an additional load on their cognitive focus and capacity (Ericsson & Simon, 1981). They may lose their train of thought, as happens in everyday life when interrupted. If I had prodded my students to talk when they were stuck, there is a chance that my leading questions could have helped them become unstuck.

In fact, Timothy developed a habit of talking to himself after he became stuck, where he stated the fact that he was stuck, explained why he was stuck, and discussed some of his preliminary ideas that related to the mathematical situation at hand. This often proved successful in leading him out of being stuck. His speech was explanatory, so it was more communicative to others than typical self-talk or Vygotsky's (1978) egocentric speech. Timothy's self-guided talk appeared to help him out of tough situations. In conclusion, prodding students to talk may influence the phenomenon of becoming stuck as an object of study, but inviting students to think out loud may be useful pedagogically.

Emotion graphs. Asking students to draw a graph of their emotions and pick out words that described their emotions while working on a task proved to be insightful for looking into their experience. With careful choices about data collection and care about the types of claims that can be gleaned, these tools are useful for research on students' experience. Satyam et al., (2018) examined the affordances of different variations of graphing as a research tool, and these findings corroborate the results reported in that work: Graphing is useful as a stimulus for helping students make sense of and discuss affective phenomena. The analysis of the graphs themselves, taken as a self-report of some phenomenon, must be done carefully.

How to get students to stay with a series of interviews. I originally chose 12 participants, with the hopes that 8 participants would complete all four interviews, i.e., that I would have no more than a 1/3 drop-out rate. I expected that participants would drop out, especially near the end of the semester when their work load increased, as they prepared for final exams. Instead, I was surprised when 11 of them continued with this work until the end (and the 12th participant had medical issues preventing her from continuing the study). In a world where it is difficult to keep participants coming back, the question is: *Why did they stay?* Yes, they were paid for being in the study, but I do not think that modest reward explains their continuing participation.

I believe my participants stayed in for the simple reason that they got something out of the interviews. A number of them said outright that they saw these interviews with the proof tasks as providing extra practice for their class. I also think the interviews were useful to them as a space to talk about their thoughts regarding the class and math in general. I believe the lesson here for research practice is to think about whether the data collection process is of current value to students, whether that be mathematically and/or emotionally valuable.

Pedagogical Issues

These results also have implications for the design and teaching of introduction to proof courses.

Curriculum design of undergraduate transition to proof courses. I argue that knowing the ways in which students develop and what they find satisfying in challenging mathematical work is useful for designing transition to proof courses at the undergraduate level. One way to go about that design would be to think about the types of developments one wishes to happen and design tasks that aid in student problem solving development. For example, if a goal is for students to come away with knowing when each proof technique makes sense, then one can design a task that asks students to prove a single statement using different proof techniques (e.g. using direct proof, then contradiction, and contrapositive) and then reflect on the advantages and disadvantages of each. A less time-consuming variation of this would be to ask students to consider two or three techniques and write down some of the advantages and disadvantages to using each, prior to implementing any of them. This same tactic could be used for designing assessment items. We can see here the need for a proving process framework, which could drive the curriculum development of courses like these, meant to help students.

Noticing when students are and are not stuck in the classroom. Distinguishing between observable behaviors that indicate a student was stuck vs. thinking silently but not stuck was difficult for me to operationalize. I found in this study that I could only make that distinction by interpreting body language and having a familiarity with the individual. This distinction is especially important for the classroom – how can we tell when a student is unproductively stuck vs. engaged in productive struggle? As math educators, the first we would like to intervene and "help," but productive struggle should be encouraged, not only among college students (Middleton, Jansen, & Goldin, 2017). In fact, as instructors, we often may not want to step in and interrupt productive struggle. Our task may center more in helping students accept productive struggle as a mathematical virtue and learn to make the responses to struggle more productive.

Interview as a vehicle for reflection and rendering knowledge. Last, I realized at some point through these interviews that there were interesting things happening in this space, beyond the foci of my dissertation. These students were being honest about how they felt they did on exams, their in-class experience, and how they felt the course – as well as other courses - were going for them. They were musing out loud about their thoughts on what they were currently learning.

My participants expressed their thoughts about mathematics generally and may not have had other people to talk about it with. Leonhard for example would routinely talk to me during the interview, sometimes going over 2 hours, and tell me his thoughts about mathematics as a whole. Now Leonhard was not the norm, however, the fact remains that this was a large public research university with only a couple mathematics advisors for the entire student body. So advice and mentorship is and was likely hard to come by. This speaks to the importance of truly listening to students and taking their experiences seriously. This is especially important considering that this is a *transition* for students. The course instructors were upfront that the math and their work would be different, but it is not clear if this is generally the case across the country. It is very, very easy for students to make ill-formed inferences about their lack of ability in mathematics and leave the STEM pipeline when there is no intervention by a mentor. How can large institutions institute opportunities for interaction of this kind for their undergraduates? A space like the Math Learning Center, where students can gather to work together and talk about mathematics, would be useful, given the current national focus on STEM education.

The interview also acted as a vehicle for reflection for the students. Some students, like Timothy, realized meta-level aspects about proving during the interview. Needing to discuss the mathematics and share their thought processes likely helped students reflect on and render their knowledge. More opportunities for this kind of reflection in mathematics education seem useful.

Alternative Explanation(s)

Here I consider one alternative explanation of this data and discuss why I believe it can be ruled out. An alternate interpretation of the development results is that students naturally became better at proving over time due to the sheer amount of relevant experience alone. In other words, students' proving grew due to practice and face time with the material and not due to changes in internal cognitive, affective, or reasoning processes. Becoming better at using tools is not necessarily reflective of deeper mathematical understanding, as Guin & Trouche (1999) noted about students using calculators as tools. I argue that the "it's mainly experience" can be ruled out because taking experience as the primary factor does not account for the variation and individual differences seen across students in this sample by the end of the semester. Some students grew in the problem solving domain while others still struggled by the end. Especially since the course design required students to work on proving tasks in class, it can be argued that all students who attended class had some base amount of experience with proving, at least more so than if students' only real experience with proving was left to outside of class time and thus less regulated. The developments seen in some students and not others and also the variations in how these developments occur, when students' experience with the course material is relatively uniform, suggests that repeated practice with tasks is not sufficient to explain growth in proving competence in this context.

