7-12-2014

Changes in Student Proving Skills and Attitudes Following a Cooperative Learning Seminar

Martha Byrne

Martha Byrne
Candidate

Mathematics and Statistics
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Michael Nakamaye , Chairperson

Kristin Umland

Cristina Pereyra

Timothy Fukawa-Connelly

CHANGES IN STUDENT PROVING SKILLS AND ATTITUDES FOLLOWING A COOPERATIVE LEARNING SEMINAR

by

MARTHA BYRNE

B.A., Mathematics, Knox College, 2002

M.S., Mathematics, University of New Mexico, 2010

DISSERTATION

Submitted in Partial Fulfillment of the
Requirements for the Degree of

Doctor of Philosophy

Mathematics

The University of New Mexico
Albuquerque, New Mexico

May, 2014

DEDICATION

For my children: your first years were harder because I was working on this, but the rest of them will be better because I finished it.

ACKNOWLEDGMENTS

It is impossible to express enough gratitude to all of the individuals who supported me through this process, but I'll try. After all, there were plenty of times when writing a dissertation seemed impossible, but I tried. I want to start by thanking Mike for keeping me laughing, and keeping me from giving up even if he didn't know that's what he was doing; Kristin for keeping me on task and being the voice of reason in the room; Cristina for her support of me and of math education; and Tim for being the most engaged outside committee member anyone could ask for and an excellent mentor to boot.

Being the lone RUME researcher at UNM wasn't easy, but the PRG made it easier. It's what helped me get from dissertation idea #1 to dissertation idea #3d, which is presented here. Particular thank you to Melissa and Milos for being great colleagues as well as friends.

I also want to thank my family and friends in Albuquerque for their shoulders, their ears, and their willingness to watch my kids when needed.

The entire Earlham community has held me up this year, and I couldn't ask for a better community to be a part of. I wish I could stay.

I also want to thank Ben for his support and his enduring patience. It's been a long road, and I wouldn't be at this end of it without him.

Finally, I'd like to acknowledge myself for my years of hard work that made this dissertation possible.

CHANGES IN STUDENT PROVING SKILLS AND ATTITUDES FOLLOWING A COOPERATIVE LEARNING SEMINAR

by

Martha Byrne

B.A., M.S.,

Ph.D.

ABSTRACT

This dissertation details research studies designed to explore undergraduate math students' beliefs and attitudes about mathematical proof, ability to compose valid proofs, and ability to read and validate purported proofs written by other students. In two studies, a cohort of seminar participants were assessed twice on their attitudes and beliefs about mathematical proof, their ability to compose proofs, and their ability to validate arguments. Between assessments, these participants worked on carefully crafted problem sets in a Cooperative Learning environment. In each study, a cohort of comparison participants took both assessments but did not engage in structured, cooperative work in the interim.

Results from both studies showed little change in participants' attitudes, and varied changes in validation skills. However, in both studies, most seminar participants' composition skills improved from pre-assessment to post-assessment. The composition results are consistent with the researcher's hypothesis that working in a Cooperative

Learning environment on carefully chosen problem sets can help students develop their proof writing abilities. Additionally, because the content area of the assessments (number theory) and seminar problem sets (functions) were distinct, the demonstrated improvement of the seminar participants supports the hypothesis that some proof skills can be transferred across distinct mathematical contexts. The composition and validation results from both studies call into question how proof composition and validation skills are related, as many participants demonstrated improved proof composition skills but did not show improvement in proof validation skills.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1 - Introduction

Proof is essential to mathematics, and proofs play many roles within the work of the mathematical community (de Villiers, 1990; CadwalladerOlsker, 2011). They are written to verify the truth of mathematical statements, to explain the reasoning behind that truth, and also persuade others that the statement is true and the reasoning behind it is correct. Proofs can also be written for discovery and intellectual challenge but are generally not done so at the undergraduate level. Because proofs are such an integral part of the field of mathematics, students graduating with their bachelor's degrees in mathematics in the US should understand the nature of proof and be able to communicate mathematics in writing (Committee on the Undergraduate Program in Mathematics, 2001). However, many students still have poorly developed skills at the time of graduation (Sowder & Harel, 2003).

Much research undertaken in the past two decades shows that students struggle with constructing and validating proofs (Almeida, 2000; Harel & Sowder, 1998; Levine & Shanfelder, 2000, Moore, 1994; Selden & Selden, 2003a, 2003b; Weber, 2001; Weber, 2003), and courses dedicated to the transition to proof are now part of the curriculum at many institutions. Several innovative course structures have been introduced for so-called bridge courses (Almeida, 2003; Bakó, 2002; Grassl & Mingus, 2004), but little dedicated research has been done on the effectiveness of such courses. However, some common themes have emerged about the necessity for and efficacy of active learning strategies, and there is a general trend away from lecture and toward more student-centered models.

In particular, this can be seen within the Modified Moore Method community (McLoughlin, 2010) and among proponents of Inquiry Based Learning (IBL).

Cooperative learning (CL) is one such active learning model. “CL may be defined as a structured, systematic instructional strategy in which small groups work together to produce a common product” (Cooper, 1990). There are six specific features that, when combined, distinguish CL from other active and IBL strategies: positive interdependence, individual accountability, appropriate grouping, student interaction, attention to social skills, and teacher as facilitator. While the efficacy of CL has been researched (Johnson & Johnson, 1991), the majority of this research has been undertaken with precollegiate populations.

Studies done on CL and active learning in the context of physics instruction (Deslauriers, et al, 2011; Heller & Hollabaugh, 1992; Heller, et al., 1992) give hope that CL could be effective in helping students acquire and develop their proof skills. The research studies presented in this dissertation were designed to test the hypothesis that working in a CL environment on carefully crafted materials could be beneficial to the acquisition and development of proof skills. Specifically, the studies were designed to examine if, after students worked in a CL environment, one could see measureable differences in 1) student attitudes about mathematics and mathematical proof, 2) student ability to construct proofs, and 3) student ability to validate student-generated arguments. These questions were addressed in the hope of motivating further research into how attitudes and skills are affected by cooperative work.

Pilot Study Methods

The researcher conducted a pilot study in the spring of 2012. Two groups of volunteer participants, a seminar and a comparison group, were assessed twice on mathematical attitudes, proof composition skills and proof validation skills. At least 11 weeks passed between assessments for all participants. During the intervening weeks, seminar participants met for eight 90-minute seminar sessions with the researcher. During these sessions, seminar participants worked in assigned, structured groups on proof-focused problem sets that had been adapted and developed by the researcher. The groups were formed and seminar sessions were conducted according to CL criteria, and the mathematical content of the problem sets focused on the concept of functions.

All participants were enrolled in proof-based courses at a large, public university. The comparison participants did not meet with the researcher between assessments and did not engage in structured group work in their courses.

Pre-assessments consisted of four portions: a background questionnaire, an attitudes/beliefs survey, a set of claims to prove, and a set of arguments to validate. The background questionnaire asked participants about their major, minor, grade level, GPA, gender, and previous and concurrent proof-based course work. Seminar participants were then interviewed about their questionnaire responses. The attitudes/beliefs survey asked participants to use a Likert-type scale to respond to 19 questions about mathematics and mathematical proof. After completing the questionnaire and survey, participants were asked to attempt to prove four claims presented as true statements. All four statements were about elementary number theoretical concepts. Participants were not timed and were allowed to go back to any argument they had set aside. Seminar participants were assessed in an interview setting and the researcher employed a think-aloud protocol during the composition section. Comparison participants produced their arguments without interaction with the researcher. The final portion of the pre-assessment consisted of four student-generated arguments attempting to prove a single claim. The arguments are those presented and discussed by Selden and Selden (2003). Participants were presented with the arguments one at a time and were asked to classify the validity of each argument and provide justification for their classification. Participants were not allowed to return to previous validations and were not allowed to progress to the next argument until a classification and justification had been provided. However, they were allowed to return to the composition portion of the assessment after completing the validation portion.

Post-assessments were almost identical to pre-assessments; the post-assessment did not include the questionnaire, and the researcher interviewed the seminar students about their experience in the research study, but the rest of the assessment was unchanged.

The attitudes/beliefs survey data from pre- and post-assessments were compared, but no patterns were apparent, and no insight was gained. To evaluate the arguments produced by the participants, the researcher adapted a tool presented by Boyle (2012) that assigned broad categories to each argument and enumerated details corresponding to each category. Participants' pre-assessment and post-assessment arguments were compared by item to check for improvement, stasis, or regression. Validation attempts were examined using a protocol adapted by the researcher from Selden and Selden's work (2003).

Classifications and their corresponding justifications were evaluated and pre-assessment and post-assessment validations were compared.

The researcher transcribed the seminar participants' exit interviews and examined and made adjustments to the CL elements of the seminar and the problem sets for the implementation study based on those interviews. Specifically, the researcher adjusted the group roles that were assigned to the group members, changed the processing questions discussed by the groups at the end of each seminar session, and reduced the number of copies of the problem sets provided to each group.

Implementation Study Methods

The implementation study for this research was conducted in the fall of 2012. Some changes were made to the seminar for this study. Group members were assigned daily roles in both studies, but the specific roles were altered for the implementation study. In both studies, groups reflected on their efficacy and cohesion at the end of each seminar session, but the framework for those discussions was also altered. Each group in the implementation study was given a single copy of each problem set, whereas each individual had been provided copies in the pilot study. Finally, some changes were made to the problem sets themselves; one problem set was removed, and specific problems were altered (for more details, see Chapter 3). However, the conceptual content of the problem sets, functions, was not changed.

The questionnaire, composition, and exit interview portions of the assessments were identical to their counterparts from the pilot study. The Likert-type attitudes and beliefs survey was abandoned and replaced with open-ended questions about attitudes and beliefs. Seminar participants were asked these questions in an interview, and comparison participants provided their responses in written form. The arguments presented in the validation portion were not altered from the pilot study, but they were presented to participants in a different order.

The researcher employed the same data analysis tools and methods in the implementation study as had been used in the pilot study. She examined the attitudes and beliefs responses for patterns among participants and for change from pre-assessment to post-assessment for individuals.

Results

Because the data from the Likert-type scale survey were not useful, no results were recorded for the attitude/belief sections for the pilot study participants. Most of the implementation participants were familiar with and felt positively about mathematical proof at the time of their pre-assessments, and the attitudes and beliefs expressed on the pre-assessments were generally the same as those presented on the post-assessments.

While results from the attitudes and beliefs portion of the assessment were largely static, the results from the composition portion were more varied. In both studies, most seminar participants' composition skills improved from pre-assessment to post-assessment despite the fact that the content area of the problem sets was different from that of the assessments. Most seminar participants exhibited greater flexibility with multiple proof methods by switching methods more frequently when they were stuck on an argument on the post-assessment than on the pre-assessment. Few of the comparison participants improved in proof composition from pre-assessment to post-assessment, and no data were gathered on the frequency of method switching for the comparison participants.

In contrast to the clear trends in composition results, validation skill results were inconsistent. The researcher saw no trends in either the pilot or the implementation study. Instead, comparison and seminar participants from both studies exhibited improvement, regression, and stasis, and several students showed both improvement and regression.

Finally, exit interview results from the two studies differed greatly. Dissatisfaction with the group work experience was high among pilot study seminar participants, but the implementation study seminar participants responded much more positively to the CL components of the study.

Discussion

The data presented in this dissertation are consistent with the researcher's hypothesis that Cooperative Learning paired with appropriate and carefully crafted materials may be beneficial to students as they acquire and develop proving skills. However, the small number of participants in this research makes it impossible to determine whether or not the cooperative seminar and problem sets were responsible for the difference between the pre-assessment and post-assessment proof-writing skills of the seminar participants. There are several other factors, such as participant motivation and relationship with the researcher that may have contributed to these results. However, these results do motivate further study on the connection between seminars such as these and the development of proof skills.

These results also give rise to other questions that merit more investigation. While there is undoubtedly a connection between composition and validation skills, the nature of that relationship is unknown and should be studied further. There was little correlation in these data between composition and validation skills and one participant made errors in composition that were similar to errors she identified as critical during her validation exercises. While validation skills are considered by some researchers to be essential to composition competence, much more research needs to be done on the connections between the skill sets. In particular, there may be a difference in validating proofs written by other people and one's own proofs

The relationship between flexibility in using a variety of proving methods and improvement in proof composition is also unclear. While the results of this study regarding successful proof production in conjunction with participants' tendency to switch proof methods mirror Hart's (1994) findings, the relationship is not fully understood. No data were gathered on the proof methods employed by the comparison participants, and more study is warranted.

The fact that the seminar students were able to demonstrate improvement in proof composition despite the fact that the mathematical content of the assessments was different from that of the seminar gives support to the notion that content-independent, transferrable proof skills exist. This researcher has plans to study this topic further in hopes of determining which skills are content-independent and the extent to which they are transferrable.

Such investigations have the potential to alter how the mathematical community approaches the instruction of proof and proof-based courses. As existing research shows, proof courses are often not meeting the needs of students. The results presented here, and the future studies they may motivate, may provide instructors with alternative classroom models that will be more effective for supporting students' acquisition and development of proof skills.

Chapter 2 – Literature Review

This review of the relevant research literature will begin by discussing current classroom practices in undergraduate proof-based mathematics courses. It will continue with a discussion of student understanding of proof including Harel and Sowder's (1998) taxonomy of proof schemes. What then follows are descriptions of how students write and validate proofs, and a discussion of the presumed, but untested, assumption that transferrable, content-independent proof skills exist. Finally, the literature defining and supporting Cooperative Learning practices is presented.

Teaching Proof

Typically, well-meaning, mathematically well educated, and well regarded professors enter classrooms and lecture to their students presenting the standard model of providing a definition, stating a theorem and immediately presenting a proof (Almeida, 2000). However, when constructing proofs, working mathematicians operate with a very different model consisting of "intuition, trial, error, speculation, conjecture, proof" (Maclane, 1994). So in general, students do not get to witness the creative nature of mathematics.

We begin introducing students to proofs and the need for proof and then start expecting to see well-constructed valid proofs based on the concepts defined and presented in class. However, because students enter college and these courses without much proof experience, the sudden call for formal rigorous proofs is generally too much. Instructors first need to take into account the pre-formal proof notions of their students; and they must be able to meet the students where they are in order to help guide them to where they ought to be. (Almeida, 1995)

Student Understanding of Proof

To understand why undergraduate math majors are often not graduating with the proof skills many consider essential to further study in mathematics (Harel & Sowder, 2003), one must first examine how undergraduate students understand mathematical proof. Students in the United States are generally introduced to proof in either a high school geometry class or in an upper-level undergraduate course, and at the point when undergraduate students are asked to produce their own proofs, they have typically spent more than a decade in computational mathematics classes in which they have been asked to provide little justification for the computations they perform. Mathematics is thus often viewed as a computational pursuit, and the abrupt transition to the creative endeavor of proof writing can be very difficult for students. Take, for example, this quote from an undergraduate major: “Of course I don’t like [proofs], but I guess it’s because you have to gather so much information in order to be able to prove it, and if you don’t know part of the information, or if it doesn’t pop out of your head right away, then you don’t know how to get it all together ” (Sowder and Harel, 2003).

Harel and Sowder's Taxonomy of Proof Schemes

Harel and Sowder, (1998) define a taxonomy of proof schemes that are held by students; proof scheme refers to what constitutes "ascertaining and persuading" for an individual (p. 244). The proof schemes are divided into three broad categories: external, empirical, and analytical. The taxonomy is not entirely hierarchical, although some categories are considered more advanced than others. However, they are not mutually exclusive, and individuals may hold multiple proof schemes simultaneously.

External conviction schemes can be classified as ritual, authoritarian, and symbolic. A student with an external-ritual scheme is concerned with whether or not arguments match what they think proofs should look like but is generally incapable of analyzing arguments with other criteria. An external-authoritarian scheme is characterized by reliance on outside sources; a proof is true because it was in the textbook, because the teacher presented it, or because someone like Euler has his name attached to it, and the holder of this type of scheme will not see the need to prove it for his/herself. An argument written under an external-symbolic proof scheme is based on algebraic manipulation of symbols done without heed to the meaning of the symbols and often results in a deeply flawed and incorrect argument.

Harel and Sowder also define empirical proofs schemes, which are divided into inductive and perceptual subcategories. Empirical-inductive proof schemes utilize inductive, instead of deductive, reasoning and example-based justification, while empirical-perceptual schemes are marginally more sophisticated. With the latter, students make observations "by means of rudimentary mental images - images that consist of perceptions and a coordination of perceptions, but lack the ability to transform or to anticipate the results of a transformation" (p. 255).

The last of the proof schemes that Harel and Sowder consider to be pre-formal are analytical-transformative schemes. While more advanced than any of the external or empirical schemes in that people holding these schemes can apply operations on the mathematical objects in play and anticipate the repercussions of such operations, these schemes are still considered pre-formal because they are not axiomatic and are constrained by one or more presumed restrictions on the part of the prover.

As presented in CadwalladerOlsker (2011), Giancarlo Rota describes proof this way: "Everbody knows what a mathematical proof is. A proof of a mathematical theorem is a sequence of steps which leads to the desired conclusion. The rules to be followed by such a sequence of steps were made explicit when logic was formalized early in this century, and they have not changed since" (p.34). This is one view of proof that CadwalladerOlsker defines a formal notion of proof, and states that in pure formalism, mathematical justification arises from the acceptance of certain undefined terms and accepted axioms and follows logical rules to demonstrate the desired conclusion. "For the formalists, the *meaning* of the mathematical proposition was irrelevant, proofs were exclusively based on syntactic constructs and manipulations" (p. 34). However, he asserts that no working mathematician truly writes purely formal proofs as all but the most trivial cases would be too cumbersome, long, and unintelligible to fulfill any of the desired roles for proofs. Therefore, mathematicians take a more practical view of formal proofs where it is admitted and accepted that "mathematics is a *human* endeavor" and that there is a subjective side to proof (p. 36) that takes into account the community in which and for which the proof is written.

Reliance on axioms as well as understanding that proof is subjective are part of the final proof scheme of Harel and Sowder (1998) - the analytical-axiomatic scheme. Holders of this scheme understand the roles of and reliance on those undefined terms and axioms, and this is the type of scheme instructors generally want to see students developing before they graduate.

How Students Write Proofs

What distinguishes the behaviors and practices of expert provers from those of novices is not fully understood. In his 1994 study of the practices of students in elementary group theory, Hart defined four levels of proving expertise and examined the proving behaviors of students along the continuum. He found that students progressing from level 0 to level 1 and from level 1 to level 2 switched proving methods with greater frequency but that students progressing from level 2 to level switched methods less frequently (pp. 59-60). While it is not the case that mathematicians don't also switch proving methods, this suggests that increased flexibility in trying different proof methods general correlates with improved proving performance among less experienced provers.

One may expect that students with advanced proof schemes and sufficient conceptual knowledge of a subject would regularly be able to produce valid proofs. However, Weber (2001) showed that this is not always the case, and that students with a solid understanding of proof construction and of relevant mathematical concepts still often fail to produce adequate proofs. One of the obstacles to producing valid proofs may be students' difficulty translating informally worded logical statements into the formal statements that provide insight into the proof (Selden and Selden, 1995).

Syntactic and semantic reasoning.

Weber and Alcock (2004) defined two types of proof production. Syntactic proof procedures involve using only relevant definitions, known results, and formal reasoning. During semantic proof production, on the other hand, authors operate outside of the context of the formal system. They think about and refer to the mathematical context of the claim using examples and instantiations of relevant mathematical constructs. Semantic reasoning is generally used to motivate the reasoning behind the truth of the statement and guide the deductive reasoning that is applied when a statement is proved.

It's important to note that syntactic proofs are not incorrect or even undesirable. In some situations the statement of the theorem, when considered in conjunction with the relevant definitions, provides a proof framework (Selden and Selden, 1995) and is easily undertaken in a syntactic fashion. However, Weber and Alcock (2004) argue that the proofs that can be produced in this manner are limited and less intuitively convincing to the prover than arguments produced via semantic reasoning.

Undergraduate mathematics instructors must assist and guide students as they navigate Harel and Sowder's taxonomy, ideally making their way to analytical-axiomatic proof schemes and operating with semantic procedures. Unfortunately, many teachers provide models of finished, formal proofs during class, but very little in the way of guidance. Take, for example, the attitude of the student that was presented earlier. She had seen formal proofs in class that seem to pop out of her professors' heads without struggle or indeed much indication of the process involved. While it is possible that those professors were able to produce such proofs without effort because of the level of the material, in presenting the finished proofs to their students, they are not providing models their students can apply. As a consequence, the students are left feeling lost but facing high expectations without much instruction or assistance. "I also feel like her expectations were very high, but clear guidance was not given in order to achieve those expectations," (anonymous comment on a teaching evaluation form from a transition to proof/discrete structures course).

How Students Validate Proofs

Many studies have been conducted on how students validate arguments (e.g. Powers, et. al., 2010; Segal, 1999; Selden and Selden, 2003; Selden and Selden 1995) and some researchers argue that proof validation is an essential skill for mathematics students and professionals because when writing a proof, the author needs to be able to evaluate his/her own work for correctness. However, most students struggle significantly with the exercise of reading through a purported proof and reflecting on the text to determine the correctness of the argument. Selden and Selden (2003) claim that composition and validation skills are closely related; "Constructing or producing proofs in inextricably linked to the ability to validate them reliably" (p. 9).

Textbooks that treat proof often have validation exercises in which students are asked to find the error(s) in a provided proof. However, most of these proofs are carefully constructed by textbook authors to contain only one error (Selden & Selden, 2003). While there is some value in this, students need more practice validating student-generated arguments in order to improve their skills at validating their own work because student-generated arguments present much more complexity and a larger variety of difficulties.

The struggles students experience while validating proofs are varied. They tend to ignore issues with logical structure and attend more to the details of the proof (Selden and Selden, 2003). Even when presented with arguments in which the details are correct and the only flaws are in the structure, the structural flaws are largely missed (Piatek-Jimenez, 2004).

As mentioned above, many students rely on empirical evidence to justify claims and are personally unconvinced by valid proofs. Segal (1999) and Weber (2010) studied students' interpretations of mathematical arguments when the students were to judge the arguments as convincing or not and as valid proofs or not. Empirical arguments were often judged as personally convincing but not proofs, which is understandable, but they also regularly rated arguments as not convincing but valid proofs. They did this in cases where the arguments were valid proofs as well as when they were not.

Content-Independent Proof Skills

Research on proof and proving often discusses proof and students' aptitude with proof as if the students' understanding of the mathematical content of the proofs were independent of the proving skills. For example, Blanton and Stylianou (2003) assert "students who engage in whole-class discussions that include metacognitive acts as well as transactive discussions about metacognitive acts make gains in their ability to construct mathematical proofs," (p. 119) and they do not limit that assertion to the realm of discrete mathematics, which was the content area of their study. Similarly, Sowder and Harel (2003) discuss proof skills without referencing mathematical context:

> Some students come to university with excellent (proof understanding, production and appreciation) PUPAs and continue to thrive in a proof environment. Others enter university with poor PUPAs and unfortunately graduate without a significant change in their proof skills and attitudes. Still others come with poor proof skills but do show some growth during their undergraduate mathematics programs. (p. 251)

Selden and Selden also take part in the content-independent discussion of proof skills. In a 2003 paper, they present a list of errors and misconceptions commonly presented by novice provers, and although they do not claim that it is a comprehensive list, it was compiled using data from abstract algebra course work and presented as applicable to all content areas. In addition, Weber (2003) suggests a route toward understanding the concept of proof based on research done in a first course on real analysis.

Many institutions offer a transition to proofs class to help prepare students for their advanced mathematics courses (Moore, 1994; Levine & Shanfelder, 2000; McLoughlin, 2010; Selden & Selden 2007). The mathematical content of these courses varies but often focuses on set theory and other discrete topics. The existence of these courses seems to represent the hope that students can develop proof skills that will be transferrable to future courses. However, little research, if any, has been done to test whether such content-independent, transferrable skills exist and what they may be.

Collaborative Teaching Practices

Chickering and Gamson (1987) laid out seven principles for good practice in undergraduate education that were designed based on the contemporary body of educational research with the purpose of guiding the practices of universities and colleges to better serve and educate students. According to Chickering and Gamson, good undergraduate educational practice encourages student-faculty contact, encourages cooperation among students, encourages active learning, gives prompt feedback, emphasizes time on task, communicates high expectations, and respects diverse talents and ways of learning.

Inquiry Based Learning (IBL).

Inquiry Based Learning is an umbrella term referring to environments in which students are primarily engaged in exploration activities and not lectured to. Many undergraduate mathematics instructors are using IBL formats and techniques in their classes (Grassl and Mingus, 2004; Leron and Dubinsky, 1995; Levine and Shanfelder, 2000; McLoughlin, 2010), and those authors present evidence that such classes can produce students who are much more proficient at constructing and validating mathematical arguments and proofs. In large physics lectures, students taught with an IBL approach for one week were drastically more successful than their lectured counterparts (Deslauriers, Schelew, and Wieman, 2011).

Deslauriers, Schelew, and Wieman's study was conducted within the theoretical framework of deliberate practice, which encompasses constructivism and formative assessment. Ericsson, Krampe, and Römer (1993) characterize deliberate practice by concern for a subject's motivation to engage in the task and improve his performance, heed paid by the subject to his/her own preexisting knowledge so that the task can be undertaken with a minimum of introduction, immediate and informative feedback by an expert or more knowledgeable person, as well as repetition of the same or similar tasks. They stress the need for more than repetition because "with mere repetition, improvement of performance was often arrested and further improvement required effortful reorganization of skill" (p. 365). Deliberate practice is a dedicated practice undertaken with guidance and instruction from a qualified teacher or tutor solely for the purpose of improving one's performance.

While Ericsson et. al. (1993) were examining the type and amount of deliberate practice involved in achieving expert and eminent performance in a variety of disciplines, Deslauriers, Schelew, and Wieman's (2011) research was conducted with the goal of getting students started down a path that could potentially lead to expert performance. How to best help students increase their proof construction and validation performances is an important question even when the students are not expected to pursue mathematics beyond the undergraduate curriculum.

Cooperative Learning (CL).

Cooperative Learning (CL) is a specific model of IBL that puts emphasis on collaborative efforts and has had success in undergraduate education in general (Cooper and Robinson, 1994) and in math and science in particular (Springer, 1998). Within the CL model, students work in small groups (ideally with 4 members) on structured tasks designed by the instructor, with the instructor moving between groups observing and intervening when deemed appropriate.

CL is distinguished by six primary principles: positive interdependence, individual accountability, appropriate grouping, student interaction, attention to social skills, and teacher as facilitator (Cooper, 1990). The combination of positive interdependence (students taking responsibility for the learning of all of the members of their group) and personal accountability (it is recommended that students' grades be based almost entirely on individual assessment) is particularly indicative of CL as students need to take responsibility for themselves and for each other. Methods such as rotating role assignment can foster both of these and largely eliminate the familiar group dynamics in which one or two students do the lion's share of the work and pull along their less engaged group mates.

Groups must also be assigned appropriately in order for them to function well, with considerations given to size and diversity of members; heterogeneous groups are generally preferred (Millis, 1992, Harskamp, et. al., 2007). Activities, projects, and tasks are structured and designed to maximize discussion and student-student interaction, while student-teacher interactions change in nature due to the role of teacher as facilitator rather than a "sage on the stage."

Finally, attention is paid to the social skills necessary to work productively in small groups and there is dedicated discussion and training on those skills. For example, groups engage in formal group processing where group members grade themselves and their group mates on social criteria periodically throughout the term of the course.

Cooperative Learning and the seven principles.

The CL structure incorporates the seven principles laid out by Chickering and Gamson, and provides both a theoretical framework and course of action for fulfilling the promises of those principles (Millis, 1992). Additionally, engagement in deliberate practice, as laid out by Ericsson et al, addresses most of the principles as well. Given the anecdotal success of IBL methods in advanced math classes and the success of deliberate practice-motivated cooperative learning methods in physics, there is sufficient groundwork laid for a study in how cooperative deliberate practice affects the proof writing and validation skills of undergraduate mathematics majors.

It is important to note that the success of IBL, and specifically CL, techniques is audience-dependent. The research of Springer (1998) and Cooper and Robinson (1994) suggests that the top performing and the lowest performing students do not see as much increased success in alternative courses as in traditionally taught courses, but the middle two quartiles of students show great improvement in alternative environments. Their research also indicates that working in CL environments may have a particularly positive effect on minority populations.

Chapter 3 - Methods

Overview

This research was conducted through two studies in 2012. Because most of the materials used in the studies were newly developed, the researcher wanted to conduct a pilot study during the spring semester and follow it with an implementation study during the fall semester of that calendar year. Thus, she would have the ability to make changes to the format and/or materials based on the results of the pilot study. The changes she made are discussed below, but the two studies were similar enough that the results can be discussed concurrently. Because the pool of potential participants for both studies was limited by the requirements of the study, and quantitative analysis methods would have been inappropriate, the researcher employed qualitative methods for both studies.

Both studies looked at changes in participants' attitudes and beliefs about mathematics and mathematical proof, proof composition skills, and argument validation skills. In both studies, students taking undergraduate proof-based courses at a large, public university were given pre-assessments at the beginning and post-assessments at the end of the semester. These participants were all pursuing degrees or minors in mathematics or mathematics education. They were primarily juniors and seniors, but one freshman and one first-year graduate student also participated.

In each study, the participants were divided into a seminar cohort and a comparison cohort. Each seminar cohort met with the researcher throughout the semester to work on problem sets in cooperative groups while the comparison cohorts did not. The researcher used the comparison cohort to examine potential changes in demonstrated proof skills of students who were not engaged in classroom-based cooperative work.

In each study, eight seminar participants and six comparison participants took the pre-assessment, but during the pilot study one seminar participant and three comparison participants withdrew after the pre-assessment, and during the implementation study one seminar participant and one comparison participant withdrew.

In each study, the assessments for the seminar and comparison cohorts were identical, but the different participants were not assessed in the same setting. The researcher was most interested in seeing if seminar participants' abilities to read and validate proofs would change after taking part in a Cooperative Learning-based seminar, so she assessed the seminar participants in individual interviews and had them employ a think-aloud protocol while constructing arguments. Conducting individual interviews with the comparison participants as well was time-prohibitive, and thus a common time was found during which they all could take their individual assessments. The comparison participants did not talk about their assessments with the researcher, except to ask for clarification.

In both studies, the seminar participants met collectively with the researcher for eight 90-minute seminar sessions between pre- and post- assessments. The participants were assigned into groups based on gender, previous proof experience, and proof composition performance on the pre-assessment. During the seminar sessions, the participants worked in their assigned groups on proof-based problem sets designed by the researcher. Each seminar participant, except for one in the pilot study, was enrolled in at least one proof-based course at the university. The other participant had independently studied proofs and was enrolled in a course in integral calculus.

The comparison participants in both studies were juniors and seniors, and each was enrolled in at least one proof-based course at the university. The participants reported that none of the courses they were enrolled in utilized any structured group work as a part of the course. Between assessments they had no interactions with the researcher.

Research Questions

Is there evidence that after working on proof-based problems in a Cooperative Learning there are measurable differences in

- an individual's attitudes about mathematical proof?
- an individual's proof composition skills?
- an individual's proof validation skills?