Limitations and Factors Influencing the Findings

This analysis was qualitative in nature. My goal was to generate theory and the small sample size was an indicator of that. Generalizations such as how most student learn how to prove are not possible with this data. Future quantitative work assessing the theory generated here will be required to answer questions of that nature.

Students' development was shaped by the nature of this specific transition to proof course. For some of the developments, there was evidence of the instructors explicitly encouraging students to engage in certain helpful practices, e.g., Ms. Frye recommending using examples to gain an intuition for a statement but not to prove it. There is a question then regarding the specificity of these results: How much do they reflect the specific features of this course? To what extent would we see these same developments in any other transition to proof course? Transition to proof has been organized in a myriad of ways across the United States (David & Zazkis, 2017), so a prototypical transition course does not in fact exist. Students using strategies like checking examples likely would have transpired regardless, however, even without the instructor's recommendations.

The developments documented in this study was also shaped by the specific proving tasks that students worked on in the interviews. This raises the question of how much the developments observed in students were shaped by the nature of those tasks. When a student did something different on a task, was it due to particulars of that task or was it indicative of some internal development? This question holds for much of scientific research (Popper, 1963) and so remains unanswerable here. However, the tasks used in this study were drawn from a reasonable population of tasks similar to those seen in class and on homework, so one can argue that development measured via tasks from the course itself would not have looked substantially different. In addition, my first research question focused on identifying *what* changes that occur and not necessarily *why*. I leave future researchers to grapple with that question.

Development may well also depend to some extent on the interview context. Would the same developments have been seen if students had worked on the tasks by themselves without my presence in the room? My presence could have added pressure and thereby impeded problem solving performance, but it also meant they could ask me factual questions easily. In addition, they knew that they would be explaining their work to me afterwards, so they have tended to write more informal written arguments which could be explained verbally when stuck. As a counteracting force though, my stature and also openness in demeanor may have contributed to making these interviews a place where students felt comfortable sharing their thoughts when problem solving. If one wishes to minimize interview presence, less intrusive data collection methods are an alternative. Technology such as Livescribe pens which record students' audio and their written work may be useful.

Timing of tasks was very important and so there are alternative choices that could be made. If I were to repeat the same study, I would ask students to draw emotion graphs immediately after they completed a task, before the debrief, in order to shorten the already short window of time between the proving and affective record of it. Another choice was in trying to capture students' emotions in the moment vs. after problems were solved. In asking students to draw a graph after the task, I documented students' emotions after the fact. However, I contend that it may not matter much what students actually felt in the moment (that is, while they were working on their proofs). Instead, what matters more was their perception of it and remembrance of it afterwards, because those remembered emotions were more likely to stick with them and affect their subsequent work. If emotions in the moment are of interest, one could measure emotion using more physiological methods, such as a heartrate monitor. This would reveal the intensity of a person's emotions but not the character of the emotion, in the way that direct observation or a person reporting their own can.

I acknowledge that these above factors – the course, the tasks, the interview setting, and timing of tasks – influenced the data. However, all these factors are not so straightforward in their effects as to determine the data one way or another.

Lastly, talk of changing one's emotions introduces an ethical dilemma. If we study emotions in mathematics education because emotions are the type of affect most responsive to change and can alter attitudes and beliefs, this implies we wish to alter students' emotions. The notion of trying to change a person's emotions feels to this author as intrusive and manipulative. Emotional responses are highly personal. Do we wish to be in the business of trying to mold students' emotions? To some extent, as instructors we already do this; we take into account students' reactions when we design a lesson. The take-away from this study is not to expect that if we do X, all students will feel Y. Rather, I argue for creating *opportunities* for satisfying moments, to at least set the conditions for them to perhaps occur, regardless of whether they do. The contrapositive holds here – if we do not provide conditions for satisfying moments to occur, satisfying moments may rarely happen and perhaps only for a few students.

Suggestions for Future Research

The empirical results reported here and the speculative propositions and frames that arose from those results suggest issues that could be explored in subsequent studies.

Empirical Work

Future work could examine how satisfying moments change over time for a student. This study produced some, but not sufficient data to make claims about change over time. It would be fascinating to see instances of students learning to enjoy challenges over time and if so, inquire about the factors that orient such change. Is enjoying challenge more of a *trait*-like aspect to an individual that is resistant to change? Another promising direction for future research is to examine satisfying moments of groups of students (for example, those working together in a small group), not just individuals. Liljedahl (2004) called for investigation of group aha moments. This is relevant for the classroom, in thinking about how to design instruction around eliciting intense positive emotions for multiple people at once. Considering how frequently satisfying moments involved fellow students, there may be potential for students to experience this together.

Lastly, this work originally sought to seek out the conditions that elicit satisfying moments, i.e. what actually triggers the feeling of satisfaction. Identifying kinds of satisfying moments, in order to get a lay of the land (so to speak) is the necessary start but getting at what truly causes satisfying moments is the next step.

Theoretical Work

Proving process frameworks. There is a strong need for a proving-problem solving framework that would support the characterization and assessment of students' proving process over time. This call is not a new one:

> A minor expansion of Carlson and Bloom's framework could potentially provide the mathematics education community a proving-process framework, complete with

additional problem-solving attributes that a prover experiences. (Savic, 2012, p. 121)

Such a framework would be useful for diagnostic purposes in the classroom as well as research: It would be an accomplishment to generate a proving-process framework, similar to Carlson and Bloom's problem-solving framework, that would accommodate beginning provers (if not more advanced accomplished students of mathematics). Such a framework would be helpful in assessing a student's proving and their phases or problem-solving attributes that need improvement. It might identify the phases (Orienting, Planning, Executing, and Checking) or problem-solving attributes (Resources, Heuristics, Affect, and Monitoring) that need work, and focus instruction on that phase/attribute (Savic, 2012, p. 122). Reliable assessments for measuring students' proof comprehension have been recently developed (Mejía-Ramos, Lew, de la Torre, Weber, 2017). A framework that covers all that the proving process entails, much like what has been developed over the years for problem solving, would be useful.