Participants

Pilot study.

Seminar.

Bill – Bill was a math major pursuing a minor in computer science. Classified as a junior, he had an atypical background. Bill had spent three semesters at a small college with a great books curriculum, one semester in the Semester at Sea program, one semester enrolled part time at a large public university, and was starting his third semester of full time studies at that same university. Bill had left his first college because he was concerned that he was not going to have any applicable job skills after graduating. He had a B+ grade average and was taking courses in complex variables and linear algebra at the time of the study. At his previous school he had taken a class on Euclid's *Elements* and Ptolemy's *Almagest*. In that course, students read and reproduced proofs, but they were never asked to compose proofs of their own. "People were randomly called and they'd have to present one of the proofs and work through it and remember the steps and explain it to everyone. The only proofs I've had to do, some basic stuff last semester in Calc III, we had to construct some proofs."

Ingrid – Ingrid was classified as a senior but did not anticipate graduating for several more semesters. She had a C+ grade average but told the researcher that some health issues had sometimes prevented her from completing her work so her grades were not reflective of her understanding. She had previously completed courses in symbolic logic and philosophy that required proofs, and she had also been enrolled in, but withdrawn from, an undergraduate advanced calculus course. At the time of the study, she was enrolled in a linear algebra course. Ingrid had initially declared a major in physics, but she was drawn to the proofs and the mathematics behind the physics and switched majors to mathematics.

Ivan - Ivan was a senior education major with a secondary focus in mathematics. He had a D+ grade average and did not expect to graduate for several semesters. He had previously completed courses in geometry and discrete structures and was enrolled in a course on number theory at the time of the study. Ivan was a non-traditional student who had worked for many years before pursuing an undergraduate degree. He had decided to pursue a degree in education because he "looked back at the jobs [he'd] had and the greatest success [he'd] had was in teaching the job," and because he believed that "truth is something that can be taught as opposed to something that just naturally comes on high [*sic*]." He specifically chose to focus on mathematics because he was drawn to its precision.

Nathan – Nathan was a senior majoring in pure mathematics with a minor in economics. He had a C+ grade average and graduated at the end of the semester in which the study was held. Nathan had the greatest experience with proof-based courses of the seminar participants in this study. He had previously completed courses in discrete structures and advanced calculus, as well as two courses in abstract algebra. At the time of the study, he was enrolled in both a course on number theory and the second semester of advanced calculus. Nathan had decided to major in mathematics while he was in high school because he had a lot of confidence in his mathematical abilities and because his teachers encouraged him to pursue the field.

Omar – Omar was the least experienced participant in the study. He was classified as a junior and had declared a double major in math and philosophy with a minor in computer science. He had previously taken a course on symbolic logic, but had no other proving experience. He had entered college intending to pursue a music major but the program "didn't work out for a multitude of reasons," and he dropped out of school. He came back to the university to pursue a degree in mathematics largely because of the encouragement from his wife. He enjoyed reading books about mathematics in his spare time, so it seemed like a good field to pursue. However, his previous mathematical education had not been strong, so he started taking basic algebra classes at a local community college. At the time of the study, he was enrolled in a calculus course. He said that he'd enrolled in the study because the ideas seemed interesting and he wasn't afraid of jumping in even if he wasn't really ready for it.

Ursula – A junior, Ursula was an applied math and linguistics major. She had previously taken a course in discrete structures and had a B+ grade average. Ursula was loaned a calculus textbook when she had been taking a course on college algebra and was inspired by the material in that text, so she decided to major in mathematics "until it stops being fun." She expressed an intention to pursue graduate studies in math or computer science after completing her undergraduate work.

Zach – Zach had transferred to the researcher's institution in the fall semester of 2011. Previously he had been at another large, public university, but he had been out of school for eight years before starting at the new school. He was closest to a math major when he came back to school, so even though his preference would have been to pursue a degree in physics, he opted for the math major and minor in physics. He had a C+ grade average and had decided to return to school because he found it very difficult to get by without having a bachelor's degree. At the time of the study, he was enrolled in an advanced calculus course.

Other – One other seminar student enrolled in the study and took the pre-assessment but withdrew from the study before any of the seminar sessions.

Comparison.

0296 – 0296 was a third year undergraduate who was classified as a senior. He was pursuing a math major with a minor in Spanish and had a B+ grade average. He had completed proof-based courses in abstract algebra, advanced calculus, number theory, discrete structures, and linear algebra, and he was enrolled in the second semester of both abstract algebra and advanced calculus at the time of the study.

4586 – 4586 was a senior with a double major in pure math and physics. He had a B+ grade average and had also taken proof-based courses in abstract algebra, advanced calculus, number theory, discrete structures, and linear algebra. He had also completed a course in complex analysis and was enrolled in the second semester of both abstract algebra and advanced calculus at the time of the study.

6772 – 6772 was a senior with a double major in applied math and fine arts and a minor in physics. She had a B+ grade average and had taken courses in discrete structures and linear algebra before enrolling in the study. At the time of the study, she was enrolled in an advanced calculus course.

Other – Three other comparison participants enrolled in the study and took the pre-assessment, but withdrew from the study before taking the post-assessment.

Implementation study.

Seminar.

Ethan – Ethan was a junior with a B+ grade average who was in the process of switching from an electrical engineering major to a secondary math education major. He had previously taken a course in geometry, and at the time of the study he was enrolled in a course on the history of mathematics that required proof composition. He said he was changing majors because his upper level engineering classes seemed like more of the same material he'd already been studying, and he was bored. However, he'd always enjoyed tutoring younger students in mathematics, and he'd enjoyed teaching when he had been in the military.

Greg – Greg was a senior with a psychology major and a minor in mathematics. He had a B grade average and had previously completed courses in abstract algebra and advanced calculus. He was enrolled in the advanced calculus course for the second time (to impove his grade), and a vector analysis class at the time of the study. Greg graduated at the end of the semester in which the study was conducted. He planned to take some time off and then go to graduate school to become a physician's assistant. He started as a biology major but changed to psychology because he was more confident in his ability to complete the degree. He chose his math minor because, as he put it, "I was really good in math until the theory part."

Nadia – Nadia was a junior secondary education major with a math minor, and she had an A grade average. Nadia had not previously completed any proof-based course, but she was enrolled in courses on discrete structures and the history of mathematics at the time of the study. In middle school, Nadia had been part of an enriched mathematics program that introduced her to non-standard problems and mathematical proof. She was hoping to pursue a secondary minor in English as well because she liked the contrast between having right answers in math and subjective answers in English.

Nick – Nick was a junior computer engineering major who was enrolled in a proof-based course on discrete structures. He enrolled in the study in the hopes that it would help him do well in his discrete structures class. Nick had a B grade average.

Tammy – Tammy was enrolled in her first semester of an applied math masters degree program at the time of the study, but she had been admitted deficient and was required to take prerequisite undergraduate math courses during her first year. She had a bachelors degree in chemistry and a masters degree in teaching, but had never taken undergraduate or graduate proof-based courses. She had taught high school chemistry and math for eight years and her enjoyment of the math instruction inspired her to return to graduate school. She was enrolled in an advanced calculus course at the time of the study.

Travis – Travis was a freshman planning to pursue a major in mathematics and a minor in biochemistry. In high school, he had taken a college-level course in advanced calculus at a large, public university. He was planning to major in math because he had always been interested in it and had always been good at it. At the time of the study, he was enrolled in a course on discrete structures and a course in vector analysis.

Usher – Usher was a senior math major with a minor in history. He had a B+ grade average and had previously taken courses in geometry and discrete structures. At the time of the study, he was enrolled in courses in abstract algebra and advanced calculus. He wanted to become a high school history and English teacher, but he felt that he would have an easier time getting a job as a math teacher, so he decided to major in math at the recommendation of his high school calculus teacher.

Other – One other seminar participant enrolled in the study, took the pre-assessment, was assigned to a group and attended two of the first four seminar sessions but then withdrew from the study.

Comparison.

1865 – 1865 was a senior math major with a minor in computer science. He had a B+ grade average and had taken courses in number theory and graph theory prior to enrolling in the study. At the time of the study, he was enrolled in abstract algebra and advanced calculus courses.

3099 – 3099 was a fifth year undergraduate double majoring in applied math and chemistry and pursuing a minor in world dance. She had a B+ grade average and had previously taken a course in advanced calculus. At the time of the study, she was enrolled in a course on Fourier analysis.

5105 – 5105 was a senior pure math major with a minor in English. She had a B+ grade average. 5105 had previously completed courses in number theory, discrete structures, advanced calculus, the history of mathematics, and linear algebra. At the time of the study she was enrolled in courses on abstract algebra and Fourier analysis. She was planning to pursue a masters degree in pure mathematics after completing her undergraduate degree.

5635 – 5635 was a senior math and Spanish double major with a B grade average. She had previously taken two semesters of advanced calculus and was enrolled in a course on discrete structures at the time of the study.

6293 – 6293 was a junior applied mathematics major with a minor in computer science. He had not previously taken any proof-based courses and was not enrolled in any at the time of the study.

Other - One other comparison participant enrolled in the study and took the pre-assessment but withdrew from the study and did not take the post-assessment.

Assessment Administration.

Seminar participants.

The assessments of the seminar participants were conducted in the presence of the researcher, and a think-aloud protocol was employed as participants attempted to construct the proofs. Because she was interested in participants' abilities to compose and validate proofs independently, the researcher conducted individual assessments of the seminar participants. Pre-assessments for the pilot study were conducted during the first four weeks of the spring 2012 semester, and post-assessments were administered during the final week and the first week following the spring 2012 semester. Pre-assessments for the implementation study were conducted during the second and third week of the fall 2012 semester; post-assessments were conducted during the final two weeks of the fall 2012 semester. All seminar participant assessments were recorded on video. For the implementation study, participants were provided with a LiveScribe Pen during both assessments and asked to do the composition portion of the assessment with it. This provided a second audio recording of that portion of the assessment, but did not significantly alter the assessment conditions for the participants.

Comparison participants.

The assessments of the comparison participants were conducted in the presence of the researcher. Pre-assessments for the pilot study were conducted during the second week of the spring 2012 semester, and post-assessments were taken during the final week of the spring 2012 semester. Pre-assessments for the implementation study were conducted during the second and third week of the fall 2012 semester; post-assessments were conducted during the final two weeks of the fall 2012 semester. For the implementation study, participants were provided with a LiveScribe Pen during both assessments and asked to do the composition portion of the assessment with it.

Assessments and Analysis

To answer the research questions, all study participants were administered a pre-assessment including a survey on the participants' attitudes and beliefs about mathematics and mathematical proof, three true number theoretical statements for the participants to prove, and four student-generated arguments for the participants to validate (See Table 3.1). Additionally, they were all asked to fill out a questionnaire about their mathematical backgrounds.

The seminar participants were assessed individually with the researcher observing. For these participants, a think-aloud protocol was employed during the proof construction, and the validations were discussed. The comparison participants were assessed individually, but at a common time and did not engage in discussion with the researcher about any portion of the assessment.

Area Assessed	**Format**
Participant background	Questionnaire with questions about year in school, GPA , and previously completed math courses.
Attitudes and beliefs about mathematical proof	(Pilot Study) 14 statements to be rated on a 5 point Likert-type scale. (Implementation Study) Eight open-ended questions. • Seminar participants responded in an interview. • Comparison participants provided written responses
Proof Composition Skills	Three true claims from number theory.
Proof Validation Skills	Four student-generated arguments attempting to prove a single claim.
(Seminar participants only) - Experience with Cooperative Learning	(Pilot Study) 11 open-ended questions. (Implementation study) Nine open-ended questions.

Table 3.1 - Assessment areas and format.

For the comparison participants, the post-assessment was identical to the pre-assessments. The seminar participants' assessments were also identical with the exception of an additional exit interview about their experiences as study participants that was conducted at the end of the post-assessment.

Questionnaire.

All participants were asked to fill out a demographic and mathematical background questionnaire at the start of the pre-assessment (see Appendix 1). The purpose of the questionnaire was to gather basic background data about the participants.

Background interview.

Seminar participants in the pilot study and the implementation study were then asked a series of questions about their mathematical backgrounds. They were asked when and why they decided to study mathematics and why they wanted to participate in the study.

Attitudes/beliefs.

Pilot study survey.

Five statements about mathematics and 14 statements about mathematical proof were presented to the participants who were asked to rate each statement on a 5-point Likert-type scale, with 1 indicating strong disagreement and 5 indicating strong agreement. The statements about mathematics came from the "Beliefs about Mathematics, Mathematics Learning, and Mathematics Teaching" survey developed by White, et. al. (2006, pg. 41), and the statements about mathematical proof were adapted from Almeida's survey (2000, pg. 872).

When comparing pre-assessment and post-assessment ratings for the 19 statements, no patterns were apparent, and without the ability to run statistical tests because of the small number of participants, the researcher could not make use of the data.

Attitudes/beliefs pilot study survey analysis.

For pilot study seminar and comparison participants, the numerical answers provided by the participants for each survey item were recorded. The researcher then noted changes from pre-assessment to post-assessment and looked for patterns in the data. However, no patterns were apparent, and the researcher determined that open-ended questions on the topics covered by the survey items would be preferable, so the Likert-type survey was adapted.

Implementation study interview and survey.

During the pre- and post-assessment interviews, the seminar participants in the implementation study were asked eight questions about mathematical proof. Based on the work of Weber (2010), and CadwalladerOlsker (2011), the researcher decided to look at how convinced participants were by rigorous proofs, and what the participants thought were the roles of mathematical proof. She also wanted to assess participants' familiarity with and knowledge of mathematical proof as well as their personal experiences with and level of enjoyment of proof. She adapted statements from the Almeida (2000) survey into interview questions that she designed to highlight these four areas: the roles of proof, conviction that proven results are valid, knowledge about proof, and personal experiences with proof. A final question was added to allow students to self-assess their own particular struggles with writing proofs (see Table 3.2) to further assess participants' personal experiences.

Interview Question	Targeted Area
How does mathematical proof differ from other kinds of proof?	Roles Conviction Knowledge
What is the purpose of writing proofs of theorems that are already known to be true?	Roles
Once you have seen a rigorous proof of a theorem, how confident are you that the theorem is true?	Conviction
Why does empirical evidence not count as proof?	Knowledge
Do you prefer proving or disproving claims? Why?	Personal
What do you like/dislike about writing proofs?	Personal
How confident are you in your ability to construct proofs?	Personal
What are the challenges you struggle with when constructing proofs?	Personal

Table 3.2 - Implementation Study Attitudes Questions

The comparison participants were presented with the same questions on the pre- and post-assessments and were asked to provide written responses to the questions. They were allowed to take the survey with them and bring their written responses back to the researcher within two days. Most participants took advantage of the offer.

A recording malfunction occurred during three of the seminar participants' post-assessments resulting in a loss of audio data for Greg, Tammy, and Travis. The researcher was able to schedule appointments with Tammy and Travis approximately five months after the post-assessments, but she was not able to reconnect with Greg. During the appointments with Tammy and Travis, the researcher conducted the Attitudes/Beliefs Interview again with the additional question "Do you think that any of your answers to these questions would have been different when I asked them of you in December?"

During the analysis (discussed below), these interviews were analyzed in the same manner as the successfully recorded interviews from the post-assessment. There are no post-assessment data for Greg for the Attitudes/Beliefs Interview.

Attitudes/beliefs implementation study interview and survey analysis.

Seminar participants' interview transcripts and comparison participants' written responses were coded with respect to conviction, proof roles, proof knowledge, and personal experience. For example, any participant response that mentioned how convincing that participant found formal proofs was coded as "conviction," and any mention of the purpose of writing proofs or of reading proofs was coded as "proof roles." Similarly, when participants talked about the relationship between empirical evidence and proof, why counter-examples are sufficient for disproving, or why proven mathematical theorems are proved forever but scientific proof can be overturned, the discussions were coded as "knowledge." Finally, any time a participant mentioned his or her own feelings about proof, preferences for certain kinds of proof, or personal challenges, those comments were coded as "personal experience."

Individual answers were compared from pre- and post-assessments and changes were documented. For example, some participants mentioned preferring disproving to proving on the pre-assessment, and a preference for proving on the post-assessment.

Composition.

The composition section of the pre-assessments and post-assessments consisted of three proof prompts in basic number theory. Each was presented as a true theorem to the participants. Elementary number theory prompts were chosen so the necessary concepts would likely be accessible to all of the subjects regardless of prior background.

While there is a general assumption in the research literature that content-independent, transferrable skills exist (e.g. Blanton & Stylianou, 2003; Sowder & Harel, 2003; Selden & Selden, 2003; Weber, 2003), this assumption has not been tested. Therefore, the researcher designed prompts to test additional proof skills that are commonly required for proofs in a variety of mathematical contexts. The specific skills tested were participants' ability to avoid a correct but appealing converse argument; to use indirect proving methods to construct a proof; to break a claim into pertinent subclaims and construct subarguments to form a proof; to use the details of an unfamiliar definition to form the basis for a proof; to recognize the need to construct two arguments to establish the validity of a biconditional claim; and to recognize the need for and apply the results of a previously proved claim.

The first statement, item C1, on the composition portion of the assessments was "if m^2 is odd, then m is odd." While this statement can be proved directly, the indirect proofs are straightforward and shorter. Being able to prove statements by indirect methods, such as proof by contradiction and proof by contrapositive, is an essential skill. Participants familiar with indirect proof methods could apply them to this prompt. At the same time, the converse of this statement, "if m is odd, then m^2 is odd," is more accessible and is established by a quick, direct proof, but it is not logically equivalent to the original claim. Students often erroneously equate conditionals and their converses (Selden & Selden, 2003), so this statement was also used to test the participants' ability to distinguish between the two and avoid the incorrect formulation. It was assumed that all participants would be familiar with even and odd integers and the mathematical context would be accessible to all.

The second statement, item C2, presented to participants was "if n is a natural number, then n^3-n is divisible by 6." Divisibility by 6 does not lend itself easily to examination while divisibility by 2 and divisibility by 3 are more common and easier to tackle. This item was included in the assessment to test participants' ability to break a statement into pertinent subclaims, and construct and combine subarguments to establish the validity of the claim. It was assumed that participants would be familiar with factoring.

A number of participants attempted to prove item C2 by induction, which is possible and does not require the construction of subclaims and subarguments. Therefore, the researcher was not able to assess participants' familiarity or facility with subclaims and subarguments. This will be further discussed in Chapter 6.

The final statement, item C3, included a definition as well as a claim. The definition provided during the pilot study conducted in the spring of 2012 was "A *triangular number* is defined as a natural number that can be written as the sum of consecutive integers, starting with 1." Each participant was asked to read the definition and discuss it with the researcher before attempting to prove the statement "a number, n, is triangular if and only if $8n$+1 is a perfect square." During the discussion about the definition, the researcher made sure that the participants could demonstrate understanding of the definition by providing examples of triangular numbers. The participants were also provided with a hint, "you may use the fact that $1 + 2 + \cdots + k = \frac{k(k+1)}{2}$." This item was designed to test three particular proving skills: knowledge of the logical implications of an "if and only if" statement, the ability to work with an unfamiliar definition, and the ability to use previously established results in the construction of an argument. This last skill was needed in the production of a valid argument for this statement because one direction of the biconditional relies on recognizing $8n+1$ as an odd number, and applying item C1. The researcher assumed participants would not be familiar with working with triangular numbers, although she did not assume that none of the participants would have seen the definition previously. She assumed participants would know the definition of a perfect square, and would be able to determine that $8n+1$ is an odd number. In the instances when participants did not know the definition a perfect square, she provided one.

The provided definition of triangular numbers proved to be problematic for almost all participants in the pilot study, so it was altered for the implementation study to an alternate, equivalent definition, "A *triangular number* is defined as a natural number that can be written as the sum of all positive integers less than or equal to a given positive integer, *k*."

All participants were allowed to return to this portion of the assessment after completing the validation exercises.

Composition Analysis.

Argument assessment tool.

Arguments produced during the pilot study assessments were initially analyzed using a tool developed by the researcher that was motivated by the classification options the participants used during the validation portion of the assessment (see Table 3.3); however, the researcher found the classifications to be too coarse to adequately discuss the arguments.

Code	Description
1	This is a rigorous proof of the claim.
2	This is a rigorous proof of a different claim.
3	This is a non-rigorous proof of the claim.
4	This does not meet the standards of a proof.

Table 3.3 - First Assessment Tool Used

In particular, most of the attempted proofs produced on the assessments fell into category 4, but there was no way to distinguish between arguments that were error-free but incomplete, arguments with errors, and responses in which no argument was presented. This caused problems when comparing participants' performance for appropriate cooperative grouping as well as comparing pre- and post- assessment performance by participants.

The researcher then used an argument assessment tool adapted by Boyle (2012) from a tool developed by Stylianides and Stylianides (2009) that provided broad argument categories, details within each category, and criteria to rate valid arguments and proofs on the basis of clarity and concision (see Table 3.4).

Main Codes	Code Details	Code Evidence
Incoherent or not addressing the stated problem (A0)	1. Solution shows a misunderstanding of the mathematical content. 2. Ignores the question completely. *3. Interprets claim, provides no argument.*	• List A0 and either 1, 2, or *3*.
Empirical (example based) (A1)	1. Examples are used to find a pattern, but a generalization is not reached. 2. Only examples are generated as a complete solution.	• List A1 and either 1 or 2
Unsuccessful attempt at a general argument (A2)	1. There is a major mathematical error 2. Illogical reasoning; several holes and or errors exist causing an unclear or inaccurate argument. 3. Reaches a generalization from examples, but does not justify why it is true for all cases. 4. Solution fails to covers all cases. 5. Solution is incomplete. Argument stops short of generalizing the stated claim.	• List A2 and match the bulleted number (1-5) in the middle column with the work in the solution.
Valid argument but not a proof (A3)	1. The solution assumes claims, in other words the solution exhibits a leap of faith before reaching a conclusion 2. The solution assumes a conjecture or lists a non-mathematical statement as a conjecture. *3. Argument is sound, but does not use mathematical notation and/or language - too informal*	• List A3 and either 1, 2 or *3* & address each of the points below **
Proof (A4)		• List A4 and address each of the three clear and convincing points below. **
** for use with A3 and A4. (+/-) The flow of the argument is coherent since it is supported with a combination of pictures, diagrams, symbols, or language to help the reader make sense of the author's thinking. Diagrams are fine as long as they are accompanied by an explanation. Explanation of ideas or patterns. (+/-) There are no irrelevant or distracting points. Variables are clearly defined and any terms introduced by the author are explained. Common understood language (+/-) The conclusion is clearly stated.		

Table 3.4 - Argument Assessment Tool (Details in italics did not appear in Boyle, 2012 and were added by the researcher.)

An attempt was coded A0 if no argument was made. This was the case when nothing was written down, if what was written did not relate to the stated prompt, or if the participant simply interpreted the claim by rewriting what it would mean for the claim to be true without providing support. Purely empirical arguments were coded A1; in these arguments, no variables appeared, and the author did not attempt to make generalizations from the examples produced. To be coded A2, arguments needed to have variables present in order to indicate an attempt at generalization, but they had to be flawed in such a way that they did not actually establish the truth of the prompt.

In order to be classified as either A3 or A4, an argument needed to be free of logical errors and mathematical errors. The distinction between these two categories lies in the level of justification and formality present in the valid arguments. An argument written informally without the use of mathematical notation was classified as A3 as was an argument making a necessary assumption without justification. This assessment tool was designed to be used in a classroom environment in which teacher and students would develop a shared understanding of which claims could be made without justification, but the researcher and participants did not have that opportunity, so the researcher determined which assumptions required justification. For example, she did not require justification for the claim that the sum of two even numbers is even, but she did require that participants justify that $8n + 1$ is odd since it is not written in the general form of an odd number.

Six examples of arguments produced by participants, two for each of the three prompts, were given to two mathematics faculty members along with the assessment tool, and the faculty members were asked to code the arguments using the tool; the researcher coded the same six arguments, and the codes assigned by all three individuals were then compared to test for reliability. All coders agreed on all of the main code designations, and after clarifying that multiple code details could be applied to a single argument, agreement was reached on the code details for each argument as well. This tool was then used to code all of the arguments produced by seminar and comparison participants from the pilot study.

When using the tool to code the arguments produced during the implementation study, the researcher found that the tool lacked the details necessary to describe some of the arguments. Thus, the tool was adapted by the researcher to include two more code details (see Table 3.4). The new details were discussed with the two faculty members who had previously tested the tool for reliability.

Argument analysis.

Participants' written proof attempts were examined, and each argument was coded using the argument assessment tool shown in Table 3.4. Each argument was assigned a main code, (A0, A1, A2, A3, or A4), and for arguments coded as unsuccessful proofs (all but A4), specific errors and flaws were noted and encoded. For successful arguments (A3 and A4), three specific aspects of clarity were addressed. Examples of arguments and codes can be found below.

When subclaims or subarguments were present in the work on item C2, each subargument was coded separately using the same scheme and the full argument was coded as well. Each direction of the biconditional in item C3 was coded separately, and then the full argument was also coded.

After applying the Argument Assessment Tool, participants' arguments were examined for evidence of the tested proof skills (see Table 3.5). For seminar participants, transcripts were examined for evidence of these skills; in several cases, participants did not express in writing that they knew they would have to prove both directions of the biconditional in item C3 but expressed the knowledge out loud.

Assessment Item	Proof Skill(s) Tested
C1. Prove: If m^2 is odd, then m is odd.	A. Use of indirect proof methods. B. Avoidance of a more appealing but logically inequivalent converse argument.
C2. Prove: If n is a natural number, then n^3-n is divisible by 6.	A. Ability to identify pertinent subclaims and construct subarguments (divisibility by 2 and 3).
C3. A *triangular number* is defined as a natural number that can be written as the sum of consecutive integers, starting with 1. Prove: A number, n, is triangular if and only if $8n+1$ is a perfect square. (You may use the fact that $1+2+...+k=\frac{k(k+1)}{2}$.)	A. Use of the specifics of a definition to form a basis for a proof. B. Ability to identify the logical implications of "if and only if" statements. C. Use of previously established results (to prove $8n+1$ a perfect square implies that n is triangular, the result of item C1 needs to be applied).

Table 3.5 - Tested Proof Skills by Item

(The definition for triangular numbers was altered for the implementation study.)

Once all proofs were coded, pre- and post-assessment evaluations of each item were compared for each participant to check for improvement, regression, or stasis.

Improvement was indicated by either an increase in main code (any argument coded as an A2 was seen as better than any argument coded as an A0 or A1), evidence of a proof skill on the post-assessment that had not been seen on the pre-assessment, or fulfillment of clarity criteria that had not been met on the pre-assessment.

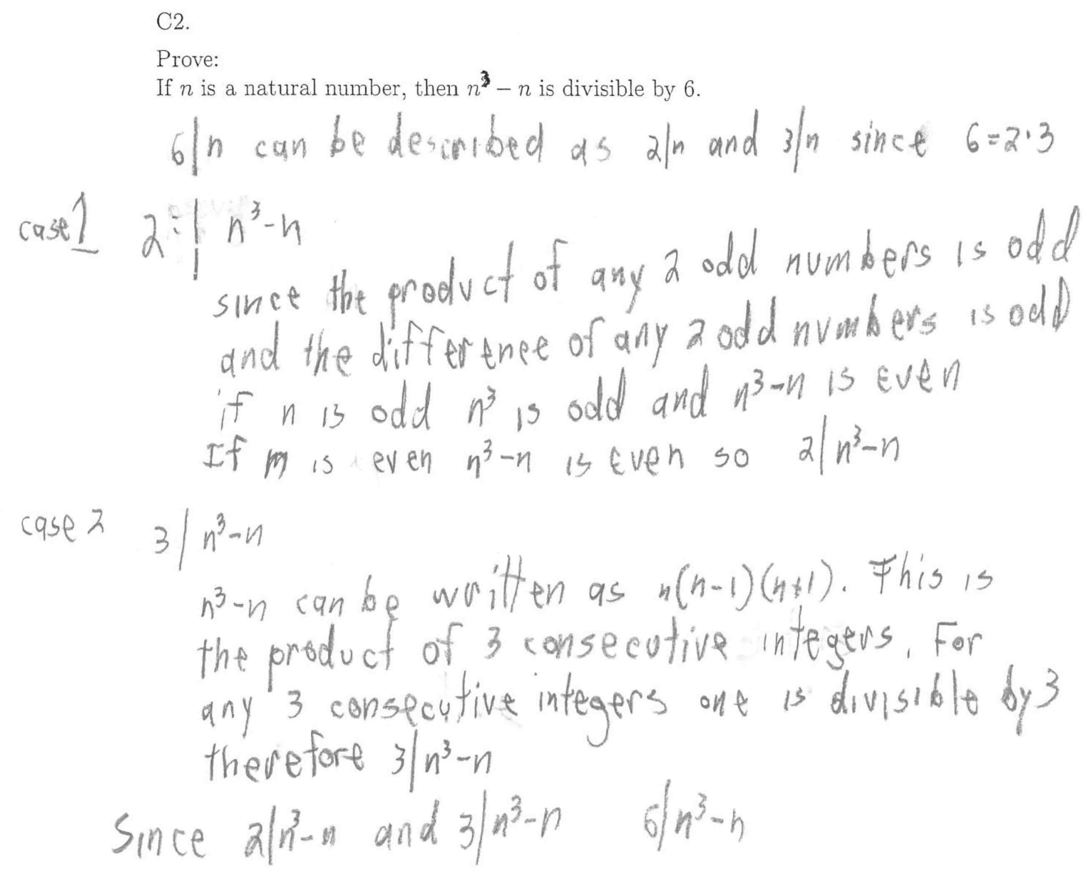
During the examination of the written work and transcripts, the researcher found that during the post-assessments, the seminar participants seemed to be much more willing to attempt different types of proof and change proving methods when stuck than they had on the pre-assessment. Because of this observation, the researcher went back to the transcripts and written work again and counted the number of times each seminar participant changed plans for each proof attempt. Any time a participant started working on a type of proof (direct, contradiction, contrapositive, or induction) or came back to an argument that had previously been abandoned, the researcher counted the action as a method switch.

When it became apparent that there was a general increase in switching methods from pre- to post- assessment for the seminar participants, the researcher used an open coding scheme to examine the seminar transcripts for evidence that the seminar supported this tendency.

Examples of coded arguments.

Ivan, item C2.

In Ivan's argument for divisibility by 2, he states "the difference of any 2 odd numbers is odd," which is mathematically incorrect (see Figure 3.1). However, he wrote the correct


C2.

Prove:

If n is a natural number, then $n^3 - n$ is divisible by 6.

$6|n$ can be described as $2|n$ and $3|n$ since $6=2\cdot 3$

case 1 $2 | n^3-n$

since the product of any 2 odd numbers is odd

and the difference of any 2 odd numbers is odd

if n is odd n^3 is odd and n^3-n is even

If m is even n^3-n is even so $2|n^3-n$

case 2 $3 | n^3-n$

n^3-n can be written as $n(n-1)(n+1)$. This is

the product of 3 consecutive integers. For

any 3 consecutive integers one is divisible by 3

therefore $3|n^3-n$

Since $2|n^3-n$ and $3|n^3-n$ $6|n^3-n$

Figure 3.1 - Ivan's Post-Assessment Work on Item C2

statement later in the subargument, and there are no other errors present. Therefore, the main code assigned to the subargument was A2. Ivan clearly attempted a general proof, but the error invalidated the argument. It was a mathematical error, so the corresponding detail code of 1 was added to the main code, and the subargument was coded A2.1. It is likely that the error in this argument was merely a transcription and not a conceptual error, but since Ivan made the same statement verbally before writing it down, the researcher could not make that assumption.

Ivan's subargument for divisibility by 3 is a valid argument, but it is too informal to be classified as a proof, so it was assigned a main code of A3. The level of formality expected in a proof was established during the seminar. The subargument meets all three clarity criteria, so the subargument was coded A3.+++.

When considered together, the two subarguments constitute a flawed attempt at a general argument with no errors other than the mathematical error in the divisibility by 2 subargument. Therefore, the full argument was also coded as an A2.1. Ivan clearly demonstrated the ability to break a claim into subclaims and construct subarguments, so the final code for Ivan's post-assessment work on item C2 was A2.1.A.

Greg, item C3.

Item C3 provided participants with the definition of a triangular number and asked them to prove that n is triangular if and only if $8n + 1$ is a perfect square. Because this is a biconditional statement, participants needed to prove both directions of the implication.

Greg proved that for n triangular, $8n + 1$ is a perfect square (see Figure 3.2), and the main code assigned to this direction was A4. On its own, line 6 is an incorrect mathematical statement: $4k^2 + 4k + 1 = (2k + 1)^2$, not $(2k + 1)$ as is written; however, when read in conjunction with line 5, the statement is that the square root of $4k^2 + 4k + 1$ is $(2k + 1)$, which is also what is reflected in the transcript of the session. So even though his notation is incorrect and could be confusing, it did not affect the

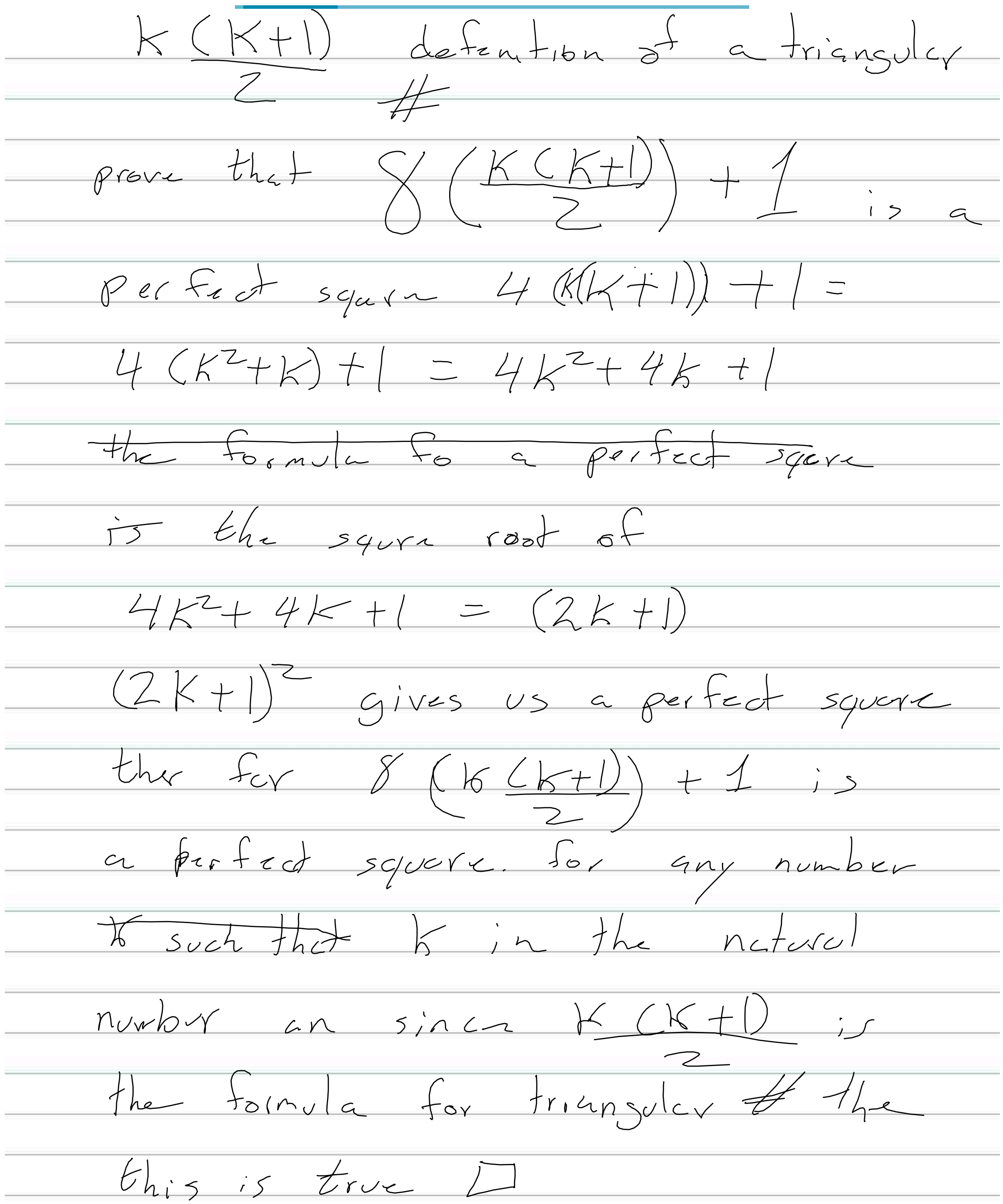


Figure 3.2 - Greg's Pre- Assessment Work on Item C3

validity of the argument. His argument is coherent, but the notation is an issue, and Greg did not use supportive language to "help the reader make sense of the author's thinking," so it failed to meet the first clarity criterion. Greg did not include any distracting or irrelevant points, but he also did not define all of his variables, since k is not specified as a positive integer and is not explicitly tied to n, so the proof did not meet the second clarity criterion: “There are no irrelevant or distracting points. Variables are clearly defined and any terms introduced by the author are explained.” Greg’s proof did not meet the third clarity criterion either since his concluding statement was “this is true,” which is not a clearly stated conclusion. Therefore, the code assigned to this direction was A4.---. Greg provided no argument for the reverse direction, so that direction was coded as A0.2.

On the whole, this was an unsuccessful attempt at a general proof, so the main code applied to the argument as a whole was A2. It was only unsuccessful because it was incomplete, so the code detail 5 was added to the main code. Greg demonstrated that he was able to use the specifics of a definition to form the basis of a proof, but he did not mention the need to prove the reverse direction of the biconditional, and he did not apply the claim in item C1, so he demonstrated only skill A of the three tested proof skills. Therefore, Greg’s pre-assessment argument for item C3 was coded A2.5.A.

Proof Validation.

In the validation section of the assessments, participants were asked to read 4 student-generated arguments for the claim “for any positive integer n, if n^2 is a multiple of 3, then n is a multiple of 3.” Participants were presented one argument at a time and were not allowed to return to previously classified arguments once they had seen a new argument. The four arguments are those Selden and Selden used to examine student validation skills (2003) (see Appendix 4a); however, the participants in this study were asked to classify each argument as either a rigorous proof of the claim, a rigorous proof of a different claim, a non-rigorous proof of the claim, an argument that did not meet the standards of proof, or an incomprehensible argument. These classifications were adapted from the scale developed by Weber to examine student perception of conviction, validity, and proof (2010, pg 317). Participants were asked to explicitly identify which, if any, errors they saw, and if they classified an argument as a rigorous proof of a different claim, the participants were asked to state what alternate claim had been proven.

Because some participants had very limited prior experience with proof, the researcher discussed the distinction between a non-rigorous proof of the claim and an argument that did not meet the standards of a proof with each participant before presenting the first argument. Non-rigorous proofs were defined as arguments that established the validity of the claim but may have had minor errors or not included some pertinent justification. Arguments that did not meet the standards of a proof were described as arguments with logical errors or other mathematical errors that affected the validity of the argument. All classifications were also further clarified upon request.

Participants in the pilot study were presented the arguments in the order they appeared in the Selden and Selden paper – “Errors Galore,” “The Real Thing,” “The Gap,” and “The Converse.” However, many of the participants struggled with Item EG, “Errors Galore,” to the extent that it seemed to affect their confidence and willingness to proceed, so the researcher changed the order in which the arguments were presented for the implementation study – “The Converse,” “The Real Thing,” “Errors Galore,” and “The

Gap." The names of the arguments are also the Seldens' and were not seen by the participants.

Proof Validation Analysis.

Error Classification.

The researcher analyzed the four arguments attempting to prove the claim "For any positive integer n, if n^2 is a multiple of 3, then n is a multiple of 3," presented by Selden and Selden (2003) and enumerated the errors present in each argument (see Appendix ……). She consulted the Selden paper for this enumeration but disagreed with some of their findings. Below is a discussion of which errors, if any, were present in each argument. These errors were used as a basis to evaluate the classifications given by the study participants. If students justified a classification by citing perceived errors that were not actually errors and/or not citing the errors present in the argument, their classification was determined to be incorrect.

Item EG: "Errors Galore."

> PROOF. [1] Assume that n^2 is an odd positive integer that is divisible by 3. [2] That is $n^2 = (3n+1)^2 = 9n^2 + 6n + 1 = 3n(n+2) + 1$. [3] Therefore, n^2 is divisible by 3. [4] Assume that n^2 is even and a multiple of 3. [5] That is $n^2 = (3n)^2 = 9n^2 = 3n(3n)$. [6] Therefore, n^2 is a multiple of 3. [7] If we factor $n^2 = 9n^2$, we get $9n^2 = 3n(3n)$; which means that n is a multiple of 3.

This item could only be correctly classified as not meeting the standards of a proof. However, whenever a participant did not understand the argument, their choice would not be counted as correct or incorrect. There are many errors in the argument starting in [2]. First, the author uses n to represent two different quantities. Second, the author states an assumption in [1] about n^2, namely that it is a positive integer that is divisible by 3, but goes on to set up the equation $n^2 = (3n + 1)^2$ in [2] which applies an assumption to n, not to n^2. Third, the assumption made about n, that $n = 3n + 1$, is not related to parity nor to divisibility by 3. There is also a factoring error in [2]: $9n^2 + 6n + 1 \neq 3n(n + 2) + 1$. Sentence [3], "Therefore, n^2 is divisible by 3," is problematic because it is presenting as a conclusion what was already assumed to be true. It is also a false statement given the work preceding it. [5] exhibits some of the same errors present in [2]; while the author states a claim about n^2, the assumption that n^2 is a multiple of 3 has been applied to n instead, and the author again uses n to represent two distinct quantities. In [6] the author again concludes something that was assumed to be true. The concluding statement of the argument "which means that n is a multiple of 3" is problematic because it does not follow from any of the work preceding it.

About the errors in [2], Selden and Selden claim "because they do not affect the correctness of [3], they cannot affect whether or not the argument is a proof." However, because the participants in this study were asked to differentiate between rigorous and non-rigorous proofs of the argument, these errors are important to this study.

Item RT: "The Real Thing."

> PROOF. [1] Suppose to the contrary that n is not a multiple of 3. [2] We will let $3k$ be a positive integer that is a multiple of 3, so that $3k + 1$ and $3k + 2$ are integers that are not multiples of 3. [3] Now $n^2 = (3k + 1)^2 = 9k^2 + 6k + 1 = 3(3k^2 + 2k) + 1$. [4] Since $3(3k^2 + 2k)$ is a multiple of 3, $3(3k^2 + 2k) + 1$ is not. [5] Now we will do the other possibility, $3k + 2$. [6] So, $n^2 = (3k + 2)^2 = 9k^2 + 12k + 4 = 3(3k^2 + 4k + 1) + 1$ is not a multiple of 3. [7] Because n^2 is not a multiple of 3, we have a contradiction.

This item could have been correctly classified either as a rigorous proof of the claim or a non-rigorous proof of the claim depending on the participants' justifications for their classifications. Selden and Selden consider this an error-free proof, but this researcher feels that the author needed to explicitly define the variables used and thinks it could be written more clearly. There is an implicit assumption that n^2 is a multiple of 3 which sets up the contradiction claimed in [7], but this could also be considered a valid proof by contrapositive if the contradiction weren't claimed. To expert readers, the implicit assumption establishing the contradiction is clear, but it may not be clear to novice readers who then may judge the proof to be non-rigorous. Because the logic and algebraic manipulations are correct, this argument should have been classified as a valid argument.

Item GP: "The Gap."

> PROOF. [1] Let n be an integer such that $n^2 = 3x$ where x is an integer. [2] Then $3|n^2$. [3] Since $n^2 = 3x$, $nn = 3x$. [4] Thus, $3|n$. [5] Therefore if n^2 is a multiple of 3, then n is a multiple of 3.

The only significant error in this argument is the jump in reasoning between [3] and [4]. While the claim presented in [4] is true, presenting it here without justification amounts to stating that the original claim is true without proof. [4] follows from [3] because 3 is prime and if a prime number divides a product, then it necessarily divides one of the factors. The leap in reasoning makes this a non-rigorous proof of the theorem. However, it could also be considered as not meeting the standards of a proof. It also would have been more appropriate to define n as a positive integer, but that is a minor omission.

Item CV: "The Converse."

> PROOF. [1] Let n be a positive integer such that n^2 is a multiple of 3. [2] Then $n = 3m$ where $m \in \mathbb{Z}^+$. [3] So $n^2 = (3m)^2 = 9m^2 = 3(3m^2)$. [4] This breaks down into $3m$ times $3m$ which shows that m is a multiple of 3.

[2] and [3] establish the validity of the converse of this theorem, but [2] does not follow from [1] since the author started out assuming that n^2 was a multiple of 3, but set up [2] with n as a multiple of 3. If one is only considering [2] and [3], this is a rigorous proof of the converse of the claim, and participants classifying it this way were considered correct

depending on their justification. However, [2] absolutely does not follow from [1], and [4] concludes something about m which was unrelated to the original claim, and the conclusion reached is unfounded. Therefore, this argument could also be classified as not meeting the standards of a proof.

Proof Validation Evaluation.

The researcher began analyzing these data by recording the classifications and corresponding justifications given by each participant to determine whether the participants' classifications were correct. She then looked for patterns of improvement and/or regression by assessment item for each participation group in each study.

The researcher then examined the written justifications of each comparison participant and the transcripts of the seminar participants to evaluate individuals' ability to identify errors and reasonably classify the arguments. The researcher created a spreadsheet to record which errors were mentioned by each participant on each assessment. Errors were then coded by how many of the participants mentioned each one.

Exit Interview.

The exit interviews were conducted only during the post-assessments of the seminar participants. These interviews were designed by the researcher to explore the participants' perception of their experiences in the research study. Participants in the pilot study were asked four questions about how their participation in the research affected their confidence about proving and validating, five questions about their experience working as part of a cooperative group during the seminar, and two final questions about mathematical proof (see Appendix 5).

The final two questions, “what does proof mean to you” and “how has that changed as a result of your participation in this research,” were not included on the exit interview for the implementation study and were instead implicitly incorporated into the interview about attitudes and beliefs that replaced the Likert-type survey.

Due to the same recording malfunction that affected the Attitudes/Beliefs data, the Exit Interview audio data for Greg, Tammy, and Travis were lost. During the appointments with Tammy and Travis that were held approximately five months after the post-assessments, the researcher conducted the Exit Interview again. During the analysis (discussed below), these interviews were analyzed in the same manner as the successfully recorded interviews from the post-assessment. There are no post-assessment data for Greg for the Exit Interview.

Exit interview analysis.

The researcher examined the responses to the first question, “how has your confidence level about constructing proofs changed as a result of your participation in this research” (see Appendix 5), and grouped them according to response: increased, decreased, or did not change. She then read through the responses to the second question, “what are the factors you think most contributed to that change,” for each group to identify recurring themes. The same process was repeated for the corresponding questions about reading and understanding proofs. Responses to question five, “how did working as a member of a cooperative group affect your learning,” were also grouped by response: positively, negatively, or not at all, and rationales were examined for common themes. The responses to questions six, seven, and eight were examined individually to check that the researcher had met the goals of Cooperative Learning during the seminar. The answers to question nine, “do you feel better able to work cooperatively as a result of your participation in this research,” were examined in the same manner as those provided for question five with response groupings of yes and no. For the seminar participants from the pilot study, the responses to questions ten and eleven were examined for common themes.

Seminar.

During the seminar sessions between assessments, the seminar participants worked on problem sets in assigned cooperative groups. Utilizing results of previous research (Cooper, 1990; Harskamp, et. al., 2007; Heller & Hollabaugh 1992), the researcher formed the groups so the female participants were not outnumbered and so the groups were heterogeneous based on skill level as demonstrated on the pre-assessment. The group assignments were not changed during the course of the studies. The members of each group spent a few minutes at the beginning of each session getting to know each other and 5-10 minutes at the end of each session doing group processing exercises. Both of these activities facilitated the development of the social skills necessary for effective cooperative work, and the rotating roles the participants assumed each session were assigned to assure the group members’ personal accountability and positive interdependence (Heller & Hollabaugh, 1992, Prescott-Johnson, 1992). After a brief introduction to the problem set by the researcher each session, the participants worked primarily with each other while the researcher functioned as a facilitator, encouraging student-to-student interactions.

The problem sets dealt specifically with function concepts rather than the number theoretic concepts from the assessments. This was done so that any changes from pre- to post-assessment would reflect changes in the subjects' ability to construct and validate proofs instead of changes in knowledge of number theory based on additional practice. This presupposes the existence of transferrable, content-independent proof skills, as is common in the literature.

Function concepts were specifically chosen because participants would be familiar with the basic concept regardless of background and so the participants would get more practice working with abstract functions and developing a deeper understanding of an important mathematical concept.

Materials.

During the pilot study, the researcher gave each participant a copy of each problem set. However, the groups in that study struggled to work cooperatively; participants often worked individually and shared findings, progress, and frustrations with their group mates. The participants from that study also expressed frustration with not having definitions close at hand. To combat the tendency to work individually the researcher provided the implementation groups with only one copy of each problem set per group. She also made a definition sheet for each problem set which was provided to each participant.

Problem set on inverses and pre-images (see Appendix 6a).

The goal of this problem set was to have the participants build a deep understanding of invertible functions. Many of the problems explored examples of functions with domains and codomains that were non-standard. The problem set was also written so students could get used to the difference between pre-images and inverses. The researcher chose inverses for the first set because the context would be familiar to the participants.

This problem set was abandoned for the implementation seminar for several reasons. First, the participants' pre-conceived notions of inverses made it difficult for them to explore the questions on a deep level; they continually reverted to definitions they already knew and did not engage with the material at the intended level. Second, the problem set included too many problems geared towards building and exploring examples, and there were consequently too few questions requiring the groups to produce proofs. Finally, all of the problem sets took more time than the researcher had expected. The participants in the pilot study did not finish any of the problem sets and they did not have the opportunity to work with the final problem set for more than a day. Because the problem set on inverses was problematic, the researcher decided to remove it from the seminar and have the implementation participants work with the other three problem sets more extensively.

Problem set on the Gaussian integers (see Appendix 6b).

For the pilot study, the second problem set was *Pythagoras, Gauss, and Norm.* This set was adapted from early materials developed by Cuoco and Rotman that led to their modern algebra text (2013) and was chosen to give the participants a chance to work with a specific function with an uncommon domain and codomain and to make and prove claims about that specific function.

On problem 4, the participants were asked to use the provided function, $N: \mathbb{Z}[i] \to \mathbb{N}$ such that $N(a + bi) = (a + bi)(a - bi)$, to prove that any natural number of the form $4k + 3$ cannot be written as the sum of two squares. This problem was difficult for the participants, and they made little progress.

For the implementation study, this was the first problem set. It was renamed *Gaussian Integers and the Norm* and some problems were adjusted to address issues that arose with the first version. Problem 4 was broken into three parts to guide the participants through the proof of the claim. They were first asked to show that 3, 7, and 11 cannot be written as the sum of two squares. Then they were asked to show that the sum of two even squares is divisible by 4 and that the sum of two odd squares is even but not divisible by 4. Finally, they were asked to determine what other case needed to be checked to establish the truth of the claim and to write a formal proof of the claim. This restructuring lessened the cognitive demand on the participants and increased the likelihood that the groups would be able to successfully complete the problem.

One of the examples in problem 5 was changed so that a new example did not need to be established in problem 6. In the pilot study, participants were asked to find all $a, b \in \mathbb{Z}$ such that $N(a + bi) = 35$, but 35 is of the form $4k + 3$, so the participants had already proved that there were no such $a, b \in \mathbb{Z}$. Problem 6 in both versions of this problem set asked participants about $N(c + di) = 10$, but in the pilot study, participants had not previously been asked to look at that particular result. So in the implementation study, 5f, find all $a, b \in \mathbb{Z}$ such that $N(a + bi) = 35$ was replaced with find all $a, b \in \mathbb{Z}$ such that $N(a + bi) = 10$. While this eliminated an opportunity for the participants to recognize the problem as an instantiation of an impossible fact and to apply the results of their prior work, the time considerations were given higher priority. Problem 8 on the pilot study version is an application of the result proved in problem 7, so problem 8 was omitted from the later version. A list of definitions applicable to the problem set (see Appendix 6b) was provided during the implementation study, so the wording of problem 9 was altered.

Problem set on fixed points and a derivative-like functions (see Appendic 6c).

The third problem set for the pilot study was *Derivatives and Fixed Points* and contained content primarily developed by the researcher. It defined a function, *D*, that returned the usual derivative of a real polynomial. Participants were asked to apply *D* to specific polynomials and prove specific properties of *D*. Participants were then provided with the definition of a fixed point and asked to work with specific functions and their fixed points before being asked to prove that *D* had a unique fixed point. Two of the proof problems on this problem set asked participants to outline how they might prove the claims using different methods (direct, contradiction, and contrapositive) and then choose a method to use in their proofs. This particular practice was adopted from Schumacher's introduction to proof text (2001) and used to familiarize participants with different proof methods and facilitate the choice of productive arguments.

This problem set was adapted to be the second problem set in the implementation study, *A Polynomial Function and Fixed Points*. Participants in the pilot study had relied heavily on their prior understanding of the derivative when working on the problems, and that reliance hindered their ability to prove the claims about the function, *D*, in sufficient detail. Therefore, the researcher removed the language about derivatives for the implementation study and renamed the function as *K*. Some students noticed that *K* behaved like the derivative, but that did not lead to the same conflicts the researcher had observed during the pilot study. In order to outline the arguments by different methods, participants needed to identify the hypothesis and conclusion of the presented claims, but participants in the pilot study did not recognize that need, so implementation participants were explicitly asked to translate the claims into implication statements. Apart from rewording two problems for clarification, no other changes were made to the problem set.

Problem set on injectivity and surjectivity (see Appendix 6d).

The final problem set from both the pilot and implementation studies dealt primarily with the concepts of injectivity and surjectivity The functions presented in this problem set were the most challenging the participants worked with during the seminar as some of the inputs involved other functions as well as ordered pairs. This allowed participants to work with functions in a less familiar context as individual problems dealt with functions simultaneously as inputs and relations. The focus on the concepts of injectivity and surjectivity was intentional; through her experiences teaching undergraduate math classes, the researcher had found that students struggled with understanding the concepts despite their common use. In particular she had found that students were reluctant to separate the notion of injectivity from the "horizontal line test" which is not widely applicable. Problems 3 and 4 dealt with the relationship of injectivity and surjectivity with function composition and were based on problems 5 and 6 in Chapter 5 of Schumacher's text (2001, pg 131). The participants were given examples to motivate the formation of a conjecture, but one of the examples was incorrectly set up (see Figure 3.3). This example was fixed for the implementation study.

3. Let $A,\ B,\ C$ be sets and $f,\ g,\ h$ be functions.

a. Given $g : A \to B$, and $f : B \to C$, find $f \circ g$.

i. $A = B = C = \mathbf{R}$, $g(x) = x,\ f(x) = x^2$.

ii. $A = B = C = \mathbf{R}$, $g(x) = x + 1,\ f(x) = x^3$.

[redacted]

$g(\alpha, \beta, \gamma) = \alpha x^2 + \beta x + \gamma$, and $f : B \to C$ is given by $f(p) = D(p) = p'$.

Figure 3.3 - Problem 3a. from *–jectivity*.

The range of g in part iii is not contained in the stated codomain.

This problem set was renamed, *Injectivity, Surjectivity, and Function Composition*, for the implementation study and altered slightly. As mentioned above, the error in problem 3 was corrected, 3a.*iv* was altered to match changes made to the derivative-like function from the previous problem set, examples were added to problem 4, and the wording in 3a was altered to be less confusing (See Figure 3.4).

3. Let $A,\ B,\ C$ be sets and $f,\ g,\ h$ be functions.

a. Given $g : A \to B$, and $f : B \to C$, find $f \circ g$, and state domain and codomain for the composition.

i. $g : \mathbf{R} \to \mathbf{R}$, $g(x) = x$, $f : \mathbf{R} \to \mathbf{R}$, $f(x) = x^2$.

ii. $g : \mathbf{R} \to \mathbf{R}$, $g(x) = x + 1$, $f : \mathbf{R} \to \mathbf{R}$, $f(x) = x^3$.

iii. $g : \mathbf{R}^2, \to \mathbf{Z}[i]$, $g : A \to B$ is given by $g(\alpha, \beta) = \lfloor \alpha \rfloor + \lfloor \beta \rfloor i$, and $f : \mathbf{Z}[i] \to \mathbf{N}$ is given by $f(a + bi) = N(a + bi) = a^2 + b^2$.

iv. $g : \mathbf{R}^3, \to \mathcal{P}$, is given by $g(\alpha, \beta, \gamma) = \alpha x^2 + \beta x + \gamma$, and $f : \mathcal{P} \to \mathcal{P}$ is given by $f(p) = K(a_n x^n + ... + a_1 x^1 + a_0 x^0) = n a_n x^{n-1} + ... + a_1 x^0 + 0 a_0$.

Figure 3.4 - Problem 3a from *Injectivity, Surjectivity and Function Composition*

Cooperative group structure.

Participants worked on these problem sets in cooperative groups based on the principles of Cooperative Learning. The groups were assigned by the researcher and were not changed over the course of either study. In each session, group members were assigned roles they were to fulfill while they worked together on that session's problem set. At the end of each session, groups engaged in a processing activity in which they reflected on their group cohesion and efficacy for that day.

Group assignment.

Appropriate grouping is one of the central features of Cooperative Learning. Groups ideally have between three and five members and are as diverse as possible regarding achievement as well as gender (Cooper, 1990; Heller & Hollabaugh 1992). Research suggests that for gender imbalanced groups, it is better to have more female group members than male. This is because it has been shown female students are more likely to engage with the material and have better learning outcomes in such situations (Heller &

Hollabaugh, 1992, Harskamp et al. 2007). These principles where employed in the formation of the groups in both studies.

Pilot study.

Eight seminar participants enrolled in the study and took the pre-assessment, but one of the participants dropped out of the study before the first meeting. The researcher therefore formed two groups, one with four participants and one with three participants. The two female participants, Ingrid and Ursula, were assigned to one group, and the other assignments were made based on pre-assessment performance. Ingrid and Ursula were two of the three participants who performed the worst on the composition portion of the pre-assessment, so Omar, who performed the worst of all participants, was assigned to the other group. Ivan, Bill, and Zach had all performed fairly well on the pre-assessment composition tasks, and the researcher assigned Bill to the four person group with Ingrid, Nathan, and Ursula, and had Ivan, Omar, and Zach making up the three person group.

Implementation study.

Eight seminar participants enrolled in the implementation study and took the pre-assessment. There were two female participants and six male participants, so the two female participants were assigned to a group together. Karen and Nick had the weakest performances on the pre-assessment and so Nick was assigned to the other group. Ethan and the participant who withdrew from the study performed similarly on the pre-assessment compositions, so they were assigned to different groups with Ethan arbitrarily assigned to the group with Nick. Travis, Nadia, and Greg had all performed similarly to each other on the pre-assessment as well, while Usher had out-performed all of the other participants by a significant margin. One of Greg's principal struggles with proof composition involved getting his thoughts onto the page while Karen struggled with producing productive thoughts but clearly demonstrated that she knew what proof language was generally used. The researcher thought those skill sets would be complementary, so Greg was assigned to the group with Karen, Nadia, and the participant who withdrew. Travis was assigned to the group with Nick and Ethan. Usher was then assigned to Nick, Ethan, and Travis' group so that both groups would have four members. The participant who withdrew dropped out of the study after four seminar sessions.

Instructor as facilitator.

During the seminar sessions, the researcher acted as a facilitator according to the principles of Cooperative Learning (Cooper, 1990). During the first meeting of both studies, she discussed the tenets of Cooperative Learning and the format of the seminar to clarify expectations and to begin a conversation about the social skills needed for the group work to be successful since those social skills are critical to successful Cooperative Learning and should not be taken for granted (Towns, 1997; Jones & Jones, 2008). Immediately following that introductory discussion, and at the beginning of each subsequent seminar session, the researcher gave the groups "conversation starter" questions to foster team building (Prescott-Johnson, 1992; Towns, 1997) and then gave a fifteen to thirty minute lecture on material the participants would need in order to work on the problem sets. These lectures included discussions about the differences between direct proof, proof by contradiction, and proof by contrapositive as well as daily reminders of the definition of function. She also talked about domain, range and codomain, injectivity, and surjectivity. The researcher asked frequent questions during the lectures, and all lectures required active participation from the participants. These lectures were necessary because the parameters of the study prevented the researcher from assigning work to the participants outside of the seminar sessions.

After the lectures, participants were asked to work on the problem sets discussed above. While the groups worked, the researcher walked around the room listening to the conversations, and prompting conversation when needed. During the pilot study, the group members tended to prefer to work individually and then share solutions or concerns with their group mates rather than work cooperatively, so the researcher spent a lot of time asking the participants to talk to each other and reminding them of their roles. She did not have to encourage the implementation study groups to work together because they were engaged as groups in the work without such prompting. In both studies, the researcher engaged with the groups on the mathematics of the problem sets when the group members seemed unable to get past a hurdle or were coalescing around an incorrect idea. In both situations, she asked the group members leading questions and helped guide them past obstacles or to the recognition their error. She also engaged with the groups when they asked her questions, but the interactions were very similar in that she answered questions with more questions and tried to get the participants to reach their own conclusions.

On one occasion during the pilot study and two occasions during the implementation study, the researcher had the groups engage in validation exercises. She had each group write a completed proof on the board and asked the other group members to critique it while the presenter made adjustments to the presented proof. Only at the end of the critique by the participants did the researcher provide feedback on the argument. In each situation, the researcher only had to address issues of clarity and readability; the participants were always able to fix any serious flaws through their discussion.

At the end of each seminar session, the researcher asked the groups to complete a group processing exercise. For the first four sessions in each study, she had individuals take a few minutes to write down individual answers to the processing questions and only had them speak with their group mates about their responses after this period of individual reflection. Once the group members seemed more comfortable together, the researcher allowed the participants to skip the individual step and engage in the group processing without writing individual responses first. During the group processing, the researcher walked between the groups and listened to their conversations to check for any serious group dynamic issues that needed to be addressed, but she did not hear evidence of any such issues.

Group member roles.