Challenge of studying phenomena by looking at individual components. Fundamentally, analysis is the process of breaking complex phenomena down into more basic and separable components, studying each separately, and then putting them all together again. The underlying assumption is that the process of decomposition and recomposition supports insights into what is going on that would not be possible if the researcher simply looked carefully at the whole phenomenon of interest. However, I sensed in my analyses of both proving and satisfying moments that something that was being lost in this process, some gestalt sense of what was going on. I believe this happened for both the analysis of proving and satisfying moments because both are fuzzy constructs – they are hard to define, especially in a way that is measurable. Problem solving and satisfying moments each involve numerous highly interrelated processes. Considering the two phenomena, I undertook this study thinking I could effectively study cognition as one component and affect as another. But as the work unfolded, I found that the relations between the two could not be ignored. This study uncovers some of the connections, but future work will likely uncover more.

Going forward, theory that attends to interactions may be most illuminating. Even in the 1990s, Schoenfeld (1992) argued that "what we know little about is how these [problem solving] components interact" (p. 363), referring to "resources," "heuristics," and "beliefs" among others. Theory that targets interactions among highly interactive rather than separable components will be helpful in shedding light on how "fuzzy" phenomena work, from descriptive to more explanatory understandings.

Conclusions

As was just stated, it is difficult to ignore the connections between affect and cognition. My concluding remarks relate these two phenomena and look across the results of each analysis to review what we have learned.

The importance of task design is clear. Careful task design, whether it be in-class instruction or out-of-class homework and other assessments, can support students' development in productive ways and potentially have students feel good about doing mathematics. Similarly, poor task design can fail to support those goals, if not worse. This extends beyond the context of proving and even undergraduate education, into K-12 schooling as well. Constructing and selecting tasks that support developments we wish to see are important. In the case of satisfaction, sequencing seems crucial because of the temporal order of emotions. Curriculum is therefore an important force. Careful attention to the storyline of the mathematics (e.g. Dietiker, 2016) and what reactions certain curriculum choices elicit (e.g. Dietiker, Richman, Brackoniecki, Miller, 2016) may be the key to providing more frequent opportunities for students to feel good about mathematics and themselves as learners of that content.

Are there any affective qualities in common across successful provers? While I did not address this directly in the analysis, some observations about my participants began to coalesce. Students whose satisfying moments involved challenges were successful either to start or became successful. The students in my sample whose satisfying moments concerned accomplishments without struggle and easy tasks tended to struggle with proving and the course, as time went on. Finding joy in struggle may be important for students, as they take on more difficult tasks in life, let alone mathematics.

In addition, successful problem solvers in my sample seemed to take failure as opportunities for learning. When they got a problem wrong, they would ask to see how it worked and expressed that now they would know what to do in the future. This speaks to a larger issue of how we treat mastery, challenge, and non-success in mathematics. While this issue extends far beyond mathematics, how math is often portrayed in this world makes failure more salient and catastrophic than in other domains of human endeavor. A retooling of how we teach mathematics – and how failures can be progress – may help us resolve this.

APPENDICES

APPENDIX A: Proof Tasks

Interview 1 – Task 1

We say that two integers, x and y, have the same parity if both x and y are odd or both x and y are even. Prove the following statement:

Suppose x and y are integers. If $x^2 - y^2$ is odd, then x and y do not have the same parity.

Interview 1 – Task 2

Prove the following statement:

If a and b are strictly positive real numbers, then $(a+b)^3$ never equals $a^3 + b^3$.

Reminder: Binomial expansion of $(x+y)^3 = x^3 + 3x^2y + 3xy^2 + y^3$

Interview 2 – Task 1

Two numbers are consecutive means one number comes after the other. Prove the following statement:

If x and y are consecutive integers, then xy is even.

Interview 2 – Task 2

We say x divides y if kx = y for some integer k. Prove the following statement:

If a, b, and c are non-zero integers such that a divides b and a divides c, then a divides ($mb + nc$), for any integers m and n.

Interview 3 – Task 1

Three positive integers a, b, and c are called a <u>Pythagorean triple</u> if they satisfy $a^2 + b^2 = c^2$.

Prove the following statement:

Suppose x, y, z are positive integers. If x, y, and z are a Pythagorean triple, then one number is even or all three numbers are even.

Interview 3 – Task 2

Prove the following statement **without using induction**:

If n is an odd natural number, then $n^2 - 1$ is divisible by 8.

Interview 4 – Task 1

A <u>perfect square</u> is any number that can be written as n^2, for some integer n. Prove the following statement:

If a and b are odd perfect squares, then their sum a + b is never equal to a perfect square.

Interview 4 – Task 2

Prove the following statement:

If x, y are positive real numbers and x ≠ y, then $\frac{x}{y} + \frac{y}{x} > 2$.

APPENDIX B: Interview Protocols

Interview #1 Protocol

Logistics:
- Interview should take place in a quiet room
- Max time: 90 min
  - 20 minutes for Proof Task #1
  - 20 minutes for Proof Task #2
  - 30-45 minutes for Satisfying Moments questions, word selection, and graph
- Ask students ahead of time to bring any scratchwork and a copy of their homework

Materials for students:
- The 2 proof tasks on separate sheets of paper
- Resources for proof task for student: sheet with definitions, laptop, scrap paper
- Note: Provide the hw, example sheets, and solutions in paper form.
- Cards with emotion words on them
- Emotion graph worksheet

Materials for interviewer:
- Paper for writing down word selection
- Paper for notes

Key for this document:
Black plain text is the script, instructions and questions to be spoken to participant
Black italicized text is instructions for interviewer, not to be spoken
Red italicized text denotes purpose of question, linking everything back to research questions.

Warm-up & Basics (1st interview only)
Thanks for agreeing to help us. I'm going to ask you some questions about your experience with math courses in your past and MTH 299.
1. What mathematics courses have you taken here before 299?
2. Are you taking other math courses this semester, along with 299?
3. How is MTH 299 going for you, so far?

Proof Tasks [RQ1]

I'm trying to understand students' thought process, in how they approach proofs; it's not about the final answer. So I'm going to give you two statements I'd like you to prove. It would be helpful to me if you share your thinking with me: what you're trying to do, why you're trying to do that, etc. I want you to vocalize everything you're thinking about the problem. Pretend you're at home and you're just talking to yourself out loud.

You may know me as a TA for this course, but in this interview, because I am interested in how you are thinking, so I will not be able to help you out during the task. This means I can't answer any questions about the math or what to do next if you get stuck.