Positive interdependence is a central facet of Cooperative Learning; one of the ways to foster this sense of responsibility for the learning of one's group mates is by having assigned, rotating roles (Cuseo, 1994; Heller & Hollabaugh, 1992; Prescott-Johnson, 1992a). Such roles also help jumpstart conversation and keep all group members involved in the work. In well-functioning groups, members often spontaneously take up and trade roles (Heller & Hollabaugh, 1992), but assigning roles is beneficial to new groups and to group members who are not accustomed to cooperative group work.

Pilot study.

For the pilot study, the researcher assigned roles adapted from Heller and Hollabaugh. There was one group of three participants and one group of four participants. The roles for the group of three were Manager, Skeptic, and Checker/Recorder. The group of four also had an assigned Explainer each week. The roles rotated weekly so that each participant was required to fulfill each role, and the participants were informed that if a member was absent, the other participants would need to make sure there was always a Manager and a Checker/Recorder, and that if three members were present, the Skeptic role took priority over the Explainer role. On one day, only one member of the four-person group was present, and he worked with the other group for the session.

Each session, the groups were given folders that contained the problem sets, prior work done on the problem sets, a schedule of role assignments, and a description of each role (see Appendix 7). The description of roles included primary responsibilities for each role as well as specific sentence starters and questions that would help a participant fulfill his/her role for the day. For example, the Explainer's responsibility was "explains and summarizes," and recommended statements were "That follows because…, So basically this proof says that…, and Intuitively, this means…." The Explainer role proved to be problematic because participants were not working on the problem sets outside of the seminar sessions, and the Explainer had no opportunity to prepare. Thus, the participant in the explainer role was often not able to explain to and educate his/her group mates. That role was adjusted for the implementation study.

Implementation study.

Eight seminar students were initially enrolled for the implementation study, and two groups of four were formed. The Manager and Skeptic roles were transferred from the pilot study, and the Checker/Recorder role was adapted into a Recorder/Presenter role. Recording the groups' daily work was a demanding job and the researcher felt that no additional duties needed to be assigned to the Recorder. Even though it wasn't part of the role title, presenting the group's work was one of the Checker/Recorder's responsibilities in the pilot study as well. During the implementation study, the Recorders each used LiveScribe pens and notebooks to record his/her group's work. The Explainer role was abandoned and a new role, Yes-Man, was created. Instead of being responsible for explaining everything to his group mates, the Yes-Man was responsible for confirming the claims and assertions made by his group mates, checking for consensus and understanding, and furthering conversation one step at a time. This role was inspired by a talk given at the 2012 Joint Mathematics Meeting encouraging mathematics instructors to incorporate theater improvisation techniques and activities into their classes. In particular, it was suggested that students be asked to further discussion by saying "yes, and…" (Young, 2012). The primary statement expected of the Yes-Man was "Yes, and…" which would reinforce the idea that you can always take one more step even if you don't know how to get all the way to your conclusion.

Group processing.

The participants were given 5-10 minutes at the end of each session to discuss as a group how well they worked together during that session. The discussion was prompted by set questions presented by the researcher, and the participants were asked to write down their answers individually and then share them with their group members when everyone was ready. The questions were printed on the inside of the groups' folders. This processing exercise was meant to facilitate group cohesion and efficacy (Johnson, Johnson, & Smith, 2007), and the participants' responses were not analyzed as part of this study.

Pilot study.

The participants were asked three questions and told to be open and honest and considerate.

- What are three ways you worked well together today?
- What problems did you have interacting well as a group?
- What concrete steps could you take next time to interact as a group more effectively?

Implementation study.

Because the participants in the pilot study needed fairly constant reminders to fill their roles two group processing questions were added to those used in the pilot study for the implementation study.

- How well did you fulfill your assigned role? Explain

- Did your group mates fulfill their roles well? Explain.

Chapter 4 - Pilot Study Results

Seminar Participants Overview

Attitudes and beliefs.

Participants were given a survey with 19 statements about mathematics and mathematical proof and asked to rate each statement on a five point Likert-type scale. Participant responses to the attitudes and beliefs survey can be found in Appendix 2b As discussed in Chapter 3, the researcher could not make use of these data, and this portion of the assessment was abandoned for the implementation study.

Composition.

The composition portion of the assessments consisted of four true, number theoretic claims that the participants were asked to prove (see Table 4.1). The pre-assessment and post-assessment items were identical. The researcher used the Argument Assessment Tool she adapted from Boyle (2012) to analyze the participant-generated arguments and compared each participants' pre-assessment composition with the corresponding post-assessment composition for each item (see Appendix 3b for the AAT). Improvement on each item was defined as an increase in the main argument code, fulfillment of clear and concise criteria that had been lacking, or evidence on the post-assessment of a tested proof skill that was not apparent on the pre-assessment. For example, any post-assessment argument with a main code of A3 was seen as an improvement over a corresponding pre-assessment argument rated A2 or A1, and a post-assessment argument rated as an A3.1.--- would have been seen as a regression from the corresponding pre-assessment argument rated as an A3.1.-++.

Assessment Item	Proof Skill(s) Tested
C1. Prove: If m^2 is odd, then m is odd.	A. Use of indirect proof methods. B. Avoidance of a more appealing but logically inequivalent converse argument.
C2. Prove: If n is a natural number, then n^3-n is divisible by 6.	A. Ability to identify pertinent subclaims and construct subarguments (divisibility by 2 and 3).
C3. A *triangular number* is defined as a natural number that can be written as the sum of consecutive integers, starting with 1. Prove: A number, n, is triangular if and only if $8n+1$ is a perfect square. (You may use the fact that $1+2+...+k=\frac{k(k+1)}{2}$.)	A. Use of the specifics of a definition to form a basis for a proof. B. Ability to identify the logical implications of "if and only if" statements. C. Use of previously established results (to prove $8n+1$ a perfect square implies that n is triangular, the result of item C1 needs to be applied).

Table 4.1 - Composition Items and Tested Proof Skills

Bill improved on item C3 because the main code for the produced arguments increased from A2 to A4 (see Appendix 3c), while Ursula improved on item C2, despite the fact that item C2 was rated an A2 on both assessments, because she demonstrated an ability to construct subclaims and subarguments on the post-assessment but not on the pre-assessment.

Of the seven seminar participants, six showed distinct improvement on at least one item from pre-assessment to post-assessment (see), and of the eighteen composition item comparisons for these six participants there was improvement on twelve, stasis on five, and regression on only one. The six participants who improved all had at least one item on which the main code improved and also demonstrated at least one proof skill on the post-assessment that had not been demonstrated on the pre-assessment. The seventh participant, Zach, regressed on one of the composition items and showed no improvement on the other two.

Participant	Item C1	Item C2	Item C3
Bill	**Improvement** *Main Code*	**Stasis**	**Improvement** *Main Code* *Proof Skill*
Ingrid	**Improvement** *Main Code* *Proof Skill*	**Stasis**	**Improvement** *Main Code*
Ivan	**Stasis**	**Regression** *Main Code*	**Improvement** *Main Code* *Proof Skill*
Nathan	**Improvement** *Clarity*	**Stasis**	**Improvement** *Main Code* *Clarity* *Proof Skill*
Omar	**Improvement** *Main Code* *Proof Skill*	**Improvement** *Main Code* *Proof Skill*	**Improvement** *Main Code* *Proof Skill*
Ursula	**Improvement** *Main Code*	**Improvement** *Main Code* *(Div by 3)* *Proof Skill*	**Stasis**
Zach	**Regression** *Main Code*	**Stasis**	**Stasis**

Table 4.2 - Seminar Participants' Change in Performance

(When changes occurred on the subargument level, the subargument(s) are identified.)

Most of the seminar participants changed proof methods more frequently on the post-assessment than they had on the pre-assessment. The participants who had the greatest difference in the number of changes were those who had the weakest performances on the pre-assessment, and only two participants changed proof methods less frequently on the post-assessment than on the pre-assessment (see Table 4.3).

Participant	Total Number of Switches	
	Pre-Assessment	Post-Assessment
Bill	5	4
Ingrid	0	6
Ivan	3	3
Nathan	1	2
Omar	0	1
Ursula	2	6
Zach	4	0

Table 4.3 - Seminar Participant Switching Tendency on Assessments

The participants naturally fell into two categories on the pre-assessment based on whether or not they were able to produce any valid arguments: Omar, Ursula, and Ingrid were not, but Bill, Ivan, Nathan, and Zach all produced at least one valid argument on the pre-assessment. All members of the high group except Zach had items on the post-assessment for which they changed plans less frequently but performed as well or better. However, all members of the low group changed plans at least as many times on every item on the post-assessment as they had on the pre-assessment.

Proof validation.

For this portion of the assessment, participants were given four attempted proofs of the claim "for any positive integer n, if n^2 is a multiple of 3, then n is a multiple of 3," (see Appendix 4a for the arguments). They were asked to classify each argument as a rigorous proof of the claim, a rigorous proof of a different claim, a non-rigorous proof of the claim, or as not meeting the standards of a proof. They were also to justify the classification by citing specific errors in the arguments if they found any. The participants were also given the option of saying they did not understand the argument and not providing a classification.

Classifications were considered correct if the corresponding justifications were supported by the written arguments (see Chapter 2). For example sentences [2] and [3] of, item CV, “The Converse,” establish the validity of the converse of the initial claim, if n is a multiple of 3, then n^2 is a multiple of 3, so if a participant classified item CV as a rigorous proof of a different claim and identified this converse as the claim being proved, their classification was deemed correct; however, if a participant classified item CV as a rigorous proof of a different claim and named any other claim as the one that had been established, their classification would have been incorrect because no other statement is supported by the argument.

Participants were allowed to leave items unclassified when they felt they did not understand the author’s intended argument. Some participants chose this option after identifying and discussing errors they had seen, but other participants chose it and did not discuss which errors they had seen. In the absence of justifications for some participants, this choice was coded as being neither correct nor incorrect for the sake of consistency in coding.

Only one of the arguments, item RT, “The Real Thing,” could have been classified correctly as a rigorous proof of the claim. While it could be improved by explicitly assuming that n^2 is a multiple of 3, there are no reasoning or mathematical errors in the argument. Two of the arguments, items CV, "the converse," and EG, "errors galore," contained critical flaws and did not support the claim. Item GP, "the Gap," contained no errors but did not provide enough justification at one step.

In general, the seminar participants were more likely to correctly classify the valid proof, "The Real Thing," than the other items. On the 56 items, there were 42 correct classifications, 11 incorrect classifications, and 3 unclassified items (see Table 4.4). Only one of the incorrect classifications was provided for item RT, "The Real Thing." The greatest improvement was seen on item EG, "Errors Galore." Only three of the seven participants correctly classified that item on the pre-assessment, but all seven students were able to identify it as not meeting the standards of a proof on the post-assessment.

	EG		RT		GP		CV	
	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Bill	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ingrid	No	Yes	Yes	Yes	Yes	No	No	**No^{+}**
Ivan	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nathan	Yes	Yes	No	Yes	**No^{+}**	No	Yes	Yes
Omar	No	Yes	Yes	Yes	Yes	Yes	**No^{+}**	No
Ursula	No	Yes	Yes	**No^{+}**	Yes	Yes	Yes	Yes
Zach	Yes	Yes	Yes	Yes	Yes	No	Yes	No
0296	Yes	**No^{+}**	Yes	Yes	No	Yes	Yes	Yes
4586	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
6772	No	No	No	Yes	Yes	No	Yes	Yes

Table 4.4 - Participants' Correct Classifications

("No^{+}" indicates an item given a "5" by the participant)

However, there were three incorrect classifications of item GP, "The Gap," on the post-assessment while there had been no incorrect classifications of it on the pre-assessment, and one student who had classified item CV, “The Converse,” correctly on the pre-assessment classified it incorrectly on the post-assessment.

The gap in reasoning present in the argument in item GP was the easiest error for the participants to spot. Every one of these participants identified it on at least one of the two assessments. While Selden and Selden found that most of the detected errors in these arguments were of a local/detailed nature (2003, p24), but that was not what was observed in this study. Three other errors were noticed by more than half of the participants, but only one was of a local nature: four students commented on the erroneous definitions of odd and even multiples of 3 provided by the author of item EG. The other two were of a more global/structural nature; more than half of the participants noticed the author of item EG was both assuming and concluding that n^2 was a multiple of 3. Similarly, four students noticed that in sentence [2] of item CV, the author assumed the purported conclusion. All other errors were noticed by at most two of the participants. It is important to note that in the instances in which a participant did not mention a particular error, no claims can be made about whether or not that participant noticed or was capable of noticing it.

Exit interview.

Exit interviews were conducted with the seminar participants after they had completed the survey, composition, and validation portions of the post-assessment. The interviews consisted of eleven questions that focused on the participants’ experiences in the research study (see Appendix 5). They were asked whether their confidence level regarding constructing and reading proofs, their ability to work in cooperative groups, or their appreciation of proof had changed because of their participation. They were also asked questions about facets of Cooperative Learning the researcher employed during the seminar sessions.

Confidence.

All but one of the seven participants reported increased confidence both about writing proofs and about reading and understanding them. When asked what contributed to their confidence increase about writing proofs, the participants named several specific aspects of the research such as seeing other people work on proofs, having specific roles to play during group work, and learning what different types of proof may be useful, but no more than two participants mentioned any one thing. No two participants identified the same reason for their increase in confidence about reading and understanding proofs.

Cooperative learning.

Only two participants thought that working in a group had been beneficial for their learning, and two others thought that it had actually been detrimental. The other three participants did not think that being in a group had either hampered or helped their learning, but all three expressed a pronounced dislike of group work.

Only one participant reported consistently feeling responsible for the learning of the other members of his group while two others said they felt responsible only when they were acting in the role of manager, and no one else mentioned feeling responsible for others' learning at all. However, all the participants reported feeling accountable for their own learning. The participants mostly felt that the group processing had been beneficial to the functioning of their groups, but two participants said that it was too difficult to be really honest during the exercise for it to have been very useful.

Proof.

The participants' responses to the question "What does proof mean to you" were remarkably consistent. Every participant said that proofs were logical arguments based on axioms or other previously known statements.

Comparison Participants Overview

Attitudes and beliefs.

Participant responses to the attitudes and beliefs survey can be found in Appendix 2b. As discussed in Chapter 3, the researcher could not make use of these data, and this portion of the assessment was abandoned for the implementation study.

Composition.

Of the three comparison participants, one demonstrated evidence of a tested proof skill on one item of the post assessment that had not been apparent on the pre-assessment. However, none of the comparison participants achieved a higher argument code on the post-assessment than had been achieved on the comparable pre-assessment item. All told, of the nine item comparisons, improvement was demonstrated only on one item, while there was stasis on six and regression on two (see Table 4.5).

Participant	Item C1	Item C2	Item C3
0296	**Regression** *Clarity*	**Stasis**	**Stasis**
4586	**Stasis**	**Stasis**	**Stasis**
6772	**Stasis**	**Stasis**	**Regression** *Main Code*

Table 4.5 - Comparison Participants' Change in Performance

There is no evidence that any of the comparison participants switched proof methods at any point during either of the assessments, which is due in part to the fact that the data for these participants are only written, and most of the switching evidence for the seminar participants was present in the audio data.

Proof validation (see Appendix 4a for the arguments discussed here).

The comparison participants as a group struggled to identify the valid arguments as well as invalid ones. Item CV, "The Converse," was correctly classified as either a proof of a different claim or not meeting the standards of a proof by all three participants on both assessments, and when it was classified as a proof of an alternative claim, the participants correctly stated that the argument proved the converse claim that if n is a multiple of 3, then n^2 is as well. However, each participant made at least one incorrect classification (see Table 4.6), and two of the three participants incorrectly classified an item on the post-assessment that he/she had correctly classified on the pre-assessment. On the 24 validations, there were 17 valid classifications, six invalid classifications, and one unclassified item.

	EG		RT		GP		CV	
	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Bill	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Ingrid	No	Yes	Yes	Yes	Yes	No	No	**No+**
Ivan	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Nathan	Yes	Yes	No	Yes	**No+**	No	Yes	Yes
Omar	No	Yes	Yes	Yes	Yes	Yes	**No+**	No
Ursula	No	Yes	Yes	**No+**	Yes	Yes	Yes	Yes
Zach	Yes	Yes	Yes	Yes	Yes	No	Yes	No
0296	Yes	**No+**	Yes	Yes	No	Yes	Yes	Yes
4586	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes
6772	No	No	No	Yes	Yes	No	Yes	Yes

Table 4.6 - Participants' Correct Classifications

("No+" indicates an item given a "5" by the participant)

Again in contrast to the findings of Selden and Selden (2003), the types of errors that the comparison group detected were more of a global nature. All three participants noted the missing justification in item GP, "The Gap," and all three noted the author's assumption of the conclusion in item CV on at least one assessment. However, the only other specific error that was noted by more than one participant was of a detailed nature; 0296 and 4586 both pointed out the erroneous definitions presented in item EG, "Errors Galore." None of these three participants noticed that the author of item EG was concluding as well as assuming that n^2 was a multiple of 3, and none pointed that n should have been specified as positive in item GP.

It is important to note that a participant's failure to mention a specific error is not evidence that the participant did not notice the error or did not possess the necessary skills to correctly identify it. It is possible that students provided just enough justification for their classifications without listing all errors they saw.

Individual Analysis for Seminar Participants

Bill.

Overview.

Despite his lack of experience writing proofs prior to the pre-assessment, Bill was able to do well on the composition portion, but only after completing the validation exercise. He had previously taken a course in which he was asked to read and reproduce geometric proofs, and it is possible this familiarity enabled Bill to apply what he saw in the validation arguments to his own compositions. Bill was more successful at validating than any of the other participants; he was the only participant to correctly classify and justify every argument on both assessments – also possibly related to his prior work reading and reproducing proofs. Bill improved on two of the three composition items from pre-assessment to post-assessment. In particular, he was able to use the new definition of triangular numbers to form the basis of a proof, which he had not been able to do on the pre-assessment.

Details.

Composition.

Before the pre-assessment, Bill had never been asked to produce a proof on his own. His proving experience was limited to a prior course in which he read some Euclid and Ptolemy, and was asked to reproduce proofs presented in those texts. When first presented with the composition portion of the assessment, he was completely stuck and said "I'm only used to geometric proofs … I haven't seen a proof like this before, and I'm thinking it'd be useful if I had, because generally with these kinds of things, I look for analogies with proofs I have seen to make the necessary conceptual leap as to how to even go about proving it." At that point, the researcher had Bill complete the validation portion of the assessment before coming back to the composition.

When he did return to composition, he wrote a nearly flawless contradiction proof of item C1, his only error consisting of a definition of m^2 as odd by letting $m^2 = 2x^2+1$ for some positive integer, x, instead of letting it be equal to $2x+1$. Despite this mistake, he did demonstrate both proving skills for the item. On item C2, he was able to identify the relevant subclaims but was only able to produce one of the two necessary arguments. On item C3, he demonstrated understanding of the implications of the biconditional in saying "A number is triangular if and only if, okay, so that means I have to prove two things," and he attempted some algebraic manipulation but was unable to translate the definition into anything useful for the proof. He did not get to a point in his attempts that would have allowed for the application of the results of item C1.

Every participant was given the opportunity of returning to the compositions after completing the validation portion of the assessment, but the researcher speculates that Bill's dramatic improvement in performance was likely due to his previous experience reading and reproducing proofs.

Bill did not have the best performance of the participants on the post-assessment, but he still improved upon his pre-assessment performance. On item C1, he wrote a proof by contradiction again, but he did not use his incorrect formulation of odd numbers. However, in the post-assessment, he did not explicitly state that m^2 was odd, an assumption he later contradicted, so his proof was rated A3.2.A,B as he did still demonstrate both tested proof skills. Bill's attempt at item C2 was very similar to his attempt from the pre-assessment. He identified the pertinent subclaims, but only argued that n^3-n was even, and was not able to produce any work for the subclaim that n^3-n was also divisible by 3. It was on item C3 that Bill showed the greatest improvement. On the post-assessment, Bill was able to use the definition of triangular numbers as the basis of the proof of one of the two directions (n triangular implies that $8n$+1 is a perfect square), and he was able to produce an argument for the reverse direction as well ($8n$+1 a perfect square implies that n is triangular). However, his argument for the reverse direction assumed $8n + 1 = (2k+1)^2$ without justifying why $8n$+1 was an odd number squared. Thus, he demonstrated just two of the three tested proof skills on this final item.

Bill's overall tendency to switch proving methods was lower on the post-assessment than on the pre-assessment, but the tendency was not consistent for individual arguments. He switched methods more on items C1 and C2 on the pre-assessment than the post-assessment, but while he did not switch methods at all on item C3 on the pre-assessment, he switched methods twice (both times while proving the reverse direction of the biconditional) on the post-assessment. He was better able to recognize what methods would be productive for the first two items on the post-assessment and chose those methods more quickly. However, he was so stuck on the third item on the pre-assessment that he did not make any attempt to approach the problem with another method. On the post-assessment, he was much more comfortable with the prompts and explored both direct and contrapositive proofs on the reverse direction.

Proof validation.

On the pre-assessment, Bill was able to identify algebraic as well as logical errors. On item EG, "Errors Galore," he recognized that the argument's author did not correctly set up the purported cases. "$3m$ plus one squared does not guarantee that it's an odd positive integer that's divisible by 3 … This is also false, this doesn't guarantee that it's even." He also noted that the author was claiming and concluding the same thing: "It's tautological. It's, he's assuming what he's purporting to prove." Bill did not discuss any other errors but knew that there were enough critical flaws to render the argument invalid and classified it as a 4, an argument that did not meet the standards of a proof. On item RT, "The Real Thing," he got excited that it was being set up as a contradiction but decided that the author ultimately did not prove the claim. He felt that it was a rigorous proof of the claim "if n is not a multiple of 3, neither is n^2." This is the contrapositive of the initial claim and thus logically equivalent, but Bill did not seem to know that, so he classified this argument as a 2, a rigorous proof of a different claim. On item GP, "The Gap," Bill did not notice any errors but was uncomfortable with the jump in reasoning, "So it seems like maybe this one is good, but not rigorous. I'm not quite sure how the conclusion that 3 divides into n follows immediately from the prior step," and classified the argument as a 3, a non-rigorous proof of the claim. Bill identified major issues with item CV, "The Converse," quickly recognizing that the assumption made in the first sentence (see Appendix 4a) was not connected to the second sentence, and that the second sentence was actually an assumption of the conclusion, so he rated the argument as a 4, an argument that did not meet the standards of a proof.

Bill's classifications for items EG, GP, and CV are valid and well justified. The classification for item RT is more complicated as Bill correctly identified what was being proved by the argument but did not recognize it as a valid proof of the claim. However, his analysis was essentially correct.

On the post-assessment, Bill again pointed out that the author of item EG assumed what was purportedly being proved and that the author did not correctly set up the algebra for the assumptions being made. Specifically, he noted that "an odd positive integer that is divisible by 3" is not equivalent to $(3n + 1)^2$ and again classified the argument as a 4. On item RT, he knew that the argument established the validity of the claim but said "I think if we were in the seminar being picky that maybe a couple steps needed to be more explicitly laid out." Specifically he wanted to see n explicitly written as $3k + 1$ and $3k + 2$ in the respective cases, and he wanted the assumption that n^2 was a multiple of 3 defined at the beginning. He then classified the argument as a 3, a non-rigorous proof of the claim. On item GP, Bill again found fault with the jump, but instead of deciding it was correct but non-rigorous, he classified the argument as a 4. "I'm just not seeing something. I don't know how it goes from that to that." His analysis of item CV was very similar to his analysis of that item on the pre-assessment. "Oh no. He's assuming what he's trying to prove. He said such that *n* squared is a multiple of 3 so then *n* equals $3m$. No. No no no. Then *n* squared equals $3m$." Again he classified this item as a 4.

Thus, his classifications of items EG and CV did not change from the pre-assessment and were still correct. He recognized the proof in item RT as establishing the claim but took issue with the presentation and again, his reasoning was valid. Bill's classification of GP had changed from the pre-assessment but his concerns about the reasoning gap, and thus his justification, were valid.

Exit interview.

Bill felt that his confidence about both constructing proofs and reading them had improved as a result of his participation in the research. He specifically mentioned that seeing other people write proofs made him feel better about writing his own, and that the researcher's feedback on written arguments made him more aware of "those nitpicky problems."

Bill did not think that working in a group had been either beneficial or detrimental to his learning, but he expressed a pronounced dislike of cooperative group work. "If we were to have all worked independently, and then I were to see some of those people kind of present how they did it, I don't know that it would have been any different. … I hate working in groups." He said that when a group really works together well, it can be very beneficial but that such groups are exceedingly rare.

In addition to feeling that his group had not worked well together, Bill did not find the group processing to be very helpful in dealing with their issues. He found it hard to talk about the issues they were having with his group mates: "I think perhaps part of the problem too is the tip-toeing that's inherent in a situation like that. No one really wants to say it didn't work well and here's why if the here's why involves perhaps assigning certain blame to certain people."

Ingrid.

Overview.

Ingrid did not employ any indirect proving methods on the pre-assessment and did not switch proving methods at all, but she switched methods several times on the post-assessment. Even though she was not able to produce a valid argument for one of the items even after switching methods several times on the post-assessment, she demonstrated much more flexibility than she had on the pre-assessment, and her willingness to switch items allowed her to produce a valid proof of the first claim. In her exit interview, she said she felt more confident about constructing proofs and attributed that confidence increase to knowing more about the different methods of proving and how to employ them.

The errors in the validation arguments that caught Ingrid's attention were primarily detail errors such as arithmetic problems and variable definitions; she did not pay much heed to structural issues. In fact, on both assessments she talked about statements in the presented arguments as being out of order, but in each instance the order in which she suggested putting the statements would have introduced severe issues into the logical structure of the arguments without solving any of the issues already present.

Details.

Composition.

During Ingrid's pre-assessment, she clearly attempted general arguments, but she was unable to produce any valid ones. She began her attempt at item C1 with the converse, but she recognized that it wasn't valid and discussed the importance of directionality in implications. "I can define something that's odd as being like 2*n* minus 1 … and then square it and say that's odd. No no no no, that's going the wrong way." Previously she had stated "If *m* squared is odd, then *m* is odd. It has to go that way. You have to start with *m* squared because it just, it goes one way. It's not, you can't start on the other hand, on the other side and go back because … there are arrows. They matter. The arrow with only like, the one-sided arrow means something totally different than the two-sided arrow. She then did some invalid algebraic manipulation on m^2, letting m equal $(2n-1)^{\frac{1}{2}}$, and multiplying by "a clever one," $\frac{(2n+1)^{\frac{1}{2}}}{(2n+1)^{\frac{1}{2}}}$, and simplifying incorrectly to get $\frac{4n^2-1}{(2n+1)^{\frac{1}{2}}}$. She recognized the numerator as an odd integer and concluded that the denominator must be odd as well "because you can't divide odd by even, you won't get a thing in **Z**." She goes on to conclude that *m* must be odd as well. While correctly analyzed the expression, the fact that it was based on incorrect algebra was fatal, and most of her written language was very informal. She addressed and avoided the converse argument, but she did not demonstrate any knowledge or understanding of indirect proof methods.

Ingrid was more successful on item C2 in that she didn't make any algebraic errors, was able to prove part of the claim, and exhibited the proving skill targeted by the prompt. She established the fact that if n is even, n^3-n is also even and recognized the incompleteness of her argument but was unable to make additional progress. However, she directly addressed the fact that divisibility by 6 is equivalent to divisibility by both 2 and 3.

On item C3, Ingrid knew the implications of the biconditional; she stated "There's the if and only if, so I have to go both ways." She connected the definition of triangular number with the hint, but wasn't able to use it to form the basis of a proof, and even though she also recognized $8n+1$ as always being odd, she was not able to connect that fact to item C1.

Ingrid fared better on two of the three post-assessment composition items. On item C1, she was able to abandon her direct proof and produce a valid proof by contrapositive. On item C3, she was able to connect the hint to the definition again, but this time she was able to base an argument on that definition and produce a valid proof of the forward direction of the biconditional. She was also able to produce a valid argument for the reverse direction, but like Bill, she assumed $8n+1 = (2k+1)^2$ without justifying why. Thus, she demonstrated two of the three tested skills. It is important to point out that at the end of her work, she wrote that n was a perfect square, but her work and previous comments indicated that she knew n to be a triangular number. She did not vocalize the incorrect conclusion, so the researcher determined this to be a transcription error.

Ingrid's post-assessment attempt at item C2 was no more successful than her previous attempt had been. She still was able to break the claim into the pertinent subclaims, and she attempted arguments by direct proof, induction, and proof by contrapositive. However, none of her proving attempts was fruitful. This was due, in part, to the fact that she made a logical error during her direct proof attempt claiming that she needed to show $\frac{n(n+1)(n-1)}{m} = 6$ would only be true when m and n were both natural numbers. This is a false statement not equivalent to the original claim.

Ingrid demonstrated no knowledge of alternate proving methods on the pre-assessment and did not switch methods at all on any item. However, on the post-assessment, she was able to switch to a productive indirect method on item C1, and used several method switches to explore item C2, even though those explorations did not lead to valid arguments. She did not switch methods on item C3 but was able to mostly prove the claim using a direct method without much difficulty.

Proof validation.

On item EG, "Errors Galore," on the pre-assessment, Ingrid had some issues with the factoring errors in sentence [2] (see Appendix 4a). She noted that for $n = 1$, equality doesn't hold from left to right, but she did not notice that n was being used to represent different quantities. Her concern was that for $n = 1$, $(3n + 1)^2 = 16$ but $3n(n + 2) + 1 = 10$. She also noted that for $n = 1$ the author's definition of $n^2 = (3n + 1)^2$ did not match the assumption that n^2 was odd and a multiple of 3. However, she did not have the same concern with the later assumption that n^2 was even and the corresponding definition. In fact the biggest concern Ingrid had was that the statements in the argument seemed to be in the wrong order. She thought that if the last line, [6], was moved before the previous line, then the even case would be fine, "because then you say, you have this explanation which means that n is a multiple of 3 so therefore n squared is a multiple of 3," which is the converse of the original claim. Despite it failing to be true for her tested cases, she ultimately determined that $3n + 1$ must be a definition for multiples of 3, and classified item EG as a 3, a non-rigorous proof of the claim.

Ingrid was able to read and understand the argument in item RT, "The Real Thing," much more easily. She recognized that it was organized into the two cases in which n could be written as not a multiple of 3 and that the case of n a multiple of 3 didn't need to be done "because obviously that's not part of the negation." Ultimately, she just had some presentation concerns "I just have a few linguisticky [*sic*] bits, maybe like you should mention k is an element of the natural numbers, but like, I would call that rigorouser [than item EG]." She stopped short of classifying the argument as a 1, so it too was classified as a 3.

On item GP, "The Gap," Ingrid identified the reasoning jump without trouble, "So why is it that nn equals $3x$ implies that n is divisible by 3? ... I just didn't see the explanation between these two." She was also very concerned by the fact that in sentence [1], x was not explicitly defined to be positive. She classified this argument as a 3 as well but stressed that it was less rigorous than the two she'd previously seen.

Ingrid was happier for the most part with item CV, "The Converse," than she had been with the others, but she wanted more explanation on the final line where the author claimed that m was a multiple of 3. "I also want to know why [in sentence [4]]. Like it may be really obvious, but like I'd like to have it written out." She described the whole argument this way "so assume the thing they give you, take note of any elements of sets that aren't what you're thinking of, do some algebra, make some distinctions. But those distinctions should be explained. I mean it's proof-shaped but not rigorous," and classified it as a 3 as well. While her classifications of items RT and GP were valid and well justified, her classifications of the other two items are erroneous as neither argument works towards establishing the validity of the claim.

On the post-assessment, Ingrid read through the argument in item EG without doing much work to figure out whether or not each line was valid before going back and analyzing it. She spent a little time on her first pass trying to test the author's definition of odd and a multiple of 3 but quickly moved on. When going back through, she attended to one of the primary logical flaws: "So it said assume n squared is divisible by 3, do a thing but then get that it is divisible by 3? Ok, because it looks like they're getting what they assumed." She knew that was a critical flaw and classified the argument as a 4 without going back to attend to the other errors.

As with the pre-assessment, she did not struggle with reading and understanding item RT and classified it as a 3, but on the post-assessment she was better able to articulate her concerns. "I can't tell if that's a contradiction or not. I can't tell if it's a contradiction or a contrapositive … I'm not sure if stated method [*sic*] was one used."

On item GP, she was no longer concerned about the reasoning jump but was still very concerned about the fact that x was defined to be an integer and not a positive integer. She thought the omission was so severe that it affected the validity of the argument, and she assigned it a classification of 2, a rigorous proof of a different claim, determining that it proved the claim for any integer instead of any positive integer. However, defining x as a generic integer has no bearing on the parity of n.

Ingrid was quite confused by item CV. She liked that m was specified as being positive, but she thought some of the statements were in the wrong order. She did not specify which order she would have liked to see them in. Ultimately she decided that she didn't think it was a proof but didn't understand it either and classified it as a 4 or a 5.

On this assessment, Ingrid's classifications of items EG, and RT are justified. She was not explicit enough about why she would have classified item CV as a 4 for the researcher to determine whether or not her justification was correct. A classification of 5, not understanding the argument, cannot be considered incorrect. Since she didn't identify any of the errors, this classification cannot be considered correct either. Her classification of 2 on item GP is unreasonable: the hole in reasoning needed to be addressed, and her classification was based on an error in judgment.

Exit interview.

On the whole, Ingrid thought that working in a group had been beneficial to her learning. During the exit interview, she reported feeling more confident both about constructing and about reading and understanding proofs, and she attributed the positive changes to the practice writing and reading proofs the seminar provided. She particularly liked being in a group with other people who had different mathematical backgrounds. "I've taken mostly calculus and applied based courses, so I was really unfamiliar with a lot of the notation and how to define things in a way and like keeping separate variables separate … I think I learned to do that better."

Ivan.

Overview.

While Ivan correctly justified that item CV on the validation portion of the assessments (see Appendix 4b) did not meet the standards of a proof, he did not give any indication that he recognized the converse argument that was proved in sentences [2] and [3]. Indeed, on the post-assessment, he thought that the first three lines of the argument were valid steps in support of the theorem. On both assessments, Ivan produced a proof of the converse of the first composition claim. Taken with his validation work, it seems as though Ivan did not understand the distinction between an implication and its converse.

Details.

Composition.

Ivan started off the composition section of his pre-assessment by very confidently proving the converse of item C1 without recognizing it to be the converse. His argument is well written with no unjustified assumptions and all variables defined. He struggled more with item C2 but was able to produce a valid proof. After a flawed attempt at induction, Ivan broke the initial claim into valid subclaims (namely divisibility by 2 and 3), wrote both subarguments (one directly, one by induction), and combined his results into a valid proof of the main claim. Interestingly, the researcher received an email from Ivan later in the day with an informal argument correctly laying out how one might go about proving this claim without the use of induction.

On item C3, Ivan again mistakenly believed himself to be successful in his proof attempt. Convinced by empirical exploration that if $n = \frac{k(k+1)}{2}$, then $\sqrt{8n+1} = (2k+1)$, Ivan did algebraic manipulation until he reached a tautological equation. Since he produced a true statement, he assumed the proof to be valid, although he did wonder if he was allowed to do so. “I don’t know where I made the, I don’t know if it’s appropriate to make the logical leap in the analysis, so I substituted that in, and it turned out to work.” While producing this argument, he gave no indication of understanding what was necessary to prove a biconditional claim.

Ivan was one of two seminar participants in this part of the study who regressed from pre- to post-assessment. On item C2, he again broke the claim into subclaims and produced subarguments. His proof of divisibility by 3 was informal but error-free, but in his argument for divisibility by 2, he incorrectly stated “the difference of any 2 odd numbers is odd” (see Figure 4.1), but on the next line he stated “if n is odd, then n^3-n is even,” which is the correct statement. So while his post-assessment item C2 was coded A2.1.A because of the mathematical error, and the corresponding item on the pre-assessment was coded A4.+++.A, the regression was not related to a less sophisticated proof attempt. This was a conservative coding judgment on the part of the researcher who recognizes that this could indeed be a valid proof with a transcription error; however, since Ivan wrote and verbally stated the incorrect statement, the researcher did not feel that she was

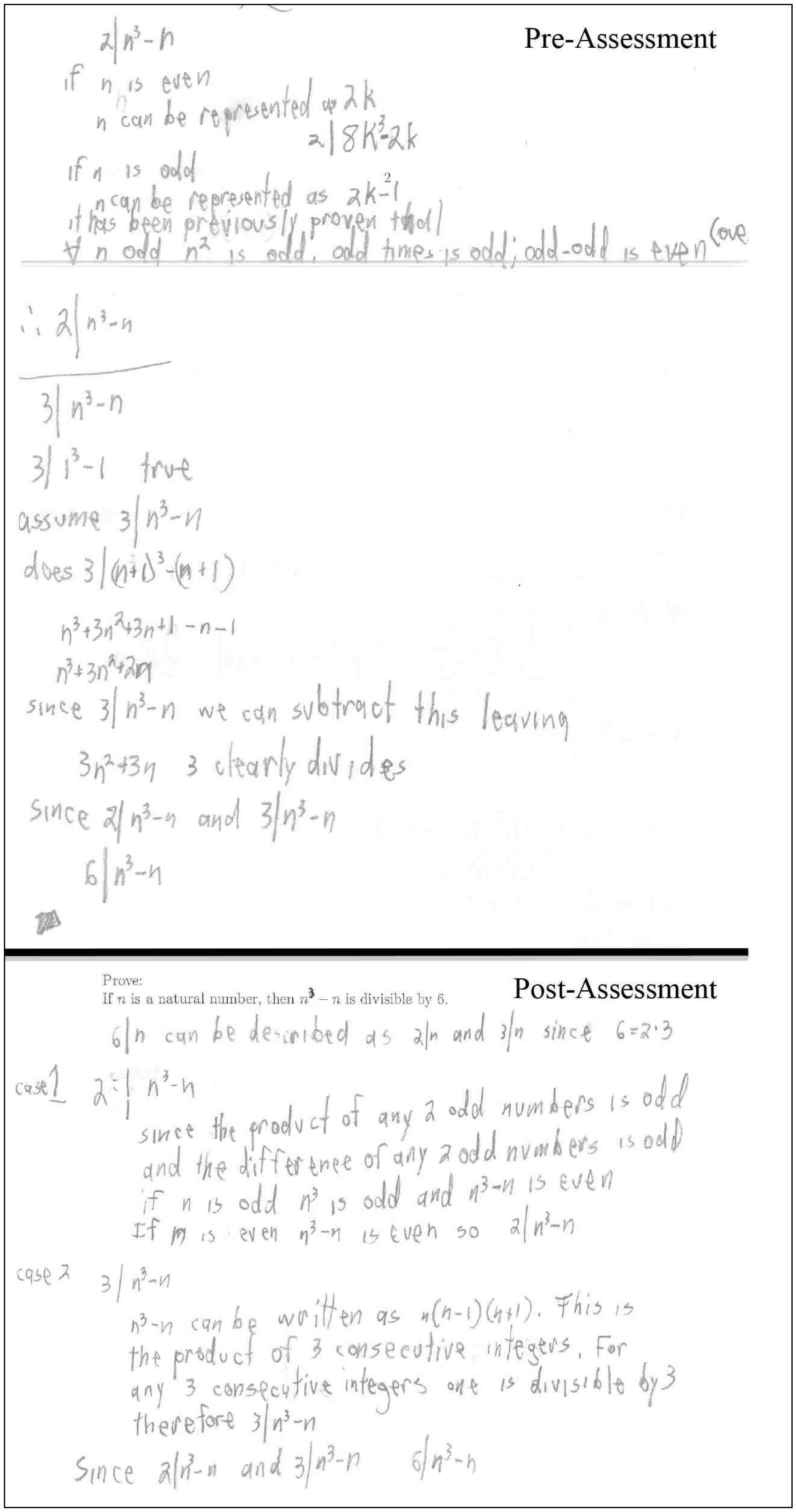


Figure 4.1 - Ivan's Pre- and Post-Assessment Work on Item C2

able to make valid assumptions about Ivan's intent. On item C1, Ivan again produced a quick, valid proof of the converse without recognizing his error. Item C3 was where Ivan demonstrated improvement. He made no unfounded generalizations or assumptions. Instead, he was able to use the definition of triangular numbers to produce a valid argument of the fact that n triangular implies $8n+1$ is a perfect square. He did some algebraic manipulation in an attempt at the reverse direction, but his explorations were not fruitful. While his work on item C3 on both assessments received a main code of A2, the absence of errors and the acknowledgement of the logical implications of the biconditional on the post-assessment clearly demonstrate improvement.

Ivan did not change methods at all on item C1 on either assessment, and he did not change proving methods on item C3 on the pre-assessment or item C2 on the post-assessment. He switched proving methods three times (direct, induction, direct, induction) on item C2 on the pre-assessment and not at all on the post-assessment. He did not switch methods for item C3 on the pre-assessment, but he did make three switches on the post assessment on that item.

Proof validation.

Ivan classified item EG, "Errors Galore," as a 2, a rigorous proof of a different claim, determining that it proved that if 3 divides n, then 3 divides n^2, but he did not give justification for why he thought it proved that. He also classified item RT, "The Real Thing," as a 2, correctly identifying that the argument proved that if 3 does not divide n^2, then 3 does not divide n either. He did not recognize that statement as the contrapositive on the initial claim and that it was thus equivalent to proving that claim. On item GP, "The Gap," Ivan was concerned about the missing justification and classified the argument as a 3, a non-rigorous proof of the claim. "I don't know, I think there should be a little more there in between nn equals $3x$ and 3 divides n… I mean it's certainly true, but … line three to four is too much of a leap."

Ivan only read the first two lines of item CV, "The Converse," before classifying the argument as a 4, not meeting the standards of a proof. He recognized that the stated assumption, "n^2 is a multiple of 3," and the second line, "then $n = 3m$ where $m \in \mathbb{Z}^+$," was faulty because the author needed that n^2 equal to $3m$. After he had decided on that classification, Ivan wrote "2nd line, does not follow, is untrue."

Ivan's classification of EG is incorrect in part because the logic flaws are so severe in that argument they prevent it from establishing the validity of any claim, and the argument definitely does not establish the claim that Ivan identified. His analysis of item RT was correct even though he did not recognize the contrapositive as being equivalent to the initial claim. On items GP and CV, Ivan's reasoning was excellent and his classifications well justified.

On the post-assessment, Ivan quickly noticed the incorrect definitions of even and odd multiples of 3 in item EG and determined that "each argument does not necessarily follow from the one before it." He did not notice the more serious error of assuming and concluding the same statement, but he still classified the argument as a 4. His analysis of item RT was very similar to his previous analysis of that argument. He approved of the logic of the argument but still rated the argument as a 2, claiming it proved that "if n is not a multiple of 3, then n^2 is not a multiple of 3," and not recognizing the equivalence of the contrapositive.

Ivan noticed the reasoning gap in item GP and was able to articulate what was problematic about it. "Well, what's missing here is that 3 is a prime number, so that … there's no way to get a product of the factors of 3 and x that will result in n, because the only factors of 3 are 1 and 3. So there's missing information. I mean it's true, but there's a fairly critical hole there." He again classified the argument as a 3.

Ivan read item CV several times and went through several classifications before deciding it was a 4. He started off thinking that it was a rigorous proof of the claim, but decided that the final statement made it non-rigorous "because the final statement, while it follows, does not actually state the truth of the matter … it definitely brings it over into non-rigorous because we're not trying to prove that m is a multiple of 3 and we don't really care." He was also concerned about m being specified as a positive integer and claimed it could have also been equal to zero. Ultimately, he decided the problems with the last line were severe enough to invalidate the entire argument.

On item EG, Ivan went from an incorrect, poorly justified classification to a correct and well justified one. His classifications of items RT, GP and CV did not change from pre-assessment to post-assessment, but only his work on items RT and GP can still be considered correct. Had the last line of the argument in item CV been a factual statement, it would not have affected the validity of the whole argument, but it is an incorrect statement, and the argument that came before it was not valid either. The "error" of defining m as positive was also not problematic. Because n was stated to be positive in the claim, m could not have been equal to zero as Ivan claimed. This was a minor issue, however. Much more concerning was the fact that Ivan accepted the first three lines of the argument as a valid proof, which they are not. His justification for the classification of this item had been correct on the pre-assessment, but it was not correct on the post-assessment.

Exit interview.

Ivan felt that his confidence about both constructing and reading proofs had increased because of the seminar, and he thought the biggest help was having to fill the different roles during the group work. He also felt that he had learned the steps of constructing a proof and how to determine what kind of proof to use.

Ivan responded positively to working in a group on the problem sets in the seminar. He felt that because the other members thought about things differently, they could all help each other get over the "brick walls" that came up while writing proofs. However, he did not feel positive about some of the Cooperative Learning elements of the seminar. For example, he said he only felt responsible for the learning of his group mates when he was acting in the role of manager, and, like Bill, he thought their group processing was not beneficial, because it was too difficult to be truly candid which is what he felt was necessary to make the exercise worthwhile.

Nathan.

Overview.

Coming into the study, Nathan had the greatest experience with proof-based courses, and performed the best on the composition portion of the pre-assessment despite being extremely uncomfortable with the assessment format. His first attempt at a pre-assessment was aborted so he could consider whether or not to continue in the study. Nathan contacted the researcher a few days later and rescheduled his assessment. Nathan was one of the few participants who did not attempt to write a proof by induction on item C2, but he also did not break the claim into subclaims, and he was unsuccessful at proving that claim on both assessments. He was, however, the only participant who successfully applied the results of item C1 to his work on item C3, which he did on the post assessment.

Details.

Composition.

At the start of his rescheduled pre-assessment, Nathan immediately began the composition portion of the assessment since he had already completed the background questionnaire and the attitudes/beliefs survey. He had also seen item C1 during the first attempt at an assessment, but he had made no progress on a written argument and the researcher did not feel that his prior exposure had an effect on his ability to prove that particular claim. During this assessment he produced a valid proof of item C1 by contrapositive after attempting a direct proof for some time. Even though his proof was valid, Nathan's argument did not meet any of the three clarity criteria. On the second item, he did unproductive algebraic manipulation and eventually abandoned the attempt. Nathan was able to prove one direction of item C3 (n triangular implies $8n+1$ is a perfect square) of the biconditional by applying the definition of triangular number and clearly demonstrated an understanding of the logical implications of the biconditional, but his algebraic manipulations of the reverse direction were unproductive. He was still quite nervous and took a break during his work on C3 to get a drink of water and calm down as well as he could. During work on all three items, Nathan had a very hard time expressing his thoughts and got increasingly nervous as he tried to do so.

Nathan was uncomfortable in the post-assessment as well, but much less so than he had been during the pre-assessment. Since having to explain his thought process had been so challenging in the pre-assessment, he asked to work silently and then explain what he had been thinking. The researcher agreed to let him do so since it was a much better arrangement for him and was unlikely to affect the quality of his proofs. On item C1, Nathan worked directly again for a while, though not as long as he had on the pre-assessment, and ultimately produced another valid proof by contrapositive which met two of the clarity criteria.

The pertinent subclaims are apparent on his post-assessment proof of item C2, but Nathan made an implicit assumption that n was divisible by 6 before beginning his algebraic manipulation, so in the end all he proved was that if n is divisible by 6, n^3-n is also divisible by 6. He did not appear to know he had made that error.

Other than the not meeting the clarity criteria, Nathan produced a flawless proof of item C3 and was the only student in the study to demonstrate the ability to apply a previous result. “Then I realized that an implication of, that this, this necessarily means that … *m* squared is odd which is what we did in the first problem that you gave me, so then I just went back, worked the algebra backwards.”

Nathan didn’t switch methods much on either assessment. On both assessments, he started every proof attempt directly. After working directly on item C1 for close to seven minutes (6:48) on the pre-assessment, he switched to proof by contrapositive and produced a valid proof very quickly (1:53). It took him less time on the post-assessment to switch away from a direct attempt. Because he did not think aloud during the post assessment, it is unknown just how much time he spent before switching, but the total time he spent on item C1 finishing with another valid contrapositive proof was less than the time he spent just working directly on the pre-assessment (6:42). On the post-assessment, he also switched proof methods one time on item C2 by reformulating the claim, but it did not help him produce a valid argument.

Proof validation.

Nathan was confused by item EG, “Errors Galore,” on the pre-assessment “for a number of reasons.” When discussing those reasons, he pointed out several of the errors present in the argument. He noted that n was being used to represent distinct quantities and that it was problematic to do so. He also pointed out that the conclusion in sentence [3] was incorrect given the work in sentence [2], and that the author reached conclusions that were already assumed. He was tempted at first to choose option 5 – I can’t classify this, because I don’t understand the argument – but ultimately he classified the argument as a 4, not meeting the standards of a proof. “It doesn’t meet the standards of proof either for those, for those same reasons, I guess. … I’m going to blame it on the proof and say 4 rather than blaming it on myself and saying 5.”

Nathan classified item RT, “The Real Thing,” as a 2, a rigorous proof of a different claim, but he decided that the argument was proving the claim “$3k + 1$ and $3k + 2$ are not multiples of 3.” He was confused by the jump in item GP, “The Gap.” “I don’t understand this jump. … since n^2equals $3x$, nn equals $3x$ implies that 3 divides n.” He spent some time trying to find a counter-example to prove that it wasn’t a valid jump but couldn’t find one and chose option 5.

Because the author of item CV, “The Converse,” assumed the conclusion in sentence [2], Nathan determined that the argument could not “count as a proof” and classified it as a 4. His classifications and justifications on items EG and CV were valid, but the alternate claim he said was proved by item GP was not established by that argument and so the classification of item GP is incorrect. His choice of option 5 on item CV cannot be considered as either correct or incorrect, but it should be noted that what he did not understand about that argument was precisely the gap in reasoning that made this argument an interesting one for participants to attempt to validate.

On the post-assessment, after looking over item EG, Nathan classified the argument as a 4 and gave his paper back to the researcher. When she reminded him to justify his classification he said “it’s meant to say that n^2 is a positive integer that’s divisible by 3, but what they wrote was that n is a positive integer that's not, that isn't a multiple of 3, and of course the problem of using the n twice like that … so that's why I said 4."

Nathan classified item RT as a 1, a rigorous proof of the claim, and didn't have any explanation because there were no errors to point out. On item GP, he initially wanted to classify the argument as a 3, a non-rigorous proof of the claim, because "it seems like there's a step missing, which I'm guessing is just the statement that n is prime, or not that n is prime, that 3 is prime." But after indicating where he would insert that statement, he decided that it wasn't critical and classified the argument as a 1. Because the author of item CV assumed the conclusion in sentence [2], he rated the argument as a 4.

Nathan's classifications and justifications of items EG, RT, and CV are valid, and while his classification of item GP is incorrect, he correctly spotted the missing justification.

Exit interview.

Nathan stated that he was more confident about the post-assessment than he had been on the pre-assessment because of his familiarity with the prompts and his comfort with the researcher, but he did not feel any more confident about constructing or reading proofs in general than he had been at the time of the pre-assessment. He also was one of two participants who felt that working in a group had a negative effect on their learning. Nathan got very frustrated working in his group and eventually gave up trying to do it well. "At the last session I think it was, it was my job to be the explainer, and Ingrid had asked for clarification, and it's like my job to be the explainer, and I was like 'Just write this down. Don't worry about it.' It's like the opposite of explaining." However, he did say that he felt that to be a result of his own social frustration and felt that having designated roles was useful.

Of all the participants, only Nathan reported regularly feeling responsible for the learning of the other members of his group, but he did not feel that responsibility at all sessions, as is evident from his interaction with Ingrid noted above.

Omar.

Overview.

Omar demonstrated the most improvement in proof composition from pre-assessment to post-assessment, but his potential for improvement was higher than his fellow participants as he had the least least experience with and exposure to proofs prior to enrolling in the study. However, he was still able to correctly classify and justify two of the validation arguments and his justifications on the post-assessment were generally more sophisticated and better communicated.

Omar was able to apply three proof skills on the post-assessment that he hadn't shown on the pre-assessment. He avoided the converse argument by writing a proof by contrapositive on the first item, and he was able to use the definition presented in the third argument to form the foundation of an argument even though he was unable to complete the argument.

Details.

Composition

Like Bill, Omar had never been asked to produce a proof of his own, but he also hadn't had the opportunity to read proofs and present them as Bill had in his prior course. Omar was enrolled in a course on integral calculus at the time of the study and had previously had a course on symbolic logic. He described his exposure to proof like this: "I've looked at [proofs], and in principle I understand the process. And I've done, we did derivations and stuff of statements within symbolic logic, so the process is familiar if the parts aren't exactly there."

Omar was not able to produce any valid arguments on the pre-assessment. He was able to give a general description of the converse argument on item C1, "I suppose my approach to this would be to cite some sort of statement about odd numbers where if you multiply an odd number by an odd number, see I don't know if that's true or not though. That's the thing. Alright, let's assume that if you multiply an odd number by an odd number then you get an odd number as a result, and if that's true then we can say that given an odd m, you're multiplying by itself, which is also an odd number, so your result would be odd. So I think that would work. I'm probably missing something. Yeah I guess that's how I'd do it."

On item C2, he explored some examples on a calculator to see if he could find a pattern but was unable to come up with a generalization other than "the way I would see this as functioning is that the n three gets you to a large enough number that 6 can go into it and then the subtraction of n provides some sort of mechanism for making it divisible. … I don't know the mechanism by which I would say that would happen." Item C3 proved even less accessible to Omar, and he was unable to fully comprehend the formulation of the claim. He did not discuss the implications of the biconditional.

The researcher was worried about how much Omar would be able to get out of the seminar with such a limited background, but he proved to be a willing and engaged participant and did progress during the course of the seminar. On item C1 on the post-assessment, Omar started by producing a rigorous proof of the converse, but he recognized his mistake very quickly. "So if m is an odd number, then m squared will be odd, which isn't necessarily what I was trying to say, is it? Well, read it backwards; you'll have what you were trying for." Then he was able to produce a valid proof by contrapositive. His proof utilizes very little notation, instead describing the process in words – likely a byproduct of his limited background.

He wasn't able to make much progress on items C2 and C3 on the post-assessment, but he was able to write down general statements and work with them and was even able to use the definition of a triangular number and the hint in item C3 to write the beginnings of an argument. However, his arguments were incomplete in both cases, and he still did not display knowledge of the structure of an "if and only if" statement.

Omar was ignorant of any indirect proving methods when he came in for the pre-assessment, and as a result he was only able to approach each claim directly. By the post-assessment, however, he had learned about some indirect methods and was able to switch from an invalid direct proof of the converse of item C1 to a valid proof by contrapositive.

He was still overwhelmed by the other two items and did not try different methods of proving on either item during either assessment.

Proof validation.

Omar wanted to classify the odd case separately from the even case as presented in item EG, "Errors Galore." He was confused about why the author let $n = 3n + 1$, so he wanted to choose option 5 – I cannot classify this because I do not understand the argument – for that case, but he thought that the even case was a 1, a rigorous proof of the partial claim. He ended up classifying the whole argument as a 3, a non-rigorous proof of the claim, because he assumed that $3n + 1$ was "some sort of thing that shows up a lot," but that he just didn't understand it. He also incorrectly stated that $3n + 1$ was always even.

On item RT, "The Real Thing," Omar talked through the whole argument describing what was going on and classified it as a 1. On item GP, "The Gap," he saw no reason why sentence [3] would imply sentence [4]:

> He's saying let n be an integer such that n^2 is equal to $3x$, ok, where x is an integer, ok. Then 3 divides into n^2. Right. Ok, fine because he's giving you x. I can live with that. Then he says since $n^2 = 3x$, then nn or n^2 is equal to $3x$. Yeah, write it any way you like. Thus 3 divides into n, and I don't feel like there is any explicit reasoning why that would be the case.

Since he doubted the validity of the implication, Omar classified the argument as a 4, not meeting the standards of a proof.

On item CV, "The Converse," Omar did not take issue with the fact that the author was using a converse argument and was happy with the argument until the final line. He was confused about why the author was making claims about m when the claim was about n, and he didn't think the claims were true. "I don't know that it explicitly states that [m] is a multiple of 3 in and of itself, so I would have to go with, I don't think it's proving a different claim necessarily, I would say 5. I don't understand his argument at all."

Omar's classification of item EG cannot be considered correct, and his reasoning about $3n + 1$ was also faulty. His classifications of items RT and GP were valid, and his choice of option 5 on item CV cannot be considered either correct or incorrect.

On the post assessment, Omar focused on sentence [2] of EG and classified the argument as a 4 since the author's written assumptions did not align with the algebraic manipulations. He said "I feel like [$3n + 1$] is just spitting out all sorts of crazy. So first line $3n + 1$ spits out even and odd values." He did not mention any other errors or concerns.

On item RT, Omar liked the fact that the author explicitly stated the forms for integers that are and are not multiples of 3 and then ran through the two cases. He again classified this argument as a 1.

The reasoning gap in item GP caught Omar's attention and he noted that while it seemed to work for this claim, he could think of other numbers it would fail for. So citing the need for more details, he classified the argument as a 3. He was happy with item CV however, and classified it as a 1 stating "it seems pretty solid. I feel like this accounts for much more than the previous one."

Omar's justifications were generally more sophisticated on the post-assessment than they had been on the pre-assessment. He correctly classified items EG, RT, and GP and justified those classifications well. However, he incorrectly classified item CV as a rigorous proof of the claim.

Exit interview.

Omar didn't think that working in a group had really affected his learning, and he was fairly dissatisfied with the experience. He thought that if any of his group mates had started at roughly the same skill level, it would have been more helpful but since the other two members of his group were more advanced, he didn't get much out of the collaboration. He said he felt like he had "spent every class running next to people" without contributing much, and he was only marginally more confident in his ability to construct proofs at the end of the study than he had been at the beginning. However, he reported that his confidence about reading and understanding proofs had increased drastically. "I would wager that my grading of those proofs is much better than it was at the start."

Ursula.

Overview.

Ursula was much more flexible on the composition portion of the post-assessment than she had been on the pre-assessment, which served her well. On the post-assessment, she was able to switch away from an unproductive direct attempt at proving the first claim to produce a rigorous proof by contradiction, and she was able to provide and recognize a valid argument for one of the subclaims in the second item, which she had not been able to do on the pre-assessment.

Details.

Composition.

Ursula began her work on the pre-assessment by writing down her givens and goals on item C1 and then attempted a proof by contradiction, thus displaying an ability to avoid the converse, as well as knowledge of indirect proof methods. However, her contradiction argument was invalid. She incorrectly formulated the contradiction of the claim, "Suppose m^2 is odd and suppose m is even for m an element of the natural numbers. Ok, for all m in the natural numbers," and she then produced a counter example that failed to establish the contradiction in general.

On item C2, Ursula initially tried to prove the claim directly, but she made a logical error when she assumed her conclusion and a mathematical error when she defined even numbers incorrectly. She did not spend long on that attempt before trying to use induction to prove the claim. She made no mistakes but got stuck during the inductive step when she simplified $(n+1)^3-(n+1)$ to $n^3-n + 3n^2+3n$. She correctly identified n^3-n as a multiple of 6 based on her inductive hypothesis, but did not see a way to work with the second part of the expression, writing in her argument "not done, but stuck." She did not attempt to split the argument into subarguments or construct subclaims.

On item C3 she was able to formulate an argument using the definition of triangular numbers, but she made a mathematical error that stymied her progress, and she abandoned her argument. She also commented on the fact that the claim was a biconditional, saying "I'm going to have to go two ways with this … because if and only if is a cute little thingy," but she did not attempt an argument for the reverse direction.

Ursula's post-assessment composition attempts were more successful. She abandoned item C1 after trying to work directly for a while and getting stuck, but after working for a while on both item C2 and item C3, she came back to item C1 and decided to write a contrapositive argument. This attempt culminated in a valid contrapositive proof of the claim.

On item C2, she worked with the contrapositive initially and then worked by induction. During that attempt, she got stuck in the same place she had gotten stuck on the pre-assessment and moved on to work on item C3 and C1 before returning. When she came back to C2, she acknowledged divisibility by 6 required divisibility by 2 and by 3 and recognized that her argument established the divisibility of n^3-n by 3. She was unable to progress beyond that point, however.

As with the other two items, Ursula worked on item C3 for a while, moved back to the other two items and then returned to work on item C3. When she looked at this item for the first time, Ursula correctly used the definition of triangular number in an attempt establish one direction of the biconditional (n triangular implies $8n+1$ a perfect square), and she acknowledged that she would need to do the other direction as well. Her attempt wasn't quite successful, however: at one point she factored out a coefficient without explaining why and ended with $8n+1$ being written as the square of a rational non-integer. She was unable to make progress on the reverse direction.

Ursula was much more flexible on the post-assessment than she had been on the pre-assessment. While she had demonstrated a familiarity with multiple proving methods during the pre-assessment by working (albeit unsuccessfully) with contradiction and induction, Ursula only switched methods twice on the pre-assessment, once on item C2 and once on item C3. On the post-assessment, Ursula switched methods once on item C1, three times on item C2, and once on item C3.

Proof Validation

On item EG, "Errors Galore," on the pre-assessment, Ursula liked the fact that it was split into even and odd cases and commented that it was set up how she would think it should be, but she didn't think the odd case was done correctly because "the math was bad." She remarked on the factoring error in sentence [2], and said "I don't see how you can show that the odd is divisible by 3 in the matter that it was done." Ultimately, she classified the argument as a 3, a non-rigorous proof of the theorem.

After looking at item RT, “The Real Thing,” Ursula decided that the cases in that argument made more sense than the cases in item EG and that the argument was presented in a better way than the other had been. She also mentioned that she tended to like proofs by contradiction and classified the argument as a 1, a rigorous proof of the claim.

As soon as she was done reading the argument in item GP, “The Gap,” Ursula said “I just don’t like this one. … How? How are you gonna go from $n^2 = 3x$, thus so $nn = 3x$ and it’s solved. It seems like it’s missing some, um the body. … the body of the argument is nonexistent. I’m reading it and I don’t think the n times n is for sure $3x$ so they haven’t convinced me.” She determined that it was not a proof because the “proof part is missing” and classified it as a 4, not meeting the standards of a proof.