Any questions I ask or notes I take means that I'm interested in what you're doing, it's not a sign that I'm judging your work or that your thinking is incorrect. I may ask questions every so often like, "What do you mean by…?" or "Why did you decide…" I may ask you "What are you thinking right now?" if you're been quiet for some time.

Proof Task #1
Here's the first task. You can do your work directly on this sheet of paper [where the task is written]. You may look at definitions in this supplementary document from class, your notes from class, or online using my computer. I'll give you 15 minutes to work on it and it doesn't matter how far you get, it'll be fine. I'll then have you stop and we can talk about what you did. You can start.

Interviewer sits near enough to see their work but a little farther away than when typically asking interview questions, in order to give student space to minimize pressure from being watched, as much as possible. Give the student max 20 minutes.

After student is done, ask the following questions:

4. Can you mark for me any places you would count as scratchwork, as in, you wouldn't include if you were typing it up in LaTeX?
5. Were there any places where you got stuck?
   Probe about their rationale at points of interest, where students paused or went in a new direction:
   How did you…?
   Why did you…?

Ask student to do the ***Emotion Word*** and ***Emotion Graph*** *(scroll down to that section) for this proof task.*

Proof Task #2. *Repeat above steps.*

After both tasks are complete:

6. How would you say you currently approach proofs right now? *Students' perception of their "typical" approach to proofs, right now* [RQ 1a]

Satisfying Moments [RQ 2]

7. Have you had any satisfying moments related to your work in MTH 299 since the last time we met? *Identifying satisfying moments without any influence in certain directions by interviewer*
8. Were there moments that felt satisfying while you were working on problems on the last homework set?
   Clarify and Probe as needed *Identifying satisfying moments recently in time*

9. How about the rest of this homework set? *Identifying satisfying moments a little farther out*
10. What's your favorite kind of problem? What kind of problem in this class feels the most rewarding? *A different approach to trying to access satisfying moments*
11. How about in class (since the last interview)? *Identifying satisfying moments longer ago*
12. Can you think of a time when you had a flash of understanding or insight? *If they had any A-HA moments, a hypothesized type of satisfying moment.*
13. Do you have moments of negative emotion, such as frustration? What moments stand out? *Moments of intense negative emotion (for sake of completeness)*
14. Do these moments (positive or negative) affect your motivation to continue to do math? If so, how? *Link between moments of intense emotion and motivation.*

The Word Selection & Emotion Graph Tasks should be about the same experience. It is up to the interviewer which experience to choose.

Word Selection Task [RQ2]

I'd like you to select words that reflect what it was like to work on <that satisfying moment>

Description of Task: 11 Words total

5 negative	5 positive	1 neutral
annoyed	curious	
disappointed	surprised	
sad	joyful	indifferent
frustrated	satisfying	
ashamed	proud	

Notes: I've written words that are "opposite" emotions in the same colors. Words in black have no corresponding word pair. Each word is written on a separate small notecard, in black. I spread the notecards out in the exact arrangement above in front of the student.

15. What words did you choose? Please circle them on this sheet. *Identifying emotions behind the problem*
16. What made you pick the words you did? *Identifying conditions*
17. Are there other words that reflect how you felt about this problem that weren't included here? *Covering any other emotions, so word choice bank doesn't restrict answers*

Emotion Graph Task [RQ 2]

I'd like you to now draw a graph that shows your emotions while working on this problem. The x-axis is time, from when you started to when you stopped working on this problem. The y-axis is emotions, where positive and negative emotions. Think of the highest mark as indicating strong positive emotions like satisfaction or excitement. The middle mark would be neutral, as in your normal resting state. The lower mark would be strong negative feelings like frustration or panic. Please also mark what triggered any ups and downs, as in turning points, in your graph, like which strategies you tried and how that corresponds with your emotional reactions.

18. Talk me through your graph here. *Have participant talk through the experience, marking turning points (conditions for changed emotion).*
    Clarify and Probe as needed

Interviews #2-4 Protocol

Logistics:

- Interview should take place in a quiet room
- Max time: 90 min
    - 20 minutes for Proof Task #1
    - 20 minutes for Proof Task #2
    - 20 minutes for Satisfying Moments questions, word selection, and graph

Materials for students:

- The 2 proof tasks on separate sheets of paper
- Resources for proof task for student: laptop, scrap paper
- Note: Provide the hw, example sheets, and solutions in paper form.
- Cards with emotion words on them
- Emotion graph worksheet

Materials for interviewer:

- Paper for writing down word selection
- Paper for notes

Key for this document:
Black plain text is the script, instructions and questions to be spoken to participant
Black italicized text is instructions for interviewer, not to be spoken
Red italicized text denotes purpose of question, linking everything back to research questions.

Warm-up

1. How is MTH 299 going for you since our last interview?

Proof Tasks [RQ1]
I'm trying to understand students' thought process, in how they approach proofs; it's not about the final answer. So I'm going to give you two statements I'd like you to prove. It would be helpful to me if you share your thinking with me: what you're trying to do, why you're trying to do that, etc. I want you to vocalize out loud how you're thinking about the problem and what you're trying to do. Pretend you're at home and you're just talking to yourself out loud.

You may know me as a TA for this course, but in this interview, because I am interested in how you are thinking, so I will not be able to help you out during the task. This means I can't answer any questions about the math or what to do next if you get stuck.

Any questions I ask or notes I take means that I'm interested in what you're doing, it's not a sign that I'm judging your work or that your thinking is incorrect. I may ask questions every so often like, "What do you mean by…?" I may ask you "What are you thinking right now?" if you've been quiet for some time.

After you are done, I will ask to talk me through what you did and why. I may ask you some questions like, "What do you mean by…?" or "Why did you decide…"

Proof Task #1
Here's the first task. You can do your work directly on this sheet of paper [where the task is written]. I'll give you 20 minutes to work on it and no matter where you get to, it's fine. You can start.

Interviewer sits near enough to see their work but a little farther away than when typically asking interview questions, in order to give student space to minimize pressure from being watched, as much as possible. If it looks like interviewer presence is causing pressure, interviewer will leave and tell participant to talk into the microphone. Give the student max 20 minutes.

If present, interviewer should try to take notes about what participant is doing – their process, including place they got stuck.

After student is done, ask the following questions:

2. Can you mark for me any places you would count as scratchwork, as in, you wouldn't include if you were typing it up in LaTeX?
3. Were there any places where you got stuck?