Ursula did like item CV, “The Converse,” but still determined that it did not meet the standards of a proof and classified it as a 4. She was concerned about the fact that it didn’t split into cases which she thought were important and debated about whether it was a non-rigorous proof or not a proof at all. She decided “it’s not a sloppy proof, it’s a significantly missing approach” but could not elaborate further. The researcher asked Ursula about why she liked it even though she didn’t think it was a proof, and Ursula had a revelation while explaining it.

> I just, it makes sense to my head. I like the way it breaks down. It’s a nice, smooth, simple, easy multiplication stuff that I’ve been doing for a long time, versus things that you are newly introduced to. Like I feel most comfortable with basic math because that’s what I’ve been doing for decades, a decade or more. …

> Wow, this is very interesting that I'm discussing this because I don't, I didn't realize what was frustrating me about the [intro to proof] class. It's when you're introduced to new concepts and then all the sudden you're expected to feel very confident in them, and you don't. And so you're reading it, and you're like I don't know what the heck's going on very well at all, I'm treading very badly. So what I like about it is … I get caught up on what I'm comfortable with, so the part of it that's here I can read, and it's nice. And that's one of the reasons I don't like the ones that are more advanced than basic arithmetic.

Ursula did correctly identify item RT as being a proof of the claim, and her classification and justification on item 3 were valid, but her other classifications were problematic. Item EG is not a proof, and the even case on its own, which Ursula thought was correct, is deeply flawed. Item CV can be correctly classified as a 4, but Ursula did not take issue with any of the errors present and justified her classification by saying the argument was missing a significant portion of what would make it valid.

Exit interview.

Ursula's confidence about constructing proofs increased over the course of the study, but she could not name any particular aspect of participating in the research that contributed to her confidence increase: "I have no idea because, I just know that for me like the last one that we did, the last session that we did, I felt like I got it a little bit more." Even with that positive final experience, Ursula felt that working in a group had a negative effect on her learning. She told the researcher that she believed the group work could have been beneficial except her group was unsuccessful. "I think that our group was, they liked a lot of alone work, and so it was really hard because they were, because the two, Ingrid and Nathan liked to have alone time to do it and then they would present an all-but-done activity. So it, I felt like there wasn't any interaction, and so it made it difficult."

Zach.

Overview.

Zach had a more negative attitude about writing proofs than the other seminar participants, and that attitude persisted throughout the study. He mentioned during the study that he did not see the point in proving things that were already known to be true, and that he had a strong preference for applied math over pure math. He was the one seminar participant from the pilot study who exhibited significant regression on the post-assessment. The researcher attributes the regression to a lack of effort on Zach's part, as he talked about wanting to remember how the claims on the post-assessment were proved (even though he had not seen correct proofs of the claims on the pre-assessment). He also did not spend very much time on the validation portion of the assessment instead making quick, unreasoned decisions about the validity of the arguments.

Details.

Composition.

On the pre-assessment he wrote a direct proof of the converse of item C1 and talked about wanting to use the prime factorization of m in conjunction with that work, but ended up abandoning that attempt. "I said that m has a prime factorization, and so on the right hand side, there was a $2n$ but m having prime factors means that there's no way to get a two over because there are no primes that multiply together to get two except two itself. … That doesn't help." He clearly knew that the converse argument did not establish the validity of the claim, and he crossed it out. He then did a valid proof by contradiction.

On item C2, he started his work by factoring n^3-n as n times n^2-1 and exploring examples. He then concluded that this factoring would give him "an even number times a multiple of 3," but he realized that at least in some cases that this was not true. He did however notice that establishing divisibility by 2 and divisibility by 3 would be sufficient for establishing the claim. He then attempted a proof by induction, but made an error in his inductive step when he computed $(n+1)^3-n$ instead of $(n+1)^3-(n+1)$. When his computations led him to a dead end, he tried proof by contradiction and then went back to testing examples. After writing several examples as products of prime powers, he gave up and moved on to item C3.

After rewriting the claim of item C3 in summation notation, Zach assumed he could "subtract one from both sides and see what happens." He then proceeded to write $n' = \sum_{i=2}^{k} i \Leftrightarrow 8n = m^2$, which is not equivalent. At this point he abandoned the summation and tried to make sense of the hint, but he was not able to use it as a basis for any argument, and he never discussed the need to prove the two directions of the claim.

On the post-assessment, Zach again tried to prove item C1 by using the prime factorization of m and m^2, but this time he did not switch to a different proof method and only succeeded in proving that if m^2 is even, then m is as well, but he thought that he had proved that m^2 is even if and only if m is, thereby also establishing the validity of not only the stated claim, but the stronger statement that m^2 is odd if and only if m is (a true statement, but its validity was not established by Zach). This was a serious regression from his performance on the pre-assessment since he failed to prove the claim. Zach also did not demonstrate the ability to use indirect proof methods, but as he believed he had constructed a valid proof without them, he did not see a need to do so.

While working on item C2, Zach made several references to wanting to remember the proof of the claim. "Doesn't look very divisible by 6, but I'm pretty sure you showed me a proof last time I was in here that it is." "So I think it was something like n is either odd or even …" "Yeah, I'm not going to remember it." However, the researcher did not share proofs of any of the assessment items with the study participants until after all participants had completed the post-assessments. Zach did acknowledge again that establishing divisibility by 2 and by 3 would establish divisibility by 6, but after deciding he wasn't going to remember the proof, he spent no further time trying to construct an argument of his own. He had spent more than ten minutes working on this item during the pre-assessment but gave up after only five and a half minutes on the post-assessment.

His attempt at item C3 was similar. After some exploration with examples, he noticed that $8n$+1 always produced an odd number squared and wrote "$8 \triangle + 1 = (2n+1)^2$" but made no further progress and then complained that he could have done better if he'd "signed up for number theory [that] semester, or if [he'd] studied the specific examples" from the beginning of the semester. Again, he spent far less time (3:53) on item C3 during the post-assessment than he had on that item during the pre-assessment (7:26). He seemed uninterested in the task of constructing his own proofs of the claims presented.

On the pre-assessment, Zach had been willing to switch proving methods when at an impasse on both items C1 and C2. In fact, on that assessment he switched 3 times on item C2 alone. He did not switch at all on item C3 in the pre-assessment. On the post-assessment he did not switch at all, instead giving up and complaining when he got stuck.

Proof validation.

On item EG, "Errors Galore," on the pre-assessment, Zach noted many of the errors in the argument. He pointed out the incorrect definitions of odd and even, the fact that n was being used to represent multiple quantities, and that the author both assumed and concluded that n^2 was a multiple of 3. He summed it up by saying "I think this is badly written and wrong," and he classified the argument as a 4, not meeting the standards of a proof.

Zach recognized the contrapositive argument in item RT, "The Real Thing," and knew it to be equivalent to a proof of the claim. "I think that's the contrapositive. So you'd want to show a implies b. They've shown that not b implies not a. Check." He classified the argument as a 1, a rigorous proof of the claim.

On item GP, "The Gap," Zach stated that the proof was clear to him but that he knew it needed more justification and the author should have said "that it works because 3 is prime." He determined this was a 3, a non-rigorous proof of the claim. On item CV, "The Converse," he noticed that the author assumed the conclusion and worked towards the converse and classified the argument as a 4. All four classifications were valid and justified well.

On the post-assessment on item EG, Zach used the errors he spotted in just the first half of the argument to determine that the argument did not meet the standards of a proof. He criticized the author's designation of n^2 as an odd multiple of 3 implying $n^2 = (3n+1)^2$ because of the use of n to represent two distinct quantities. In his words, "they've mixed the definition of odd with the definition of divisible by 3 and just made math salad out of it … and by using the same variable they have mangled any, they've hidden anything they actually know in bad definitions."

Zach recognized that item RT was a contrapositive argument based on the opening assumption and read through the argument with that in mind. He determined that it was a valid contrapositive proof and classified it as a 1. On item GP, he discussed the reasoning gap but decided it wasn't critical and classified the argument as a 1: "I think the only thing it doesn't say is that 3 is prime, like this stuff that I had so much trouble getting on the paper about that n^2 being even or odd, this is the same thing and taking for granted that someone reading a math proof would know that. Seems reasonable enough."

Zach then spent only 30 seconds on item CV before classifying it as a 1. Zach's classifications and justifications of items EG and RT are valid, but while he recognized the prominent issue with item GP, he classified it incorrectly. It seemed to the researcher that he had lost all interest in the assessment by the time he got to item CV, which may have been why he spent so little time on it and consequently classified it incorrectly.

Exit interview.

Interestingly, Zach stated in his exit interview that he was much more confident about proof validation on the post-assessment than he had been on the pre-assessment as a result of being in the study. He also said his confidence about constructing proofs had increased. “When I need to write a proof and I can sit down and take the time to do it, um, yeah, I feel like I - that that I know a whole lot more tricks, but I guess I know that there aren’t so many.”

Zach avoided the question of how the group work had affected his learning and just stated a strong preference for being in a lecture. “I prefer to be lectured to. I’d rather have somebody at the front of the class saying these are the things to know, you’re all responsible for them, and this is how much I will slow down for anyone who has questions, and if you can’t keep up, just drop.” When asked why he preferred that method of instruction he said, “Because I can keep up.”

Individual Analysis for Comparison Participants

0296.

Overview.

0296 did no better on the composition portion of the post assessment than he had on the pre-assessment, and only slightly better on the validation portion. He correctly classified and justified one of the validation arguments that he had incorrectly evaluated on the pre-assessment, but he also did not classify an argument that he had correctly critiqued on the pre-assessment.

Details.

Composition.

The proof produced by 0296 on item C1 from the pre-assessment was a detailed proof by contradiction. Not only did this participant demonstrate both tested skills and produce a proof that met all three of the clear and convincing criteria, but he also used a great level of detail in discussing the contradiction. 0296 was unsuccessful on item C2, however. He set up a proof by induction and established the base case and inductive hypothesis, but despite doing extensive algebraic manipulation was unable to finish the proof because he got stuck at $(k+1)^3 - (k+1) = (k^3 - k) + (3k^2 + 3k)$. His work up to that point was error-free.

0296 was also unable to provide an argument for item C3 on the pre-assessment. On that item he wasn't able to produce any portion of an argument, just a list of examples of triangular numbers and the first line of the proof that $8n$+1 a perfect square implies n triangular.

0296 did not improve from pre-assessment to post-assessment; in fact, the only change from pre-assessment to post-assessment was a minor regression on item C1. He was able to again produce a valid proof of item C1, this time by contrapositive, but he did not include a conclusion statement and thus his proof did not meet the third clear and convincing criterion. He again tried a proof by induction on item C2, but when he reached the point at which he had previously gotten stuck, he attempted some algebraic manipulations that did not help further his argument. On item C3, 0296 was again unable to produce any argument. He did make the connection between the hint and the definition of triangular numbers, but he could not then use the hint to form an argument.

Proof validation.

On the pre-assessment, 0296 classified item EG, "Errors Galore," as a 4, not meeting the standards of a proof, stating "there are several false claims made, and in two occasions what is stated in words is not equivalent to what is written in numbers." He underlined sentence [3], "Therefore, n^2 is divisible by 3," and wrote "not true" next to it. He also specified that the written assumptions about n^2 being even or odd and a multiple of 3 did not match the mathematical setup on the subsequent lines.

On item RT, "The Real Thing," 0296 noted that n should be explicitly equated with $3k + 1$ and $3k + 2$ and that it was more a proof by contrapositive than by contradiction. Despite having those concerns, he still classified the argument as a 1, a rigorous proof of the claim. 0296 also classified item GP, "The Gap," as a 1 even though he recognized that more justification was warranted. He wrote "only true if p is prime, point out that p is prime."

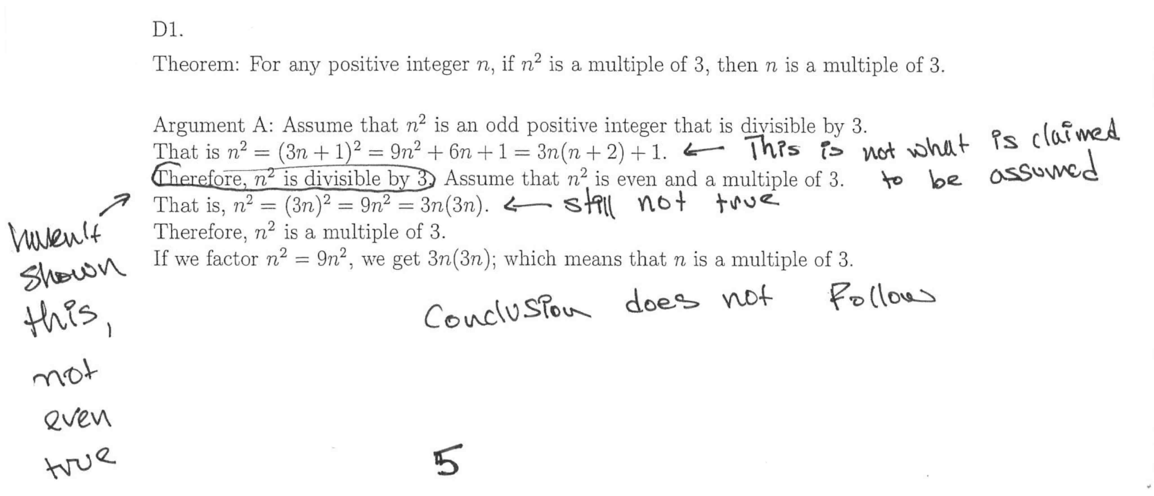
On item CV, "The Converse," 0296 recognized that the argument was faulty and classified it as a 4, providing this explanation: "This is certainly not a proof of the desired theorem. The conclusion, on top of the fact that it is not relevant, has not been shown."

Overall, 0296 did well on this portion of the assessment, but it is interesting to note that he was willing to rate arguments as rigorous proofs even when he had concerns about them. The missing justification in item GP prevents it from being able to be considered a rigorous proof so 0296's classification is invalid, but his other classifications and justifications were correct.

0296 was more confused by item EG on the post-assessment. He wrote several comments (see Figure 4.2), but ultimately decided he did not understand the argument, and chose option 5, "I cannot classify this, because I do not understand the argument." He classified item RT as a 1 and again stated that he thought it was "more of a proof by contrapositive, but he did not point out any other errors or concerns.

Even though 0296 was able to spot and fill in the reasoning gap in item GP on the pre-assessment, on the post-assessment he determined that sentence [4] did not follow and that it may not even be true. "We don't know that $\frac{x}{n} \in \mathbb{Z}$." He therefore classified this argument as a 4. He also classified item CV as a 4 pointing out that sentence [2] was "not part of the assumption" and that the concluding line was not true.

0296's classification of items RT and CV remained correct on the post-assessment, but that was not the case with item EG. On item EG, 0296's classification of 5 is neither correct nor incorrect. His classification of item GP is correct, but his justification suggests less maturity and mathematical knowledge than his pre-assessment had shown.



D1.

Theorem: For any positive integer n, if n^2 is a multiple of 3, then n is a multiple of 3.

Argument A: Assume that n^2 is an odd positive integer that is divisible by 3.
That is $n^2 = (3n+1)^2 = 9n^2 + 6n + 1 = 3n(n+2) + 1$.
Therefore, n^2 is divisible by 3. Assume that n^2 is even and a multiple of 3.
That is, $n^2 = (3n)^2 = 9n^2 = 3n(3n)$.
Therefore, n^2 is a multiple of 3.
If we factor $n^2 = 9n^2$, we get $3n(3n)$; which means that n is a multiple of 3.

Figure 4.2 - Excerpt from 0296's Post-Assessment Work on Item EG

4586.

Overview.

4586's compositions on the post-assessment were nearly identical to their pre-assessment counterparts; he drew the same unsupported conclusions and got stuck at the same points in his computations. On the validation portion of the pre-assessment, he correctly classified and justified all four arguments, but he was unable to do so on the post-assessment because he no longer recognized the argument in item RT as a valid argument.

Details.

Composition.

On the pre-assessment, this participant produced the only direct proof of item C1. While he did not demonstrate the ability to use indirect proving methods, his proof was valid and met all three clear and convincing criteria. 4586 also used induction to try to prove item C2. His argument was essentially correct, but it lacked justification and used incorrect notation, having biconditional arrows on three occasions where equal signs were warranted. The missing justification was evident at the end of the argument when he wrote "and of course, $6 \mid 3(2l + k^2 + k)$," which is not an obvious statement.

4586 produced a valid proof of one direction of item C3 on the pre-assessment, but he was unable to prove the reverse direction. He addressed the need to prove two directions, and he was able to use the definition of triangular as the basis for his proof, but he did not apply the results of item C1 to establish the second direction.

4586's performance on the post-assessment was almost identical to that on the pre-assessment. He again produced a valid, direct proof of item C1, and made essentially the same unjustified leap on item C2, this time just stating "$k^3 + 3k^2 + 2k = 3k^2 + 3k + 6l$ which is divisible by 6." On item C3, his proof of the forward direction is almost identical to that on the pre-assessment, and while he did slightly different algebraic manipulations while working on the reverse direction, he again was unable to produce an argument.

Proof validation.

On the pre-assessment, 4586 classified item EG, "Errors Galore," as a 4, not meeting the standards of a proof. He noted that having $n^2 = 3n(n + 2) + 1$ did not allow the author to conclude that n^2 was divisible by 3, and he pointed out that the proposed forms for odd and even multiples of 3 were incorrect. He classified item RT, "The Real Thing," as a 1, a rigorous proof of the claim, stating "I like this proof and can't see any flaw in it. Perhaps a direct proof would be more elegant, but as far as I can tell, this is a good, rigorous proof."

In item GP, "The Gap," 4586 saw the reasoning gap and stated that sentence [4] was not justified by the prior work because "$nn = 3x \Leftrightarrow n = 3\frac{x}{n}$ but $\frac{x}{n}$ is not necessarily an integer." He classified this argument as a 4.

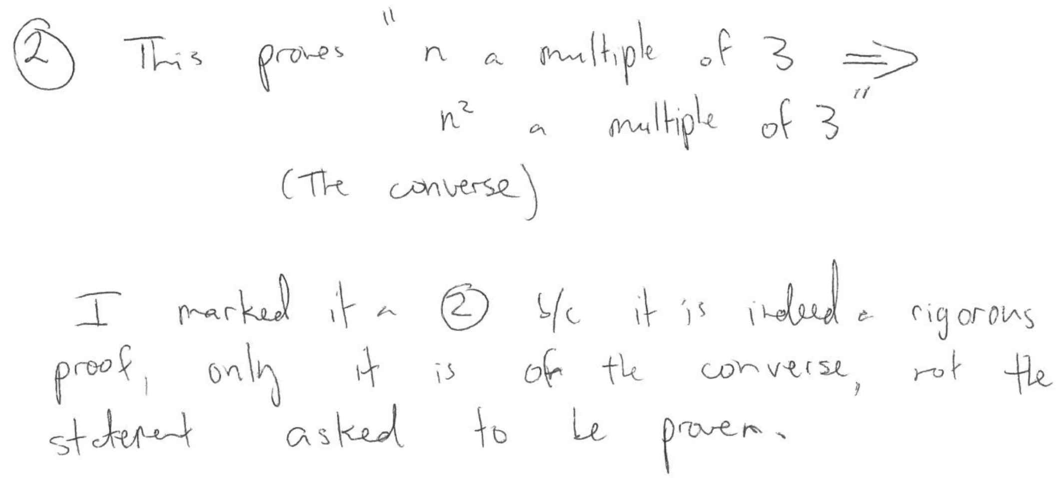


Figure 4.3 - Excerpt from 4586's Pre-Assessment Work on Item CV

On item CV, "The Converse," he noted that the argument proved the converse and decided it was a rigorous proof; therefore, he classified it as a 2, a rigorous proof of a different claim (see Figure 4.3). All four of these classifications were valid and well justified.

On the post-assessment, 4586 again noted that $n^2 = 3n(n + 2) + 1$ is not divisible by 3, and he also was concerned about using n to represent different quantities, writing that

"n^2 even and multiple of 3 $\nRightarrow$ $n^2 = 3n^2$. This is a nonsensical claim unless $n = 0$." He concluded that this argument was not at all rigorous and classified it as a 4.

4586 also classified item RT as a 4 on the post-assessment. He indicated sentence [3] and wrote "in this line you are contradicting that n^2 is a multiple of 3 for no reason. The way this proof by contradiction should go is: (i) Suppose n, not a multiple of 3. (ii) Properly assume n^2 a multiple of 3 (iii) Show that a contradiction is reached." He did not indicate any other concerns with the argument.

On item GP, 4586 was no longer concerned that the hole in reasoning wasn't true, but he did state that it needed more justification and classified the argument as a 3. Finally, he classified item CV as a 4 because the author assumed the conclusion to be true in sentence [2]. 4

586's classification of item EG remained correct, but his classification of item CV could no longer be considered valid. While the points he made would have improved the clarity of the argument, the work present establishes the validity of the claim and is not so flawed as to warrant a 4. While his classifications of items GP and CV were different than they had been on the pre-assessment, they were still valid and well reasoned.

6772.

Overview.

6772 wrote proofs of the converse of the first item on the pre-assessment and made an implicit assumption on the post-assessment that amounted to the same error, yet she was able to correctly identify the converse argument in the validation exercise as erroneous on both assessments. This simultaneous acceptance and rejection of the validity of the converse indicates that the relationship between proving and validating is complicated and that the two practices require some separate skills.

Details.

Composition.

On the pre-assessment, she wrote a detailed proof of the converse of the claim presented in item C1. She then attempted a proof by induction on item C2, but she used k to stand for two separate quantities and confused her induction hypothesis, thereby making an invalid substitution. She performed very well on the third item of the pre-assessment, however. She proved that n triangular implies $8n+1$ is a perfect square by contradiction, and she wrote a valid argument of the reverse direction. While her argument was valid, she made the assumption that $\sqrt{8n+1} = 2k+1$, citing her earlier work on the other direction. That justification was invalid here since the relationship was proved only when n was assumed to be triangular, and she provided no admissible justification, so her argument was determined to be a valid argument but not a proof.

On the post assessment, 6772 did not work explicitly with the converse of item C1, but an implicit assumption that invalidated her argument; she attempted to work directly assuming m^2 to be odd "such that $m^2 = 4(k^2 + k) + 1$," which implicitly assumes that $m = 2k+1$, and thus, she was not able to completely avoid the converse argument. Her work on item C2 on the post-assessment did not contain any errors, but she was again unable to complete the proof by induction.

6772 was the only student in the pilot study to exhibit serious regression; even though she again produced a proof of the forward direction of the claim in item C3, she was unable to provide any argument of the reverse direction.

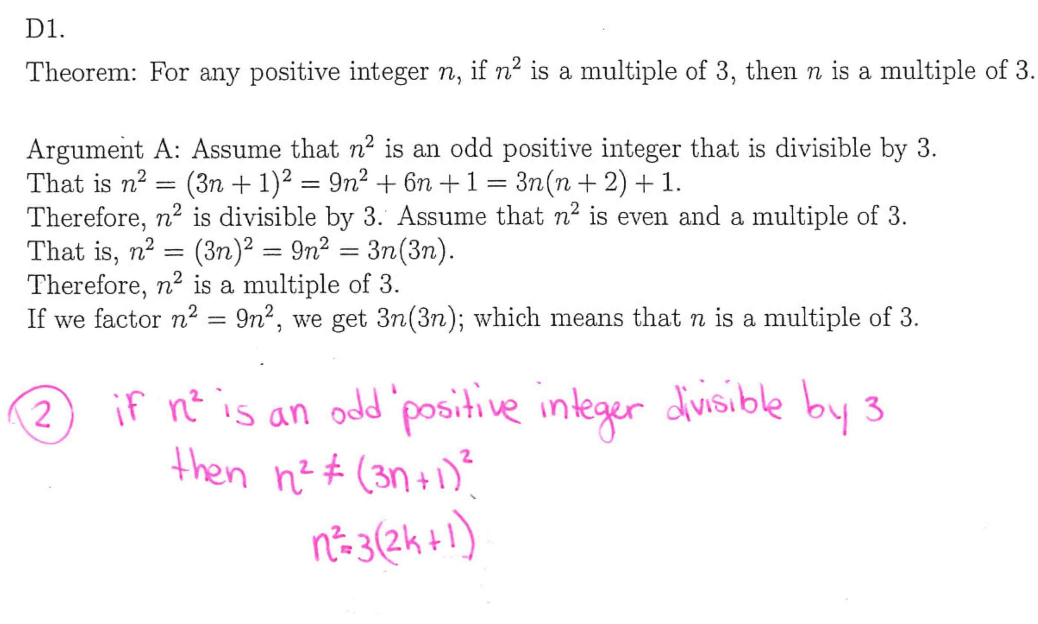
Proof validation.

On the pre-assessment, 6772 classified item EG, "Errors Galore," as a 3, a non-rigorous proof of the claim. In her justification she noted that $n^2 = 9n^2 + 6n + 1 \neq 3n(n+2) + 1$, and that even if that statement were true, it shows that n^2 is not a multiple of 3, which is contrary to what the author claimed. She also noted "the issue of n being a multiple of 3 is never addressed in the odd integer case." Since the comparison assessments were not conducted in an interview setting, the researcher did not have the opportunity to ask 6772 why she thought the argument still constituted a proof despite these errors.

On item RT, "The Real Thing," 6772 recognized the validity of the argument in establishing that if n is not a multiple of 3, then neither is n^2, and she classified the argument as a 2, a rigorous proof of a different claim. She did not recognize the equivalence of this claim with the original statement.

She classified item GP, "The Gap," as a 3 as well noting that sentence [3] did not imply sentence 4 because "$n = 3\frac{x}{n}$ however, just because both n and x are integers does not imply that $\frac{x}{n}$ is a rational." This statement is false since by definition $\frac{x}{n}$ is a rational number whenever n and x are integers, and n is non-zero, but as the rationality of n is not what is of concern, it may be reasonable to assume that she meant that one cannot conclude that $\frac{x}{n}$ is an integer.

6772 classified item CV, "The Converse," as a 2 and stated "this proof showed that if n is a multiply [*sic*] of 3, then n^2 is a multiple of 3." 6772's classifications of items RT and CV are correct and well reasoned, but her other classifications are problematic. The argument in item EG does not establish the validity of any claim and cannot be classified as any sort of proof, and 6772's justification for her classification of item GP was problematic as discussed above. However, if the assumption is made that she meant to discuss $\frac{x}{n}$ as an integer, her classification becomes valid.


D1.

Theorem: For any positive integer n, if n^2 is a multiple of 3, then n is a multiple of 3.

Argument A: Assume that n^2 is an odd positive integer that is divisible by 3.
That is $n^2 = (3n+1)^2 = 9n^2 + 6n + 1 = 3n(n+2) + 1$.
Therefore, n^2 is divisible by 3. Assume that n^2 is even and a multiple of 3.
That is, $n^2 = (3n)^2 = 9n^2 = 3n(3n)$.
Therefore, n^2 is a multiple of 3.
If we factor $n^2 = 9n^2$, we get $3n(3n)$; which means that n is a multiple of 3.

(2) if n^2 is an odd positive integer divisible by 3
then $n^2 \neq (3n+1)^2$
$n^2 = 3(2k+1)$

Figure 4.4 - 6772's Post-Assessment Work on Item EG.

6772 classified item EG as a 2 on the post-assessment. It is unclear, however, what claim she thought was being proved (see Figure 4.4). It would seem that she believes the argument is proving the claim "if n^2 is an odd positive integer divisible by 3 then $n^2 \neq (3n+1)^2$," but she may have just been pointing out an error in reasoning in the argument.

Her analysis of item RT was much clearer. She rated the argument as a 1, a rigorous proof of the claim, and had $P \rightarrow Q$ and $\neg P \rightarrow \neg Q$ written on the page indicating that she recognized the contrapositive argument and its equivalence to the original claim. She again classified item GP as a 3 but provided less justification. She indicated that x should be positive and wrote "no" at the end of sentence [4]. She also wrote $nn = 3x \nRightarrow 3|n$ but provided no other explanation.

On item CV, 6772 was conflicted and classified the argument both as a 2, stating that it proved "if n is a multiple of 3 then n is a multiple of 3 (ie 0=0)," and as a 3. She provided no explanation for her classification of the argument as a 3.

6772's classification of item EG was problematic because it was not clear but also because that argument does not establish the validity of any claim. Her analyses of items RT and GP were valid even if there was not much justification provided. Neither of her classifications for item CV can be considered correct; the alternate claim provided to justify the classification of 2 is not established by the argument, and the argument does not establish the validity of the claim, so it cannot be accurately classified as a 3 either.

The relationship between 6772's validation and composition skills merits further discussion, because even though she was able to identify a problem when validating an argument written by someone else, she was unable to avoid the same problem in her written work. On her pre-assessment, 6772 correctly validated item CV and identified that the second and third lines of the argument establish the converse of the initial claim. That is, the author of item CV showed that n^2 is odd whenever n is odd instead of proving the original claim that whenever n^2 is odd, n is as well. Even though her validation of item CV was incorrect on the post-assessment, she did correctly identify that the author had made an invalid declaration when writing $n = 3m$, because that assumption was not supported by the previous assumption that n^2 was a multiple of three. So on both assessments, she was able to spot the issues with item CV that entailed the author's assumption of the conclusion. However, on both assessments, 6772 established the converse in item C1. On the pre-assessment, she explicitly proved the converse (if m is odd, then m^2 is odd) of the presented claim (if m^2 is odd, then m is odd), and on the post-assessment, she implicitly assumed $m=2k+1$, and showed m^2 odd even though she was attempting to prove the correct claim. While 6772's pre-assessment argument for item C1 is a rigorous proof of the converse, her post-assessment argument involved only an implicit assumption; she assumed m^2 to be odd, but defined it as $4(k^2 + k) + 1$ which is not true for all odd numbers. It is possible she knew this was an invalid assumption but did not know how to start her proof and made a choice that would allow her to produce an argument.

Summary

All but one of the seminar participants improved on composition from pre-assessment to post-assessment, but their validation skills did not see a parallel improvement. None of the comparison participants demonstrated improved composition or validation skills on the post-assessment. While many of the seminar participants reported increased confidence in regards to reading and writing proofs, most also responded negatively to the cooperative work required by the seminar. Based on their reports and frustrations, some changes were made to the problem sets and the group structure for the implementation study (see Chapter 3 for details).

Chapter 5 - Implementation Study Results

Seminar Participants Overview

Attitudes and beliefs.

The researcher intended to use this interview to gauge participants' knowledge about proofs, examine what roles participants thought proofs play, and test whether or not participants were fully convinced by rigorous proofs. However, she wanted this portion of the assessment to be short so as not to overwhelm the participants before the composition and validation portions. As a result, she included a minimal number of questions and did not press participants for much explanation of their responses. The answers given touched on the intended subjects but did not provide a comprehensive view.

Knowledge.

All seven of the seminar participants recognized that empirical evidence cannot, in general, be used to prove claims, although one participant did mention that proof by examples would be fine "if you had a finite set you were talking about." Most participants also mentioned the fact that mathematical proof is based on logical rules and that once proven, mathematical theorems cannot be disproved. There was no noticeable change in the participants' answers to these questions from pre-assessment to post-assessment for any of the participants.

Conviction.