   Probe about their rationale at points of interest, where students paused or went in a new direction:

   How did you…?

   Why did you…?

Word Selection Task
Emotion Graph Task (leave the room for emotion graph)

Proof Task #2. *Repeat above steps, including Word and Graph Tasks.*

After both tasks are complete:

4. How would you say you currently approach proofs right now? *Students' perception of their "typical" approach to proofs, right now* [RQ 1a]

5. Do you think your ability to write proofs has changed since the last time we met? In what ways?

Probe as needed. Students' perception of their development since last point in time [RQ 1b]

Satisfying Moments [RQ 2]

6. Have you had any satisfying moments related to your work in MTH 299 since the last time we met?

Ask them to pull out graphs they did at home and talk me through it. Identifying satisfying moments without any influence in certain directions by interviewer

7. Talk me through this experience/your graph.
   Ask as needed
   a. Can you find me the exact problem?
   b. What was happening before this?
   c. Were you working alone or with others (if not clear)
   d. What do you think triggered "this"?

If they forgot to return graphs, give them a blank sheet to draw it.
If they forgot to do it period, ask the following:

8. Have you had any moments related to MTH 299 that felt satisfying? By satisfying, I mean a super positive feeling, like rewarding or a feeling of joy, etc.
   Clarify and Probe as needed Identifying satisfying moments recently in time

If they report no satisfying moments:

9. So you said you've had no satisfying moments (homework, class, etc.) – is this true?
10. Is math ever satisfying for you?
    Prompt for examples and try to suss out situations/properties.
11. If so, what do you think it takes for math to be satisfying for you?

Back to MTH 299

12. Have you had any moments of negative emotion, such as frustration, since our last interview? Any moments that stand out? *Moments of intense negative emotion (for sake of completeness)*

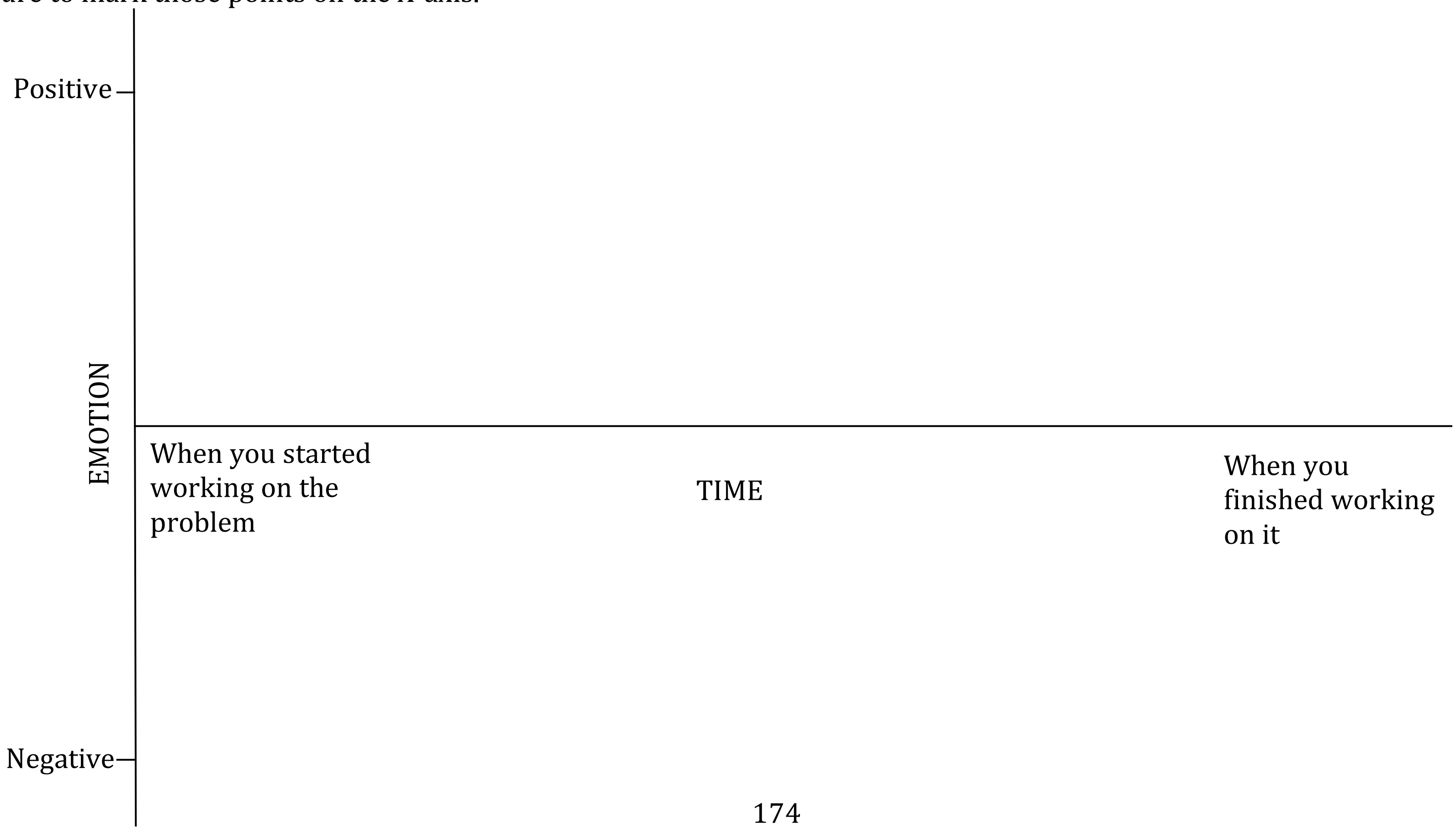
13. Do/how do these moments (positive or negative) affect your motivation to continue to do math? If so, how? *Link between moments of intense emotion and motivation.*

APPENDIX C: Emotion Graph

Please draw a graph of your **emotions** over the course of this homework problem.

The X-axis represents **time**, from when you started working on the problem to when you finished. Please mark different strategies you used on the x-axis.

The Y-axis represents your **positive and negative feelings** while working on this homework problem. Think of the highest mark as indicating emotions like satisfaction or excitement; the middle mark would be neutral, your "resting state"; and the lower mark would be feelings like frustration or panic. If there were points during the problem when your feelings changed, be sure to mark those points on the X-axis.



REFERENCES

REFERENCES

Alcock, L., & Weber, K. (2005). Proof validation in real analysis: Inferring and checking warrants. *The Journal of Mathematical Behavior*, *24*(2), 125-134.

Alcock, L., & Weber, K. (2010). Referential and syntactic approaches to proving: Case studies from a transition-to-proof course. In F. Hitt, D. Holton, & P. Thompson (Eds.), *Research in collegiate mathematics education VII* (pp. 93-114). American Mathematical Society.

Bae, Y., Smith, J. P., Levin, M., Satyam, V. R., & Voogt, K. (in press). Stepping through the proof door: Undergraduates' experience one year after an introduction to proof course. *Proceedings of the 21st Annual Conference on Research in Undergraduate Mathematics Education*. San Diego, CA.

Bakeman, R., & Gottman, J. M. (1997). *Observing interaction: An introduction to sequential analysis*. Cambridge, England: Cambridge University Press.

Baker, D., & Campbell, C. (2004). Fostering the development of mathematical thinking: Observations from a proofs course. *Problems, Resources, and Issues in Mathematics Undergraduate Studies*, *14*(4), 345-353.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191-215.

Barnes, M. (2000). "Magical" moments in mathematics: Insights into the process of coming to know. *For the Learning of Mathematics*, *20*(1), 33–43.

Bell, A. W. (1976). A study of pupils' proof-explanations in mathematical situations. *Educational Studies in Mathematics*, *7*(1-2), 23-40.

Bieda, K. N. (2010). Enacting proof-related tasks in middle school mathematics: challenges and opportunities. *Journal for Research in Mathematics Education, 41*, 351-382.

Bieda, K. N., Ji, X., Drwencke, J., & Picard, A. (2014). Reasoning-and-proving opportunities in elementary mathematics textbooks. *International Journal of Educational Research, 64,* 71-80.

Blanton, M. L., Stylianou, D. A., & David, M. M. (2003). The nature of scaffolding in undergraduate students' transition to mathematical proof. *International Group for the Psychology of Mathematics Education*, *2*, 113-120.

Blåsjö, V. (2012). A definition of mathematical beauty and its history. *Journal of Humanistic Mathematics, 2*(2), 93-108.

Bleiler, S. K., Thompson, D. R., & Krajčevski, M. (2014). Providing written feedback on students' mathematical arguments: Proof validations of prospective secondary mathematics teachers. *Journal of Mathematics Teacher Education*, *17*(2), 105-127.

Buxton, L. (1981). *Do you panic about maths?: Coping with maths anxiety*. London: Heinemann.

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

Cellucci, C. (2015). Mathematical beauty, understanding, and discovery. *Foundations of Science*. http://doi.org/10.1007/s10699-014-9378-7

Cobb, P., & Wheatley, G. (1988). Children's Initial Understandings of Ten. *Focus on Learning Problems in Mathematics*, *10*(3), 1-28.

David, E. J., and Zazkis, D. (2017). Characterizing introduction to proof courses: A survey of R1 and R2 institutions across the U.S. In A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, & S. Brown (Eds.), *Proceedings of the 20th Annual Conference on Research in Undergraduate Mathematics Education* (pp. 528–535). San Diego, CA.

Dawkins, P. C., & Karunakaran, S. S. (2016). Why research on proof-oriented mathematical behavior should attend to the role of particular mathematical content. *The Journal of Mathematical Behavior*, *44*, 65-75.

de Villiers, M. D. (1990). The role and function of proof in mathematics. *Pythagoras*, *24*, 17-24.

DeBellis, V. A., & Goldin, G. A. (2006). Affect and meta-affect in mathematical problem solving: A representational perspective. *Educational Studies in Mathematics*, *63*(2), 131-147.

Dietiker, L. (2016). Generating student interest with mathematical stories. *Mathematics Teacher*, *110*(4), 304-308.

Dietiker, L., Richman, A., Brakoniecki, A., & Miller, E. (2016). Woo! Aesthetic variations of the" same" lesson. In *Proceedings of the 38th annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*. Tucson, AZ: The University of Arizona.

Dreyfus, T., & Eisenberg, T. (1986). On the aesthetics of mathematical thought. *For the Learning of Mathematics, 6*(1), 2–10.

Dubinsky, E., Elterman, F., & Gong, C. (1988). The student's construction of quantification. *For the Learning of Mathematics*, *8*(2), 44-51.

Dweck, C. S. (2006). *Mindset: The new psychology of success*. New York, NY: Random House LLC.

Epp, S. S. (2003). The role of logic in teaching proof. *The American Mathematical Monthly*, *110*(10), 886-899.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215–251.

Ericsson, K. A., & Simon, H. A. (1981). Protocol analysis. *Communications of the ACM, 49*(2), 117–122. http://doi.org/10.1145/1113034.1113039

Evans, J. (2002). *Adults' mathematical thinking and emotions: A study of numerate practice*. London, England: Routledge.

Fawcett, H. P. (1938). T*he nature of proof; a description and evaluation of certain procedures used in a senior high school to develop an understanding of the nature of proof.* Oxford, England: Teachers College, Columbia University.

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

Garofalo, J., & Lester Jr, F. K. (1985). Metacognition, cognitive monitoring, and mathematical performance. *Journal for Research in Mathematics Education, 16*(3), 163-176.

Glaser, B., & Strauss, A. (1967). *The discovery of grounded theory: Strategies for qualitative research*. Chicago, IL: Aldine Publishing Company.

Gómez-Chacón, I. M. (2000). Affective influences in the knowledge of mathematics. *Educational Studies in Mathematics*, *43*(2), 149-168.

Guin, D., & Trouche, L. (1999). The complex process of converting tools into mathematical instruments: The case of calculators. *International Journal of Computers for Mathematical Learning, 3, 195–227.*

Hadamard, J. (1945). *Psychology of invention in the mathematical field.* Princeton, NJ: Princeton University Press.

Hanna, G., & Jahnke, H. N. (1996). Proof and proving. In *International handbook of mathematics education* (pp. 877-908). Dordrecht, Netherlands: Springer.

Hannula, M. S. (2002). Attitude towards mathematics: Emotions, expectations and values. *Educational Studies in Mathematics*, *49*(1), 25-46.