Weber (2010) found that students are not always convinced of the validity of a claim even when they have seen what they determine to be a valid proof of the claim. In order to see if the participants in this study exhibited that same lack of conviction, the participants were asked "once you have seen a rigorous proof of a theorem, how confident are you that the theorem is true?" However, since this question was asked out of the context of evaluating arguments, the researcher is not convinced the participants' answers accurately captured their conviction.

On the pre-assessment, three of the participants mentioned the need to really understand the proof in order to be convinced by it. Two of the participants mentioned needing some authority backing for the presented argument, and the remaining two expressed the need for personal experience with the concepts involved and based their conviction on that experience more than on the presented proof. On the post-assessment, two participants still said that they would be convinced of the truth of the theorem provided that they understood the proof, and two said that they need to see examples to be fully convinced that theorems are true.

Roles of proof.

In response to the question, "What is the purpose of writing proofs of theorems that are already known to be true," the seminar participants focused almost exclusively on the purpose of the act of proving and only two participants mentioned the purpose of the proofs themselves. Those two participants brought up different roles: convincing others of the truth of a claim, and demonstrating one's own understanding to others.

All of the participants discussed the purpose of proving, and two main themes were apparent. Three participants said that through proving you gain a deeper understanding of the material, and all seven participants said that you prove known results in order to practice proving.

Personal experiences.

On the pre-assessment, all but one participant expressed a preference for disproving over proving because they found disproving easier; the remaining participant stated that he had no preference. On the post-assessment, two of the participants who had previously said they preferred disproving because of its ease changed their minds. One expressed a preference for proof by contradiction, and the other stated that he no longer had a preference in general.

All participants listed aspects of proving that they liked as well as aspects of proving they did not like. Most liked the satisfaction or sense of accomplishment felt after completing a proof. There were two themes that arose from the expressions of less pleasant aspects of proof: the precision required and trouble getting started. The aspects of proof the participants disliked were also brought up as specific things they struggled with when trying to prove claims.

Composition.

The composition portion of the assessments consisted of four true, number theoretic claims that the participants were asked to prove (see Table 5.1). The pre-assessment and post-assessment items were identical. The researcher used the Argument Assessment

Assessment Item	Proof Skill(s) Tested
C1. Prove: If m^2 is odd, then m is odd.	A. Use of indirect proof methods. B. Avoidance of a more appealing but logically inequivalent converse argument.
C2. Prove: If n is a natural number, then n^3-n is divisible by 6.	A. Ability to identify pertinent subclaims and construct subarguments (divisibility by 2 and 3).
C3. A *triangular number* is defined as a natural number that can be written as the sum of consecutive integers, starting with 1. Prove: A number, n, is triangular if and only if $8n+1$ is a perfect square. (You may use the fact that $1+2+...+k=\frac{k(k+1)}{2}$.)	A. Use of the specifics of a definition to form a basis for a proof. B. Ability to identify the logical implications of "if and only if" statements. C. Use of previously established results (to prove $8n+1$ a perfect square implies that n is triangular, the result of item C1 needs to be applied).

Table 5.1 - Composition Items and Tested Proof Skills

Tool presented in chapter 3 to analyze the participant-generated arguments and compared each participants' pre-assessment composition with the corresponding post-assessment composition for each item. Improvement on each item was defined as an increase in the main argument code, fulfillment of clear and concise criteria that had been lacking, or evidence on the post-assessment of a tested proof skill that was not apparent on the pre-assessment. Any post-assessment argument with a main code of A3 was seen as an improvement over a corresponding pre-assessment argument rated A2 or A1, and a post-assessment argument rated as an A3.1.--- would have been seen as a regression from a corresponding pre-assessment argument rated as an A3.1.-++. For example, Ethan improved on item C1 because the main code for the produced arguments increased from A2 to A4, while Travis regressed on item C2 because even though the main code for both attempts was A2 (see Appendix 3d), he did not demonstrate the ability to construct subarguments on the post-assessment as he had on the pre-assessment.

As a group the seminar participants in the implementation study were weaker on the pre-assessment than their counterparts in the pilot study had been. This was due, in part, to the fact that three of them had not taken a proof class previously. Of the seven seminar participants, five of them demonstrated improvement on at least two of the three composition items, but one of those five also demonstrated some regression on one item.

One participant, Ethan, showed improvement on only one of the three items with no regression evident. The seventh participant saw neither improvement nor regression on any item, but he was the highest performer of all seminar participants on the pre-assessment and only had room for improvement on one of the three items. Of the 21 validation items from both assessments, results showed improvement on 13 items, regression on one, and stasis on eight (see Table 5.2).

All of the seminar participants switched proving methods at least as often on the post-assessment as they had on the pre-assessment; in fact Usher was the only participant who did not switch methods more frequently on the post assessment (see Table 5.3). Tammy, who showed the greatest improvement, also exhibited the greatest number of changes in proving methods.

Participant	Item C1	Item C2	Item C3
Ethan	**Improvement** *Main Code* *Proof Skill*	**Stasis**	**Stasis**
Greg	**Improvement** *Main Code* *Proof Skill*	**Stasis**	**Improvement** *Main Code* *Proof Skill* *Clarity*
Nadia	**Improvement** *Main Code* *Proof Skill*	**Improvement** *Main Code* *(Div by 2)*	**Improvement** *Main Code* *Proof Skill*
Nick	**Improvement** *Main Code*	**Stasis**	**Improvement** *Main Code* *Proof Skill*
Tammy	**Improvement** *Main Code*	**Improvement** *Main Code*	**Improvement** *Main Code* *Proof Skill*
Travis	**Improvement** *Main Code* *Proof Skill*	**Regression** *Proof Skill*	**Regression** *Main Code* *(T=>S)* **Improvement** *Main Code* *(S=>T)*
Usher	**Stasis**	**Stasis**	**Stasis**

Table 5.2 - Seminar Participants' Change in Performance

(When changes occurred on the subargument level only, the subargument(s) are identified.)

Seminar Participant	Total Number of Switches	
	Pre-Assessment	Post-Assessment
Ethan	0	4
Greg	0	5
Nadia	3	11
Nick	1	3
Tammy	1	10
Travis	1	4
Usher	5	5

Table 5.3 - Seminar Participant Switching Tendency on Assessments

Proof validation.

For this portion of the assessment, participants were given four attempted proofs of the claim “for any positive integer n, if n^2 is a multiple of 3, then n is a multiple of 3,” (see Appendix 4a for the arguments). They were asked to classify each argument as a rigorous proof of the claim, a rigorous proof of a different claim, a non-rigorous proof of the claim, or as not meeting the standards of a proof, and to justify the classification by citing specific errors in the arguments if they believed any were present. The participants were also given the option of saying they did not understand the argument and therefore could not provide a classification.

Classifications were considered correct if the corresponding justifications were supported by the written arguments (see Chapter 2). For example, sentences [2] and [3] of item CV, “The Converse,” establish the validity of the converse of the initial claim, so if a participant classified item CV as a rigorous proof of a different claim and identified this converse as the claim being proved, their classification was deemed correct; however, if a participant classified item CV as a rigorous proof of a different claim and named any other claim as the one that had been established by those two lines, their classification would have been incorrect as no other statement is supported by the argument.

Participants were allowed to leave items unclassified when they felt they did not understand the author's intended argument. Some participants chose this option after identifying and discussing errors they had seen, but other participants chose it and did not discuss which errors they had seen. In the absence of justifications for some participants, this choice was coded as being neither correct nor incorrect for the sake of consistency in coding.

Only one of the arguments, item RT, "The Real Thing," could have been classified correctly as a rigorous proof of the claim. While it could be improved by explicit assumption definition, there are no reasoning or mathematical errors in the argument. Two of the arguments, items CV, "the converse," and EG, "errors galore," contained critical flaws and did not support the claim. Item GP, "the Gap," contained no errors but did not provide enough justification to be considered a rigorous proof.

The seminar participants were much better as a group at identifying valid proofs of the claim than they were at classifying the arguments that did not establish the claim. All the participants classified item RT, "The Real Deal," correctly on both the pre- and post-assessments, and only one of the 14 classifications of item GP, "The Gap," was invalid. Overall, of the 56 items, there were 43 correct classifications, ten incorrect classifications, and three unclassified items (see Table 5.4). Item EG, "Errors Galore," was very confusing for the participants initially; all three instances in which participants were too confused by an argument to make a classification occurred on that item on the pre-assessment.

	CV		RT		EG		GP	
	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Ethan	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Greg*	Yes	Yes*	Yes	Yes*	**No**$^{+}$	Yes*	Yes	Yes*
Nadia	Yes	No	Yes	Yes	Yes	Yes	No	Yes
Nick	No	No	Yes	Yes	No	No	Yes	Yes
Tammy	Yes	Yes	Yes	Yes	**No**$^{+}$	Yes	Yes	Yes
Travis*	Yes	No*	Yes	Yes*	**No**$^{+}$	No*	Yes	No*
Usher	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1865	Yes	Yes	Yes	Yes	**No**$^{+}$	Yes	Yes	Yes
3099	No	Yes	No	Yes	Yes	Yes	**No**$^{+}$	Yes
5105	Yes	Yes	Yes	Yes	**No**$^{+}$	Yes	No	No
5635	No	No	Yes	No	No	No	Yes	Yes
6293	No	No	Yes	Yes	**No**$^{+}$	No	**No**$^{+}$	**No**$^{+}$

Table 5.4 – Participants' Correct Classifications

(* - no audio was recorded for these participants' post-assessments and no justifications were recorded, so correctness was based solely on the numerical classification provided by the participant)

("No^{+}" indicates an item given a "5" by the participant)

Even though six of the ten incorrect classifications were for item CV, "The Converse," most of the participants correctly identified the errors present in that argument. Four of the seven participants pointed out that the author had assumed the purported conclusion, and all but one of the participants noted that the conclusion of the argument, "this breaks down into $3m$ times $3m$ which shows that m is a multiple of 3," was either not supported by the argument or not related to the original claim. The participants were also very likely to note the erroneous definitions in item EG, "Errors Galore," and the missing justification in item GP, “The Gap.” In fact, all seven participants were troubled by the transition from sentence [3] to sentence [4] in item GP, where the author of the argument concludes that 3 divides n from the stated fact that 3 divides n^2 without providing justification. It is important to note that in the instances in which a participant did not mention a particular error, no claims can be made about whether or not that participant noticed or was capable of noticing it.

Exit interview.

Exit interviews were conducted with the seminar participants after they had completed the survey, composition, and validation portions of the post-assessment. The interviews consisted of nine questions that focused on the participants’ experiences in the research study (see Appendix 5). They were asked whether their confidence level regarding constructing and reading proofs, or their ability to work in cooperative groups had changed because of their participation. They were also asked questions about facets of Cooperative Learning the researcher employed during the seminar sessions.

Due to an equipment malfunction, no audio was recorded during Greg’s, Tammy’s and Travis’s exit interviews. Follow-up interviews were conducted with Tammy and Travis four months after the original post-assessments. The researcher was not able to contact Greg, so no follow-up interview occurred for that participant.

Confidence.

Four of the participants thought that their participation in the research study had led to increased confidence about writing proofs. They cited a variety of reasons for this increase, including working in groups, the roles they had to play in the groups, and the time spent on the different types of proof and proof frameworks in the seminar. The other two participants did not feel that their participation had either helped or harmed their confidence about writing proofs.

Attitudes about confidence in reading and understanding proofs was similarly split: three participants felt that their participation in the research had led to an increase in confidence regarding reading and understanding proofs, but the other participants did not agree. All three of those who did notice a confidence change attributed it in part to working in a cooperative group.

Cooperative learning.

All of the participants felt that working in a cooperative group had been beneficial to their learning, and most mentioned the benefits of different participants bringing their own ideas and perspectives to the table. They felt that the different perspectives allowed them to fill gaps in their own knowledge, learn different ways of reasoning about concepts, and learn to be more critical about what was right so that the group could move forward.

There was only one interviewed participant who did not feel responsible for the learning of the other members of his group. However, he attributed that to the fact that he was less knowledgeable than his group mates and felt he could not offer much to them. On the other hand, everyone reported feeling accountable for his/her own learning, and all of the interviewed participants found the group processing to be at least somewhat beneficial; however, several participants said that their conversations lacked depth that would have made the processing more helpful. All of the interviewed participants also expressed that they would be better able to work in cooperative groups in the future.

Comparison Participants Overview

Attitudes and beliefs.

The researcher intended to used this survey to gauge participants' knowledge about proofs, examine what roles participants thought proofs play, and test whether or not participants are fully convinced by rigorous proofs, but without the opportunity to ask the comparison participants follow-up questions since the comparison participants were not interviewed, the written surveys did not elicit as much information as the interviews with the seminar participants had.

Knowledge.

These participants all expressed the idea that mathematical proof needs to be general and not empirical, and they all professed to understand that once a mathematical theorem has been proven its validity cannot be questioned or revised. In discussing empirical evidence, all of the participants mentioned that a single counter-example is all that is needed to disprove a claim, so examples alone cannot be used to prove claims about infinite sets. There was no significant change in response from pre-assessment to post-assessment for any of the participants.

Conviction.

The comparison participants focused on how to determine they had seen a valid proof when answering the question about whether or not they were convinced by rigorous proofs. Two participants responded about needing to be able to follow the proof and needing the proof to use accessible language, and another participant discussed how her confidence is affected by the source of the proof in question.

Roles of Proof.

In their responses, these participants focused on the purpose of the act of proving rather than on the roles that proofs can play. None of the participants was put off or frustrated by the act of proving known results, and all of them saw proving as a useful exercise – though for different reasons, including preparing to become a mathematician, increasing comprehension of the mathematics involved, and pure enjoyment. There was little difference in responses from the pre-assessments and those from the post-assessments.

Personal experiences.

Three participants stated a preference for proving over disproving on both assessments, and each cited the fact that the challenge of proving is rewarding as a reason for that preference. One participant expressed no preference on the pre-assessment and a slight preference for proving on the post-assessment. The other participant misunderstood the question in a way on both assessments that suggested she did not understand what it means to disprove a claim.

Two themes arose in the participants' responses about what they liked about writing proofs: the challenge inherent in the task and the creativity required. The participants also mentioned many different struggles they have; no two participants listed the same problem. What participants reported as likes, dislikes, and struggles did not change substantially from pre-to post-assessment.

There was stark disagreement between two of the participants on the benefits of writing proofs in a mathematical context that is relatively new. One expressed this practice as one of the good things about writing proofs, "I like writing proofs because that is a great way to learn the material." However, another mentioned more than once that confusion about the mathematical context makes it "almost impossible" for her to write proofs. There is a profound difference in these two mindsets that should be explored further in other research.

Composition.

Only one of the comparison participants did as well or better on every composition item on the post-assessment as on the pre-assessment. The other four participants showed improvement on at least one item but also regressed on at least one item (see Table 5.5). Some of the written work produced by the comparison participants clearly shows that they changed proving methods, but there were likely other changes that happened that did not show up in the written work, so method changes were not recorded for this set of participants.

Participant	Item C1	Item C2	Item C3
1865	**Stasis**	**Improvement** *Main Code*	**Improvement** *Clarity*
3099	**Improvement** *Main Code*	**Stasis**	**Regression** *Proof Skill*
5105	**Stasis**	**Stasis**	**Regression** *Proof Skill* **Regression** *Main Code (T=>S)* **Improvement** *Main Code (S=>T)*
5635	**Improvement** *Main Code* *Proof Skill*	**Regression** *Main Code*	**Regression** *Main Code*
6293	**Stasis**	**Regression** *Main Code (Div by 2)*	**Improvement** *Main Code* *Proof Skill*

Table 5.5 - Comparison Participants' Change in Performance

(When changes occurred on the subargument level, the subargument(s) are identified.)

Proof validation.

This group of participants struggled more with the validation exercise than any other group of participants; they correctly classified just over half of the items. Of the 40 items, there were 22 valid classifications, 12 invalid classifications, and six unclassified items (see Table 5.6). Item RT, "The Real Deal," was the least confusing and least troublesome for these participants. Everyone understood the argument and was able to provide a justified classification, and only two of the ten classifications were invalid. Conversely, item EG, "Errors Galore," was the least accessible: there were only four valid classifications of the item, and three invalid classifications. The other three students chose option 5, "I cannot classify this because I do not understand the argument."

	CV		RT		EG		GP	
	Pre-	Post-	Pre-	Post-	Pre-	Post-	Pre-	Post-
Ethan	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Greg*	Yes	Yes*	Yes	Yes*	**No**$^+$	Yes*	Yes	Yes*
Nadia	Yes	No	Yes	Yes	Yes	Yes	No	Yes
Nick	No	No	Yes	Yes	No	No	Yes	Yes
Tammy	Yes	Yes	Yes	Yes	**No**$^+$	Yes	Yes	Yes
Travis*	Yes	No*	Yes	Yes*	**No**$^+$	No*	Yes	No*
Usher	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
1865	Yes	Yes	Yes	Yes	**No**$^+$	Yes	Yes	Yes
3099	No	Yes	No	Yes	Yes	Yes	**No**$^+$	Yes
5105	Yes	Yes	Yes	Yes	**No**$^+$	Yes	No	No
5635	No	No	Yes	No	No	No	Yes	Yes
6293	No	No	Yes	Yes	**No**$^+$	No	**No**$^+$	**No**$^+$

Table 5.6 – Participants' Correct Classifications

(* - no audio was recorded for these participants' post-assessments and no justifications were recorded, so correctness was based solely on the numerical classification provided by the participant)

("No$^+$" indicates an item given a "5" by the participant)

In general, the comparison participants were most likely to point out the assumption of the conclusion on item CV, "The Converse," and the gap in reasoning on item GP, "The Gap." All but one participant noted the former, and all participants mentioned the latter. Even though this group struggled more than any of the others on this exercise, this was the only group of participants in which a majority noticed that the author of item EG, "Errors Galore," used n to represent multiple quantities. As many participants mentioned that error as mentioned the erroneous definitions used by the author of that argument. It is important to note that in the instances in which a participant did not mention a particular error, no claims can be made about whether or not that participant noticed or was capable of noticing it.

Individual Analysis for Seminar Participants

Ethan.

Overview.

Ethan was unable to produce any valid arguments on the pre-assessment and was only able to prove one of the three claims on the post-assessment, and his validation attempts on the post-assessment were no more successful than those on the pre-assessment. However, even though he improved on the composition portion and not on the validation portion, he reported increased confidence in reading and understanding, but not in writing, proofs during the exit interview.

Details.

Attitudes and beliefs.

On the pre-assessment, Ethan qualified the level of conviction he found reading rigorous proofs with a need to apply the theorem to his previous experiences with mathematics.

"I'm convinced based on just my background and experience I've had before, not just because of that one proof, but because I've worked with the concepts that the proofs are teaching my whole mathematical life." His level of conviction was also qualified on the post-assessment: he said that he would be convinced of the truth of a proven theorem provided that he had been able to understand the proof. Such understanding would necessarily be connected to his previous background and experiences, but the conviction was not quite as limited on the post-assessment, as he seemed to express that he could be convinced of the truth of a theorem on new content rather than just content he was already familiar with. On both assessments, Ethan stated that he found disproving to be easier than proving and thus preferable.

Composition.

Ethan was one of the few seminar participants in the implementation study who had prior experience in an undergraduate proof-based class, as he had taken college geometry. He started off by attempting to prove item C1 (if m^2 is odd, then m is odd as well) using mathematical induction. However, induction is inappropriate in this situation, and he was unsuccessful at producing an argument.

Ethan struggled with item C2, (for all natural numbers, n, $n^3 - n$ is divisible by 6), as well, and it is interesting to note that while induction would have been appropriate to apply here, Ethan did not mention it or make any attempt to use it. He tried several examples and partially factored n^3-n, but he also made some statements the researcher was not able to make sense of such as "I'll just write series. I know how to write sums with epsilon, but I don't know how to write series," and "it's gonna be 0, 2, one was 1, one was 0, right? Start at 3, add 2. Let x equals 3 to infinity.” Beyond the examples and the work the researcher was unable to make sense of, Ethan was not able to make progress, and he expressed frustration with the mathematical content of the claims: “these are all very algebraic proofs, this is not really the type of proof I’ve done. I don’t know how to translate them. I know these can be translated to a geometrical figure.”

He was also unable to produce any argument for item C3, (n is a triangular number if and only if $8n + 1$ is a perfect square). He looked at specific examples but had “no clue” and abandoned the attempt. He did not address the biconditional at all when looking at this item.

One of the things that hampered Ethan in these early attempts as well as his work throughout the semester was his lack of comfort with generally accepted notation. He often preferred to make up his own notation, which made it very difficult at times for him to communicate with his group mates and with the researcher. This was a problem he was already aware of during the pre-assessment. “I’m not good with formal languages. I’ve not made it something I really care about learning until recently. I want to learn about the concept. I don’t care much how you’re supposed to write it down. … Until recently, and I guess I’m discovering that picking that way to help you write it down can help you to learn it a little bit differently.” Despite his declaration that he was discovering the usefulness of conventional notation, he struggled with using it throughout the seminar, which occasionally led to confusion among his group mates.

Ethan was more successful on the post assessment. He began by writing a proof of the converse of item C1, but he recognized that it wasn’t a valid argument because he was proving the wrong thing and then wrote a valid proof by contrapositive, although he still talked about not knowing how to write things mathematically.

He began looking at item C2 directly, but then he briefly considered a contrapositive or contradiction proof before attempting a proof by induction. He skipped the inductive hypothesis, though, and went straight to trying to prove that $(n+1)^3-(n+1)$ was divisible by 3, and he was unable to progress beyond the expansion of the expression.

On item C3, Ethan explored examples and saw that when *n* was triangular $8n+1$ always returned an odd square, and he then generalized that $8n+1$ was always odd. He was unable to then apply the results of item C1 and instead began discussing an indirect proof, which would begin with assuming $8n+1$ was not a perfect square. He talked about wanting a way to write triangular numbers in general, but he did not connect that with the hint provided and made no further progress. Again, he did not address the biconditional.

Ethan did not change proof methods on any item during the pre-assessment, but he switched at least once on each item on the post-assessment. Even though his switches did not result in valid proofs of items C2 and C3, they demonstrated greater flexibility while proving.

Proof validation.

On the pre-assessment, Ethan noticed very quickly that the author of item CV, "The Converse," started with the conclusion instead of with the hypothesis, "they automatically stated that *n* was a multiple of 3, right at the beginning, but that's what they're supposed to be proving, so that right there tells me it's illogical." Despite deciding the argument was illogical, Ethan still believed it to be a proof of something. He struggled with whether to classify the argument as a 2, a rigorous proof of a different claim, or as a 3, a non-rigorous proof of the claim. Ultimately, he couldn't name what alternate claim the argument was proving and classified it as a 3. After doing so, he asked the researcher for further clarification on the distinction between options 3 and 4. He did not change his classification as a result of the clarification.

Ethan understood the argument presented in item RT, "The Real Thing," but was concerned that the author did not actually contradict any of the explicit assumptions in the argument despite claiming to have reached a contradiction and classified the argument as a 3. On item EG, "Errors Galore," Ethan classified the argument as a 4, not meeting the standards of a proof, but he did not provide much justification other than to note that while n^2 is assumed to be a multiple of 3 the way it is defined makes it impossible to be a multiple of 3. Ethan was happy to see that sentence [1] of item GP, "The Gap," started off correctly, but he was concerned about the reasoning gap between sentence [3] and sentence [4] and unsuccessfully tried to determine what the missing justification was. "I'm trying to figure out the missing logic. I mean I know they're missing a step here between this step and this (draws a line from sentence [3] to sentence [4]), … They still have n^2. This is the same thing as this. They missed a step here. I'm not sure what they're missing, but they're missing something." He classified the argument as a 3.

Ethan's classification of item CV was incorrect; he was right to be concerned about the assumption of the conclusion, and that assumption invalidates the argument, but it cannot be classified as a 3. His classifications of the other three items were valid and adequately justified.

On the post-assessment, Ethan classified the argument in item CV as a 2, a rigorous proof of a different claim. While it can be considered a proof of the converse, that was not the claim he thought the argument proved. He again pointed out that the author had assumed the conclusion in sentence [2], but decided that the remainder of the argument was fine and that it proved the claim that if $n = 3m$, then m is a multiple of 3. The author does conclude that m is a multiple of 3, but the argument does not support that conclusion.

His analysis of item RT was very similar to what it had been on the pre-assessment; he was able to follow the argument, but he did not like that the author did not explicitly state what contradiction was reached. He also mentioned that the author should have clearly stated the cases and classified the argument as a 3. The incorrect definition of an odd multiple of 3 in item EG was enough for Ethan to classify that argument as a 4. "If n is 3, then that's not odd, and therefore that's, and not necessarily divisible by 3 either. Yeah, that's the big part. It's not necessarily divisible by 3, so it doesn't really prove anything. Because the … element they choose to use is not linked to being divisible by 3."

On item GP, Ethan was unhappy with the reasoning gap and determined that the argument was missing some steps but was set up correctly, so he classified it as a 3. As discussed above, Ethan's reasoning about item CV is incorrect; the argument does not establish the claim he said it proved. However, his classifications for items RT, EG, and GP are all valid and well justified.

Exit interview.

Ethan did not feel that his confidence regarding writing proofs increased over the course of the seminar, but he did feel more confident about reading and understanding proofs. He said that working though the strategies of direct, contradiction, and contrapositive proofs with other people helped him to stay focused and to follow the logic of proof in general. He also found working in a group to be more fun than what he'd generally experienced in previous math classes because it kept him more engaged, and he found it motivating to "watch other people struggling with the same thing that you're struggling with, and to have them, to see them overcome it."

Ethan felt strongly responsible for the learning of his group mates because he knew he needed to rely on them as well. "If they were lost, then I explained before, I'm depending on them to keep me from getting lost, so yeah, I really cared that they knew what they were talking about … It was important to me to know they were following along." Along with his group mates, Ethan found the group processing to be moderately beneficial, but he thought that it would have had a much bigger impact on their efficacy if they had been given the opportunity to review their processing conversation at the start of the following session because they had often forgotten their own suggestions and comments.

Greg.

Overview.

Due to an equipment malfunction, the only audio data recorded during Greg's post-assessment is what the LiveScribe Pen captured during the composition portion. As a result, no post-assessment data for his attitudes and beliefs or exit interview exist, and the justifications for his validation classifications are also missing as they were only communicated orally.

On the pre-assessment, Greg was one of the strongest participants on the validation portion, correctly justifying his classification on three of the four arguments. The fourth argument, item EG, "Errors Galore," he did not classify. While no justification data was recorded for the post-assessment, his classifications of all four arguments were plausible and assumed to be correct because of his strength on the pre-assessment.

Details.

Attitudes and beliefs.

On the pre-assessment, Greg expressed a preference for disproving over proving because he considered producing a counter-example to be simpler than producing a proof; however, he did not see much reason behind proving things that are already known to be true. "Well the problem is because I already know things are true … it seems like they don't need much to prove that it's true because I already know it should be by experience, I guess."

Composition.

On item C1 on the pre-assessment, Greg wrote an unnecessarily complicated argument for the converse. He had $m^2 = 4n^2 - 4n - 1$ (there is a mathematical error here that did not change the nature of the argument) but wanted it to be in the form $2n$-1, which was how he defined odd numbers. “We can define an odd number to be any number obtained by adding two other numbers and subtracting one.” It was clear from his work that he meant an odd number is formed by adding another number to itself and subtracting one. However, he wanted to use a different variable in the m^2 equation to distinguish between the *n*’s and went through a convoluted process resulting in the equation

$$m^2 = 4u^2 - 4u - 1 = 2(2u^2 - 2u) - 1 = 2n - 1$$

and the conclusion that m^2 is odd. Greg did not demonstrate the ability to use indirect proof methods or to avoid the converse argument.

He attempted to prove item C2 using contradiction, but he incorrectly set up the contradiction and set about trying to disprove the claim “If *n* is a natural number, then n^3-*n* is not divisible by 6 for all natural numbers *n*,” which he did by counterexample by setting $n = 1$. This was a logical error involving the incorrect negation of a statement with a universal quantifier.

Greg was able to use the definition of triangular numbers to form the foundation of a proof of one direction of the biconditional in item C3. He successfully proved that if *n* is triangular, then 8*n*+1 is a perfect square, but he never acknowledged the existence of or need for the second direction and did not have the opportunity to apply the results of item C1.

Greg’s post-assessment was much more successful than his pre-assessment. On item C1, he did a proof by cases, examining the consequences of *m* being odd and *m* being even and correctly concluding that m^2 can only be odd if *m* is. While this proof did not demonstrate the use of indirect proving methods, the converse argument was successfully avoided. Greg was still not able to provide a valid proof of item C2 on the post-assessment; he worked on an induction proof at first but got stuck and decided to work on item C3 before trying again.

When he returned to item C2, he worked with some examples and then did an incorrect proof by contradiction that in his words did not satisfy him. He was not successful because he incorrectly formulated the contradiction statement and negated "if it's not a natural number, then it's not divisible by 6" with a single counterexample. He did, however, improve on item C3. On the post-assessment, he was once again able to produce a valid proof of the forward direction of the biconditional, and he also acknowledged and provided an argument of the reverse direction. However, his argument assumed $8n + 1 = (2k + 1)^2$ without justification, and he often used implication arrows when equal signs were warranted.

On the pre-assessment, Greg did not switch methods at all. On the post-assessment, he was able to provide arguments for items C1 and C3 without much difficulty and did not switch methods on those items. However, as he struggled with item C2, he switched tactics on 5 occasions. Even though he didn't find his argument satisfying, Greg did think he had proved item C2, which is when he stopped switching methods. He did mention he thought it should have been proved by induction and that he just couldn't figure it out.

Proof validation.

Greg focused mainly on the final line of the argument in item CV, "The Converse," on the pre-assessment. "This is saying that m is a multiple of 3, but that's not proven here, and that's not the question either." He classified the argument as a 4, not meeting the standards of a proof because he felt that he did actually understand the argument. He did not provide much explanation, but he recognized item RT, "The Real Thing," as a rigorous proof of the claim and thus classified it as a 1.

The use of n to represent multiple quantities in item EG, "Errors Galore," was very confusing to Greg. "I don't know, it just doesn't really seem to be proving, like saying that n^2 is equal to $3n^2$. So what is n? … This doesn't say for example that n^2 is n^2. It's saying, it's just like adding a coefficient out of nowhere." This confused him enough that he chose option 5 – I cannot classify this, because I do not understand the argument.

On item GP, "The Gap," Greg classified the argument as a 4, citing the reasoning gap between sentences [3] and [4]. "So from here to there, there's like no explanation." Greg's choice on item EG cannot be considered either correct or incorrect, but his classifications of items CV, RT, and GP are all correct.

Due to an equipment malfunction, no audio was recorded during the validation portion of Greg's post-assessment. As such, his justification data are missing as he wrote his classifications on the assessment but only discussed his reasoning verbally. He classified item CV as a 3, item RT as a 1, item EG as a 4, and item GP as a 3. His classification of item CV is incorrect because that argument does not support the claim, but the others are valid even though it is possible that Greg's reasoning about one or more of those items was problematic.

Exit interview.

Because of an equipment malfunction, the exit interview with Greg was not recorded, and as the researcher was unable to schedule a follow-up interview for him, there are no exit interview data for this participant.

Nadia.

Overview.