Hardy, G. H. (1940). A mathematician's apology. Cambridge: Cambridge University Press. http://doi.org/10.1017/CBO9781139644112

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

Harel, G., & Sowder, L. (2007). Towards a comprehensive perspective on proof. In F. Lester (Ed.), *Second handbook of research on mathematical teaching and learning* (pp. 805–842). Washington, DC: NCTM.

Herbst, P. G. (2002). Establishing a custom of proving in American school geometry: Evolution of the two-column proof in the early twentieth century. *Educational Studies in Mathematics*, *49*(3), 283-312.

Higgins, J. L. (1970). Attitude changes in a mathematics laboratory utilizing a mathematics-through-science approach. *Journal for Research in Mathematics Education*, *7*, 43-56.

Houston, K. (2009). *How to think like a mathematician: A companion to undergraduate mathematics.* New York: Cambridge University Press.

Iannone, P., & Inglis, M. (2010). Self efficacy and mathematical proof: Are undergraduate students good at assessing their own proof production ability? In *Proceedings of the 13th Conference on Research in Undergraduate Mathematics Education*. Raleigh, NC.

Inglis, M., & Aberdein, A. (2014). Beauty is not simplicity: an analysis of mathematicians' proof appraisals. *Philosophia Mathematica*, *23*(1), 87-109.

Inglis, M., & Alcock, L. (2012). Expert and novice approaches to reading mathematical proofs. *Journal for Research in Mathematics Education*, *43*(4), 358-390.

Jacobs, V. R., Lamb, L. L., & Philipp, R. A. (2010). Professional noticing of children's mathematical thinking. *Journal for Research in Mathematics Education*, *41*(2), 169-202.

Jones, K. (2000). The student experience of mathematical proof at university level. *International Journal of Mathematical Education in Science and Technology*, *31*(1), 53-60.

Karunakaran, S. (2014). Comparing bundles and associated intentions of expert and novice provers during the process of proving. (Doctoral dissertation).

Knuth, E. J. (2002). Secondary school mathematics teachers' conceptions of proof. *Journal for Research in Mathematics Education*, *33*(5), 379-405.

Knuth, E. J., Choppin, J., & Bieda, K. (2009). Middle school students' productions of mathematical justification. In M. Blanton, D. Stylianou, & E. Knuth (Eds.), *Teaching and learning proof across the grades: A K-16 Perspective* (pp. 153-170). New York, NY: Routledge.

Ko, Y. Y., & Knuth, E. J. (2013). Validating proofs and counterexamples across content domains: Practices of importance for mathematics majors. *The Journal of Mathematical Behavior*, *32*(1), 20-35.

Leighton, J. P. (2009). Two types of think aloud interviews for educational measurement: Protocol and verbal analysis. Paper presented for symposium *How to Build a Cognitive Model for Educational Assessments* at the 2009 annual meeting of the National Council on Measurement in Education (NCME), San Diego, CA.

Liljedahl, P. G. (2004). The Aha! experience: Mathematical contexts, pedagogical implications. (Doctoral dissertation).

Lo, J. J., Grant, T. J., & Flowers, J. (2008). Challenges in deepening prospective teachers' understanding of multiplication through justification. *Journal of Mathematics Teacher Education, 11*(1), 5–22. http://doi.org/10.1007/s10857-007-9056-6

Lobato, J., Hohensee, C., & Rhodehamel, B. (2013). Students' mathematical noticing. *Journal for Research in Mathematics Education*, *44*(5), 809-850.

Malmivuori, M.-L. (2001). The dynamics of affect, cognition and social environment in the regulation of personal learning processes: The case of mathematics. (Doctoral dissertation).

Mariotti, M. A. (2006). Proof and proving in mathematics education. *Handbook of research on the psychology of mathematics education: Past, present and future*, 173-204.

Mason, J., Burton, L., & Stacey, K. (1982). *Mathematical thinking.* London, England: Addison-Wesley.

McLeod, D. (1988). Affective issues in mathematical problem solving. *Journal for Research in Mathematics Education, 19*(2), 134–141.

McLeod, D. (1992). Research on affect in mathematics education: A reconceptualization. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (Vol. 1, pp. 575–596). New York, NY: Macmillan.

McLeod, D. B. (1994). Research on affect and mathematics learning in the JRME: 1970 to the present. *Journal for Research in Mathematics Education, 25*(6), 637–647.

McLeod, D., Craviotto, C., & Ortega, M. (1990). Students' affective responses to non-routine mathematical problems: An empirical study. In G. Booker, P. Cobb, & T. de Mendicuti (Eds.), *Proceedings of the Annual Conference of the International Group for the Psychology of Mathematics Education with the North American Chapter 12th PME-NA Conference* (pp. 159–166). Mexico.

Mejía-Ramos, J. P., & Inglis, M. (2009). Argumentative and proving activities in mathematics education research. In *Proceedings of the ICMI study 19 conference: Proof and proving in mathematics education* (Vol. 2, pp. 88-93).

Mejía-Ramos, J. P., Lew, K., de la Torre, J., & Weber, K. (2017). Developing and validating proof comprehension tests in undergraduate mathematics. *Research in Mathematics Education, 19*(2), 130-146.

Middleton, J. A., Jansen, A., & Goldin, G. A. (2017). The complexities of mathematical engagement: Motivation, affect, and social interactions. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 667-699). Reston, VA: National Council of Teachers of Mathematics.

Middleton, J., & Spanias, P. (1999). Motivation for achievement in mathematics: Findings, generalizations, and criticisms of the research. *Journal for Research in Mathematics Education, 30*(1), 65–88. Retrieved from http://www.jstor.org/stable/749630

Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics, 27*(3), 249–266. http://doi.org/10.1007/BF01273731

NGA, & CCSSO. (2010). *Common Core State Standards for Mathematics*. Washington D.C.: National Governors Association Center for Best Practices (NGA) and Council of Chief State School Officers (CCSSO). http://www.corestandards.org/Math/

Op't Eynde, P., De Corte, E., & Verschaffel, L. (2007). Students' emotions: A key component of self-regulated learning? In *Emotion in education* (pp. 185-204).

Op't Eynde, P., De Corte, E., & Verschaffel, L. (2006). "Accepting emotional complexity": A socio-constructivist perspective on the role of emotions in the mathematics classroom. *Educational Studies in Mathematics*, *63*(2), 193-207.

Pedemonte, B. (2007). How can the relationship between argumentation and proof be analysed? *Educational Studies in Mathematics*, *66*(1), 23-41.

Pedemonte, B., & Reid, D. (2011). The role of abduction in proving processes. *Educational Studies in Mathematics*, *76*(3), 281-303.

Piaget, J. (1964). Development and learning. In R. Ripple & V. Rockcastle (Eds.), *Piaget Rediscovered* (pp. 7–20). New York, NY: W. H. Freeman. Retrieved from http://www.psy.cmu.edu/~siegler/35piaget64.pdf

Piaget, J. (1971). The theory of stages in cognitive development. In D. R. Green, M. P. Ford, & G. B. Flamer, *Measurement and Piaget.* New York, NY: McGraw-Hill.

Poincaré, H. (1952). *Science and method*. New York, NY: Dover.

Polya, G. (1945). *How to solve it: A new aspect of mathematical method.* Princeton, NJ: Princeton University Press.

Popper, K. R. (1963). *Conjectures and refutations: The growth of scientific knowledge*. New York, NY: Harper & Row.

Recio, A. M., & Godino, J. D. (2001). Institutional and personal meanings of mathematical proof. *Educational Studies in Mathematics*, *48*(1), 83-99.

Rota, G.-C. (1997). The phenomenology of mathematical beauty. *Synthese, 111*(2), 171–182.

Satyam, V. R. (2016). The importance of surprise in mathematical beauty. *Journal of Humanistic Mathematics*, *6*(1), 196-210.

Satyam, V. R., Levin. M., Grant. T. J., Smith, J. P., Voogt, K., & Bae, Y. (in press). *Graphing as a tool for exploring students' affective experience as mathematics learners*. In *Proceedings of the 21th Annual Conference on Research in Undergraduate Mathematics Education.* San Diego, CA.

Savic, M. (2012). Proof and proving: Logic, impasses, and the relationship to problem solving. (Doctoral dissertation).

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

Schoenfeld, A. H. (1992). Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. In D. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334-370). New York, NY: Macmillan.

Schoenfeld, A. H. (1985a). Making sense of "out loud" problem-solving protocols. *The Journal of Mathematical Behavior, 4*, 171–191.

Selden, A., & Selden, J. (2013). Proof and problem solving at university level. *The Mathematics Enthusiast*, *10*(1&2), 303-334.

Selden, A., & Selden, J. (1987). Errors and misconceptions in college level theorem proving. In *Proceedings of the Second International Seminar on Misconceptions and Educational Strategies in Science and Mathematics* (pp. 457–470).

Selden, A., & Selden, J. (2003). Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education*, *34*(1), 4-36.

Selden, A., & Selden, J. (2007). Teaching proving by coordinating aspects of proofs with students' abilities. (Report No. 2007-2.) Retrieved from http://files.eric.ed.gov/fulltext/ED518762.pdf

Senk, S. L. (1989). Van Hiele levels and achievement in writing geometry proofs. *Journal for Research in Mathematics Education*, *20*(3), 309-321.

Sherin, M., Jacobs, V., & Philipp, R. (Eds.). (2011). *Mathematics teacher noticing: Seeing through teachers' eyes*. New York, NY: Routledge.

Silver, E. (1985). Research on teaching mathematical problem solving: Some underrepresented themes and needed directions. In *Teaching and learning mathematical problem solving: Multiple research perspectives* (pp. 247–266).

Sinclair, N. (2004). The roles of the aesthetic in mathematical inquiry. *Mathematical Thinking and Learning, 6*(3), 261–284. http://doi.org/10.1207/s15327833mtl0603

Sinclair, N. (2006). *Mathematics and beauty.* New York, NY: Teachers College Press.

Sinclair, N. (2009). Aesthetics as a liberating force in mathematics education? *ZDM*, *41*(1-2), 45-60.

Smith, J. P., Levin, M., Bae, Y., Satyam, V. R., & Voogt, K. (2017). Exploring undergraduates' experience of the transition to proof. In A. Weinberg, C. Rasmussen, J. Rabin, M. Wawro, & S. Brown (Eds.), *Proceedings of the 20th Annual Conference on Research in Undergraduate Mathematics Education* (pp. 298–310). San Diego, CA.

Staples, M., Bartlo, J., & Thanheiser, E. (2012). Justification as a teaching and learning practice: Its multifacted (potential) role in middle school classrooms. *Journal of Mathematical Behavior*, *31*, 447-462.

Stylianides (2007). The notion of proof in the context of elementary school mathematics. *Educational Studies in Mathematics, 65,* 1-20.

Stylianides, A. J., Stylianides, G. J., & Philippou, G. N. (2004). Undergraduate students' understanding of the contraposition equivalence rule in symbolic and verbal contexts. *Educational Studies in Mathematics*, *55*(1-3), 133-162.

Stylianides, G. J., Stylianides, A. J., & Philippou, G. N. (2007). Preservice teachers' knowledge of proof by mathematical induction. *Journal of Mathematics Teacher Education*, *10*(3), 145-166.

Stylianides, G. J., Stylianides, A. J., & Weber, K. (2017). Research on the teaching and learning of proof: Taking stock and moving forward. In J. Cai (Ed.), *Compendium for research in mathematics education* (pp. 237-266). Reston, VA: National Council of Teachers of Mathematics.

Thomsen, D. E. (1973). The beauty of mathematics. *Science News*, *103*(9), 137–138.

van Hiele, P. M. (1959). The child's thought and geometry. In *Classics in mathematics education research* (pp. 60–68). Reston, VA: National Council of Teachers of Mathematics.

Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.

Weber, K. (2010). Mathematics majors' perceptions of conviction, validity, and proof. *Mathematical Thinking and Learning*, *12*(4), 306-336.

Weber, K., & Alcock, L. (2004). Semantic and syntactic proof productions. *Educational Studies in Mathematics, 56*(2–3), 209–234. http://doi.org/10.1017/CBO9781107415324.004

Wells, D. (1990). Are these the most beautiful? *The Mathematical Intelligencer*, *12*(3), 37-41.

Zeki, S., Romaya, J. P., Benincasa, D. M. T., & Atiyah, M. F. (2014). The experience of mathematical beauty and its neural correlates. *Frontiers in Human Neuroscience*, *8*(68), 1-12. https://doi.org/10.3389/fnhum.2014.00068