On the post-assessment, when asked if she had a preference for proving or disproving, Nadia said she had a preference for proof by contradiction, which was reinforced by the fact that each of her proof attempts on that assessment began as proofs by contradiction, though she did switch proof methods when she hit impasses. Nadia's confidence about writing and reading proofs increased as a result of the study, and she credited the roles she had to play during the seminar for the increase.

Details.

Attitudes and beliefs.

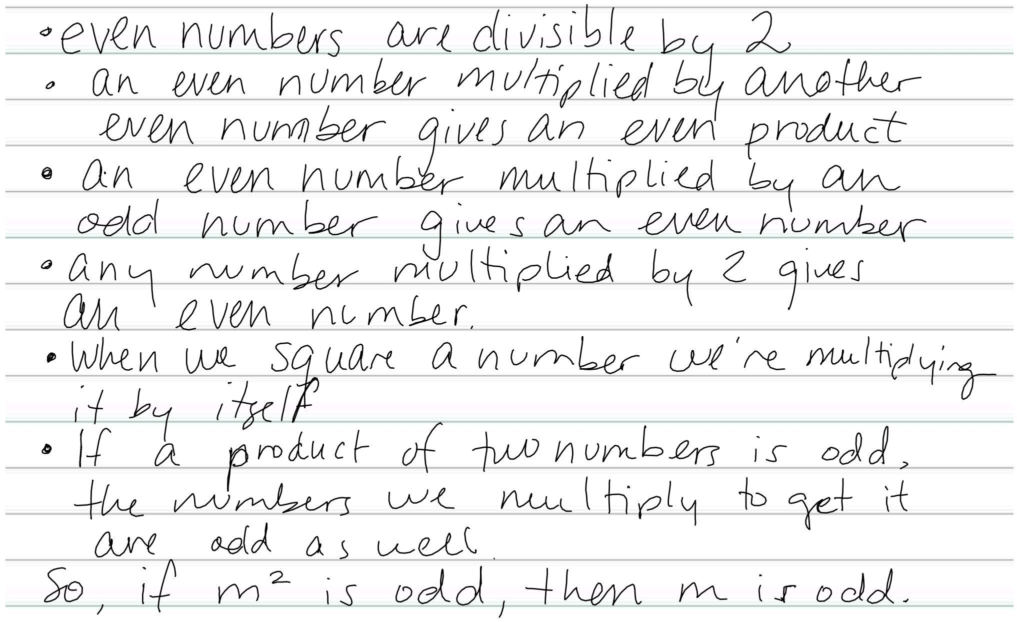
On the pre-assessment, Nadia's expressed level of conviction of the truth of a theorem she'd seen a proof of depended on whether or not she understood and could follow each step. "Sometimes it depends on what axioms they've used, because I can be like, wait, where'd you get that? But if I get all the axioms and I get all the deductive leaps, usually I'm okay. Usually that's enough for me." Also on the post-assessment, she felt that she would be usually convinced by a proof, but that on occasion seeing an example would help take her from almost convinced to completely convinced.

Like most of the other participants, on the pre-assessment Nadia expressed a preference for disproving over proving because of its ease; however, on the post-assessment she said that her preference was for proof by contradiction:

> I find it easier and because it seems like, even if there is a direct proof for something, if you're not seeing it right away and you just start by contradiction, a lot of time the contradiction proof seems to have the direct proof inside of it. You just kind of cancel out what you started with, and then you can just cross it out and be like oh yeah, there's that direct proof. But if not, then at least you're in um, if there isn't a direct proof, then at least you're not even worrying about that, you're just trying to do it by contradiction.

Composition.

On the pre-assessment on item C1, Nadia gave an informal, written description of a valid argument, but her argument lacked any symbolic expression and the explicit assumptions that would have provided the level of formality expected in a mathematical proof (see Figure 5.1). Participants in this study were advanced enough that they were allowed to assume multiplication by an even number results in an even product without justification,



- even numbers are divisible by 2
- an even number multiplied by another even number gives an even product
- an even number multiplied by an odd number gives an even number
- any number multiplied by 2 gives an even number.
- When we square a number we're multiplying it by itself
- If a product of two numbers is odd, the numbers we multiply to get it are odd as well.

So, if m^2 is odd, then m is odd.

Figure 5.1 - Nadia's Pre-Assessment Work on Item C1

so Nadia's list of statements establishes the validity of the claim, but as noted above, it is lacking formality. Her description is a description of a contrapositive argument, but without expressing explicit assumptions and following the typical contrapositive proof framework; thus, she did not demonstrate knowledge of indirect proving methods even though she was able to avoid the converse argument.

On item C2, Nadia identified the pertinent subclaims and was able to provide an informal written argument for divisibility by 2 (see Figure 5.2). She was unable to provide an argument for divisibility by 3.

On item C3 she was able to use the definition of triangular number to form the basis of an argument for one direction of the biconditional (that *n* triangular implies $8n+1$ is a perfect square), but she made a mathematical error since she dropped the +1 term when she substituted $n = \frac{k(k+1)}{2}$ into 8n+1. As a result, she computed 8n, which is not a perfect square for any triangular number. She did not notice her error and was unable to complete an argument. She also did not acknowledge the existence of or need for the reverse direction.

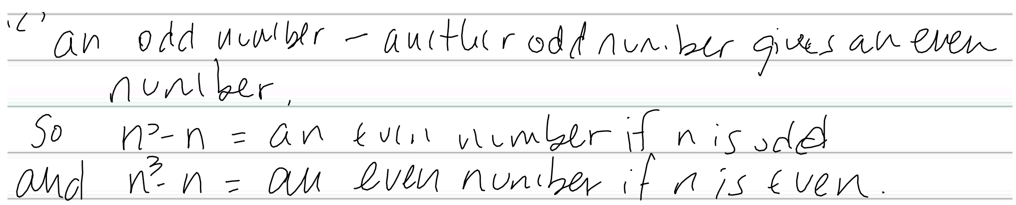


Figure 5.2 - An Excerpt from Nadia's Pre-Assessment Work on Item C2

On the post-assessment, Nadia was able to write a formal, rigorous proof of item C1. She did scratch work to start off and then produced a valid proof by contradiction. On item C2 on the post-assessment, she was still unable to provide an argument for divisibility by 3, but her argument for divisibility by 2 was formal enough to be coded as a proof even though it still did not meet all of the clear and convincing criteria.

She also managed to avoid the mathematical mistakes that had previously stymied her attempts at item C3 and produced a valid proof of the forward direction of the biconditional using the definition of triangular numbers as a foundation. While she was unable to provide an argument for the reverse direction, she did acknowledge the need for it and attempt a general proof. Her algebraic manipulations were just not very helpful to her.

Nadia switched from direct proof to contradiction to contrapositive on item C1 on the pre-assessment and went through the same switches on the post-assessment before settling back into contradiction, which ultimately resulted in a valid proof. She made one switch on item C2 on the pre-assessment when she put the proof aside to work on item C3 before returning to item C2, but on the post assessment she switched three times from direct to contradiction to putting the proof aside and back to a direct proof attempt. Finally, she did not do any switching when working with item C3 on the pre-assessment, but she changed methods and stepped away from the proof five times on the post-assessment.

Proof validation.

On the pre-assessment, Nadia focused on the converse argument embedded in item CV, "The Converse." "They said let n be a positive integer such that n^2 is a multiple of 3, but then they defined n as a multiple of 3, and then they went on to say that that multiple of 3 squared would give you a multiple of 3 instead of going the other way around." She then classified the argument as a 2, a rigorous proof of a different claim, and said that it proved that n a multiple of 3 implies that n^2 is a multiple of 3.

Nadia also saw that the argument in item RT, "The Real Thing," proved that if n is not a multiple of 3, then n^2 isn't either, but she did not seem to know that was equivalent to the original claim, and she classified the argument as a 2 as well. On item EG, "Errors Galore," Nadia pointed out the flawed definitions and also stated "you can't assume something that you're trying to prove" while indicating that the author had both assumed and concluded that n^2 was a multiple of 3. She classified the argument as a 4, not meeting the standards of a proof.

Nadia read through the argument in item GP, “The Gap,” out loud, and when she got to the end of sentence [4] said “I don’t know if that’s true. It seems to me like n would be divisible by root 3, but not necessarily by 3.” After considering whether the leap was true but unjustified or untrue, she decided that she didn’t believe the implication was true and classified the argument as a 4. Nadia’s classifications of items CV, RT, and EG were valid and well justified, but her justification for item GP was problematic since if the jump between sentences [3] and [4] in that argument were not mathematically sound, it would render the claim untrue as well, and she knew the claim to be true.

Nadia actually read the arguments out loud on the post-assessment, which gave the researcher an opportunity to look at Nadia’s validation process (see Appendix 4d for the transcript). Even though Nadia had noticed and mentioned the converse argument in item CV on the pre-assessment, she did not seem to notice it on the post-assessment. She read through the argument line by line and frequently made comments about how well she was following the argument and could see what the author was doing. She would also periodically explicitly check warrants, “Then n equals $3m$ where m belongs to our positive integers. Uh, yes. I can see that because if n is a positive integer, then m has to be a positive integer as well.” She seemed only to check for warrants when she was not completely sure of the validity of a particular part of a statement. She did not check to see if the statement $n = 3m$ was warranted, only that the definition of m was valid. The only part of the argument that concerned her was the concluding line, which she did not understand. Because she found that line confusing, she read it several times. She classified the argument as a 3, a non-rigorous proof of the claim because she could follow the logic and thought it was correct up until the very end. She did not believe that the author had actually proved that m was a multiple of 3.

Nadia spent more time reading the argument in item RT. She checked back with previous statements in the argument more often than she had on item CV. But each time she paused to ask a verification question or check a warrant, she was able to find the support until she reached the concluding statement. "Because n^2 is not a multiple of 3, we have a contradiction. So here's where it seems inadequate because n^2 is not a multiple of 3, that's not a contradiction." She determined the argument was a non-rigorous proof of the claim, a 3, because the assumption that n^2 was a multiple of 3 was not explicitly stated at the beginning of the proof.

When she started reading sentence [2] of item EG, Nadia commented that the author seemed to be doing a proof by even and odd cases. She spotted the factoring error in that sentence and kept reading. When she got to the conclusion of the first case, she expressed her concern about its validity. "Therefore n^2 is divisible by 3. Is divisible by 3, which doesn't really follow from that either because if n^2 is equal to 3 times something plus 1, then it's not divisible by 3. … We'll put a little x to say it did not follow." She paused again after reading sentence [5] to question its validity as well. She did not think the author had presented a representation of n^2 that was even and a multiple of 3. This led her to go back to the odd case and she realized that that representation wasn't necessarily odd. At this point, Nadia determined that the argument did not meet the standards of a proof, but she continued reading "for the heck of it." At that point she noticed that n^2 was defined incorrectly in terms of n and substituted in k to see what would happen. After working through the argument with k, she incorrectly stated that the author had proved that k was a multiple of 3, but it was clear to her that the divisibility of n was absent from the argument. She classified it as a 4.

On item GP, Nadia read through the argument directly until she got to sentence [4]. At that point, she discussed whether or not that line was supported by the work that had come before and whether or not that was a valid conclusion. "Not necessarily because in order to say that, you have to say n is of the form $3x$ over n. Well, I guess so because $\frac{x}{n}$ then is just some number, and 3, do we know that's an integer though? … Well x is an integer, and n is an integer, but we don't know if $\frac{x}{n}$ is an integer." She decided to classify the argument as a 3, even though she was not entirely convinced that implication was correct.

Nadia's analyses of items RT, EG, and GP were correct and well reasoned, but her classification of item CV was invalid.

Exit interview.

Nadia's confidence regarding writing proofs increased over the course of the study, and she felt that spending time on the different types of proof and proof frameworks in the seminar was one of the primary reasons for that increase. She also thought that having to be the recorder was very helpful because she couldn't just tag along with what her group mates were doing. Nadia also felt more confident reading and understanding other students' proofs, but she would have liked to spend more time working on that in the seminar. "I wish there had been more times when … we'd had more time to look at other peoples [proofs] … When you're so involved in your own, you're like okay. I know this is rigorous, but then having to go through somebody else's process I think is really helpful." Validation activities were rare in the seminar, and Nadia thought that playing the role of the skeptic had the biggest impact on her validating skills.

> If you're writing something that the skeptic or just whoever feels like being skeptical today is like I don't think you can do that because of this, and then that, you also have that perspective on what you're doing. And then I think when you're doing that same thing to other people's when you're thinking critical the whole time, then it becomes easier to look at a proof that's already written and do that process.

Nick.

Overview.

Nick had not taken any undergraduate proof-based math courses prior to enrolling in the study. On the pre-assessment, Nick was unable to produce any argument for any of the claims on the composition portion, and he thought that all four validation arguments were valid and established the truth of the claim. His validation skills on the post-assessment were not much better, and while he was able to make some progress on arguments for two of the claims on the composition portion of the assessment, he wasn't able to complete either.

Details.

Attitudes and beliefs.

During the attitudes and beliefs survey interview on the pre-assessment, Nick indicated that he held an authoritarian proof scheme (Harel & Sowder, 1998): “If it’s been proved by other people, I can take that as true, like in class he tells us, this is a very famous proof that has been done several times, and I’m like, ok, so this is definitely true.” On the post-assessment, he had moved, at least in part, away from his authoritarian proof scheme which was illustrated by the fact that he expressed he would be convinced of the truth of the theorem provided that he understood the proof.

Composition.

Nick was unable to provide any argument on any of the items on the pre-assessment. While perhaps unsurprising, as he had not previously taken any proof-based courses, other participants who had similar backgrounds were in fact able to provide some sort of argument attempts. When he was unable to get started on an argument, the researcher encouraged him to do the validation portion of the assessment before returning to the composition portion. This had been done with one of the pilot participants as well, and all participants were allowed to return to their composition work after completing the validations. Upon returning to his compositions, Nick talked through a single example illustrating the claim in item C1 but did not write anything down. His performance on the other two items was similar; he told the researcher that he was “going over a couple examples in [his] head,” and he did not write anything down on either item.

Nick did not fare much better on the post-assessment, but he did show some improvement. He did some work towards proving the converse of item C1, but he recognized his error, saying “I think I’m trying to prove if m is $2k+1$, then m squared is $2k+1$, which is not the question … Yeah, that doesn’t make any sense.” He also talked about doing a proof by contrapositive: “So we can prove using the contrapositive that if m is not odd, then m squared can’t be odd. I think,” but he did not write any part of that argument down.

Nick thought about item C2 for a while, but the two notes he wrote on the page did not begin to constitute an argument for the claim. On item C3 Nick attempted general arguments of both directions and was correctly able to state the contrapositives of both implications, but his arguments were far from complete. He was able to use the hint to write a formula for triangular numbers, but he was not able to use it to start a proof.

The researcher counted Nick’s switch from composition to validation and back as one change in proof method, but other than that, Nick did no switching on the pre-assessment. He still did not switch much on the post-assessment, but he did switch methods once on item C1 and twice on item C3. He did not switch at all on item C2 on either assessment.

Proof validation.

On the pre-assessment, Nick classified item CV, “The Converse,” as a 1, a rigorous proof of the claim, saying “I don’t see what else he could’ve, or they could’ve done.” Nick looked at item RT, “The Real Thing,” twice. On his first attempt he remarked on the fact that the argument was a purported proof by contradiction but got bogged down in the computations and had trouble following the argument. Nick stated that he thought the proof was correct but didn't understand how the author did it, and so he chose option 5 – I cannot classify this, because I do not understand the argument – but reading the argument in item EG, "Errors Galore" helped Nick understand what the author was trying to do in item RT, and so he returned to look at it again. On his second attempt, he understood it a lot better but still didn't think it was very clear; however, because he understood it better, he classified it as a 1.

Nick found item EG very clear. He liked how the computations were carried out, and he found the even case especially clear. He classified the argument as a 1. On item GP, "The Gap," Nick didn't see any errors, but it felt to him like something was missing and that the author had made large reasoning jumps, so he classified the argument as a 3, a non-rigorous proof of the claim. Nick's classifications of items RT and GP were valid, but as the arguments in CV and EG do not establish the validity of the claims, his classifications were incorrect.

On the post-assessment, Nick's classification of item CV did not change. He classified the argument as a 1 and said "it seems they did a really good job with the arithmetic. Makes sense … except for the last sentence." When asked what the issue was with the last sentence, Nick was unable to elaborate and reiterated that he thought it was rigorous.

He decided that the author of the argument in item RT had proved the contrapositive and had not produced a proof by contradiction, so he classified it as a 2, a rigorous proof of a different claim, stating that "they assumed that n is not a multiple of 3, and they proved that n^2 is not a multiple of 3." This is an accurate statement of the contrapositive; it isn't clear whether Nick knew the contrapositive to be equivalent to the claim and rated the argument as a 2 because it wasn't a contradiction proof as stated in the argument, or if he did not know that the claim proved was equivalent to the original statement.

On item EG, Nick stopped as soon as he found the erroneous designation of $n^2 = (3n+1)^2$ as an odd multiple of 3. He called it a logical fallacy and decided that it invalidated the entire argument. He then classified the argument as a 4, not meeting the standards of a proof. He once again thought the author of item GP had omitted necessary justification between sentences [3] and [4] and classified the argument as a 3. "I think I'm going to say this is a non-rigorous proof of the claim because I think they could have gone into more detail in this area."

While Nick's post-assessment classification of item CV was still incorrect, his justification for his valid classification of item RT was stronger than it had been on the pre-assessment, and while it is true that the argument in item EG does not meet the standards of a proof, the single erroneous definition is insufficient evidence, so his justification isn't strong enough to consider this classification correct. However, his classification and justification for item GP were still valid.

Exit interview.

Nick's confidence about constructing proofs increased during the semester, and he attributed that increase to having been a member of a cooperative group during the seminar, and he thought that the group work was generally beneficial to his learning. "I think it helps a lot because if, when you know you understand something one way, someone may understand it a different way, and once you get those different understandings, it can help you gain a greater understanding of the concept." However, he did not feel that his validation skills had improved at all.

Nick was the only interviewed participant who did not feel responsible for the learning of the other members of his group. However, he attributed that to the fact that he was less knowledgeable than his group mates and felt he could not offer much to them.

Tammy.

Overview.

Tammy was a first-year graduate student enrolled in a Master's program in applied math but needed to take a few undergraduate prerequisite classes and was enrolled in her first proof-based mathematics courses since high school at the time of the study. Even with her limited background, her validations were among the best of the implementation study participants on both assessments. She was unable to provide any arguments for the composition items on the pre-assessment, but she improved greatly during the study, and she was able to write proofs for item C1 and the forward direction of item C3 on the post-assessment.

Details.

Attitudes and beliefs.

Tammy was not, in general, convinced by proofs on their own. On the pre-assessment she said that she'd be convinced by a proof in a textbook because of its source, and she also mentioned that she always wanted to see concrete examples. “I assume that if it’s in the book, it must be true, so I can understand it, but I’d say it depends on how complicated it is. A really simple one, I’d say yeah, I buy that. A really complicated one, I’d say I think I need more understanding before I can really buy that, but it’s in the book, so I can trust it.” By her re-interview Tammy had abandoned her reliance on the text, but still needed to see examples in order to be convinced of the truth of a proven claim.

Like most of the other seminar participants in the implementation study, on the pre-assessment Tammy expressed a preference for disproving over proving because she found it easier. On the post-assessment, she said that she still preferred disproving but found it less satisfying than producing a proof:

> If you can come up with a counterexample, that’s a whole lot easier, I mean it’s very concise. Look, it is, it doesn’t work, there you go. So I guess if you can come up with that, that would be my preference to, to um, just because it doesn’t take a lot of explaining away, it just is in and of itself, but I guess that’s less satisfying than proving something is so … I guess to me a truth is more compelling than a non truth.

Tammy’s post-assessment responses were recorded several months after the conclusion of the study because of an equipment malfunction which occurred during the original post-assessment, but during her re-interview, Tammy stated that she did not believe any of her expressed attitudes would have been different at the time of the initial assessment than they were at that time.

Composition.

Tammy was the only non-undergraduate participant. She was just beginning her graduate studies in applied mathematics, but she had been admitted "deficient" and was enrolled in undergraduate courses as a result. She had majored in Chemistry as an undergraduate and had not previously taken any proof-based courses.

On the pre-assessment, she was not able to provide any arguments. She did some algebraic manipulation on item C1, but she was working with the converse, which she recognized as invalid. "I've taken my conclusion and shown that my hypothesis is true." She was not able to do any work in the correct direction. So while she was able to avoid the converse argument, she was not able to use alternative proof methods such as proof by contradiction or contrapositive. On item C2 she did not get beyond recognizing that natural numbers are positive; again, she provided no argument, and on item C3, she acknowledged the biconditional saying "I know if and only if you have to prove the forward and the backward," and she talked about it being a candidate for induction but did not start an argument.

On the post-assessment, Tammy showed improvement on all three items. She did a valid, direct proof by cases of item C1, but she still did not demonstrate knowledge of alternate proof methods on that item. Tammy was very committed to the assessment and worked for a long time on items C2 and C3, going back and forth between the two and employing several different proving methods. Despite working with examples and attempting direct, contradiction, and contrapositive proofs of item C2, she was not able to produce a complete argument. She was, however, able to produce a valid proof of one direction for item C3; she used the definition of triangular numbers as a basis for her proof of n triangular implies 8n+1 a perfect square. She did some algebraic manipulations for the reverse direction that were error-free but she was still not able to use that work to produce a proof.

Proof validation.

On item CV, "The Converse," on the pre-assessment, Tammy expressed concern that the author assumed n^2 to be a multiple of 3 in sentence [1] but wrote $n = 3n$ in sentence [2]. She debated about whether or not the argument was actually a proof of a different claim because the author's final conclusion was about m being a multiple of 3 but decided that wasn't the case. She asked for clarification on what the standards of a proof were and classified the argument as a 4, not meeting the standards of a proof, because she didn't "think step number two [came] from step number one."

Tammy could clearly see what the author of item RT, "The Real Thing," was intending to do, but she had a problem with the fact that the contradiction was with an unstated assumption. She determined that if the author had started the argument with the assumptions "n^2 is a multiple of 3 and n is not a multiple of 3," the proof would have been rigorous, but as written it could not be considered rigorous. Thus, she classified the argument as a 3, a non-rigorous proof of the claim.

Tammy chose option 5 – I cannot classify this, because I do not understand the argument – on item EG, "Errors Galore." She was confused by the erroneous designation of

$n^2 = (3n+1)^2$ as an odd multiple of 3 and by what she saw as a simultaneous assumptions that n^2 was even and odd.

Tammy expressed two primary concerns with the argument in item GP, "The Gap." She did not like that x was defined as an arbitrary integer and not specified as a positive integer, and she did not like the gap in reasoning between sentences [3] and [4]. Ultimately she classified the argument as a 3, but she also specified that it was only a 3 provided that the leap from sentence [3] to sentence [4] was valid. Tammy's classifications and justifications for items CV, RT, and GP were valid, and her choice of option 5 on item EG could not be considered either correct or incorrect.

On the post-assessment, Tammy noted that the author of the argument in item CV assumed the conclusion in sentence [2]. She saw this as a sufficiently critical flaw to invalidate the argument and classified it as a 4. On item RT, Tammy could not decide whether she thought the argument was rigorous or not, but she did not doubt that it was a proof of the claim. She recognized that the author was implicitly assuming that n^2 was a multiple of 3, but Tammy wanted that assumption to be made explicit. Since she had no other qualms about the argument, her indecision was based on that unstated assumption. She classified the argument as either a 1 or a 3, which the researcher allowed.

Tammy was very confused by item EG, but she could not avoid classifying the argument because she felt she knew what the author was attempting to do. Ultimately, she determined that the argument had a lot of problems, mostly centering on the erroneous definitions of even and odd multiples of 3.

> They were going direct proof with assuming that that was multiple of 3, and so they assumed it was either odd and multiple of 3 or even and a multiple of 3, but they didn't, but that isn't necessarily even because if n was 1, and so sort of they were on a track of saying well assume this and then show that it's a multiple of 3, but then they, their assumptions, they didn't actually do [that] … but I think they were trying to get there.

She thus classified the argument as a 4.

On item GP, Tammy pointed out that in sentence [1] both n and x should be designated as positive integers, not just arbitrary integers, and she was unhappy with the leap from sentence [3] to sentence [4]. She decided that the leap was incorrect and constituted a logical error. Therefore, she classified the argument as a 4. All four of Tammy's post-assessment classifications and justifications were valid.

Exit re-interview.

Tammy felt that participating in the study had led to an increase in her confidence regarding writing proofs, and she gave credit for that increase to spending time on the different types of proof and proof frameworks in the seminar. However, she did not feel any more confident about reading and understanding proofs than she had at the beginning of the study.

She did say that the group work had been beneficial to her learning but made a point of saying that she would not generally find group work to be so helpful. She was very aware of the fact that the seminar was for research purposes and made a more concerted effort to keep the whole group on track than she felt she would have in a traditional class setting.

> Because it was for your research study, I really tried to one, make sure that other people were understanding and two, make sure that I understood as opposed to maybe in a class structure where I would have grinned and nodded and moved along for the sake of time or the sake of whatever, and so, in that case I think [working as a member of a cooperative group] did increase my learning and was maybe the combination of the group and the setting. You know those things like if you're being watched you behave differently, and I know that in a class setting, I know I would have behaved differently.

However, she did say that her teaching background had contributed to her sense of responsibility for the learning of her group mates and that she was consistently going back to make sure her fellow members were understanding the work that was being done.

Travis.

Overview.

Travis was the least successful student in the implementation study in terms of change from pre-assessment to post-assessment; Travis's composition performance on the post-assessment showed both improvement and regression over that of the pre-assessment. While he was able to construct a proof for item C1, which he had not been able to do on his first attempt, on item C3 he failed to construct an argument for the direction that he had previously been able to prove, though he was able to make progress on the reverse direction, which he had not been able to do before. He demonstrated regression and no improvement on the validation portion of the assessment: he correctly classified and justified three arguments on the pre-assessment and only one on the post assessment.

Details.

Attitudes and beliefs.

Travis was the only participant in the implementation study who did not state a preference for either proving or disproving, but he did say the he liked writing proofs because “whatever you proved you know. You know it holds. You don’t have to take somebody’s word on it.” While he was fully convinced of the truth of claims he had proven for himself, Travis was less convinced by the proofs of others unless he could understand their arguments. “If I understand the proof and can follow the logic, then I’m pretty confident of the theorem only because I can understand it. If I can’t understand it, I can’t really say anything about it because I don’t understand it.”

Travis’s post-assessment responses were recorded several months after the conclusion of the study because of an equipment malfunction which occurred during the original post-assessment, but during his re-interview, Travis stated that he did not believe any of his expressed attitudes would have been different at the time of the initial assessment than they were at that time.

Composition.

On item C1 on the pre-assessment, Travis attempted a direct proof, but he implicitly assumed m to be odd and then concluded the same thing. However, he was aware that there may have been issues with his argument: “I don’t know if this is cheating…I don’t think that’s a really good proof, but that’s doing my best.”

Travis identified the pertinent subclaims on item C2: “you know it’s divisible by 6 if it’s divisible by 2 and 3,” but he did not attempt the subarguments. Instead, he attempted a proof by induction, but he used *n*=2 as his base case, which was problematic because even if he had been able to complete the argument, it would have only established that $n^3 - n$ is divisible by 6 when $n \geq 2$. However, he did not progress beyond expanding and factoring the polynomial in the inductive step; stopping with the equation $(n+1)^3 + (n+1) = n(n+1)(n+2)$.

On item C3, he used the definition of triangular numbers to provide a base for his valid proof of one direction. While he mentioned the need for the second direction – “I guess the rest of the proof would have gone back the other way because, well it’s if and only if” – he did not attempt an argument for the other half of the biconditional.

Item C1 went much better for Travis on the post-assessment. He was able to avoid the converse argument entirely, and he used a direct proof by cases to prove the claim. His proof was lacking in clarity because he did not clearly define his variables, and he did not state a conclusion.

On item C2, he failed to acknowledge the subclaims and thus did not demonstrate a skill he had previously shown; this was captured as a regression in the coding scheme, but as he was working with a proving method that did not require the application of subclaims or the construction of subarguments, this researcher does not feel that this demonstrates a regression in proving ability. Additionally, he progressed further with his induction argument, correctly establishing the base case at n=0 and invoking his inductive hypothesis, reducing the problem to a need to establish that $3n^2 + 3n$ was also divisible by 6. He was unable to produce an argument for that, however.

Travis exhibited regression on item C3 that was more severe. He attempted to do an induction argument to establish the forward direction of the biconditional, but he inducted upon n, which was inappropriate and logically invalid. While he was able to use the definition of triangular numbers as a base for the faulty argument, he was unable to produce a complete argument for that direction as he had in the pre-assessment. He even had written down that $8\left(\frac{k(k+1)}{2}\right) + 1 = \cdots = (2k+1)^2$, but he did not recognize this as proof of that direction (see Figure 5.3), even though he was able to use that argument as a proof on the pre-assessment. He did attempt a general argument at the reverse direction, which he had not done previously, but as he was unsuccessful, the improvement there was tempered by the regression on the other direction.

On the pre-assessment, Travis switched proof methods one time on item C3 and none on the other two items. On the post-assessment, he switched methods one time on each of item C1 and item C2, and he made two switches on item C3.

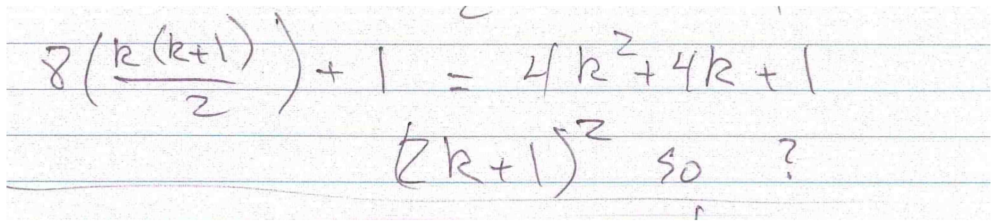


Figure 5.3 - An Excerpt from Travis' Post-Assessment Work on Item C3

Proof validation.

On the pre-assessment, Travis got confused on item CV, "The Converse," in sentence [3] when the author wrote $9m^2 = 3(3m^2)$. He didn't know why that was included, and he pointed out that the author concluded that m was a multiple of 3 but needed to show that n was, so he classified the argument as a 4, not meeting the standards of a proof. Travis then classified item RT, "The Real Thing," as a 1, a rigorous proof of the claim, without explanation.

On item EG, "Errors Galore," he was confused about why the author broke the argument into odd and even cases and thus chose option 5, "I can't classify this because I don't understand the argument." On item GP, "The Gap," he got stuck at the jump from sentence [3] to sentence [4], decided that the step was unsubstantiated, and classified the argument as a 4.

On item CV, even though his classification was valid, Travis's justification was problematic because he did not seem to take issue with the implicit assumption of the conclusion. However, not stating a concern does not mean he did not see the issue, so we take this classification to be correct. His classifications of items RT and GP were correct, and his justification on item GP was reasonable. His choice of option 5 on item EG cannot be considered either correct or incorrect.

Due to another equipment malfunction, no audio was recorded during the validation portion of Travis's post-assessment. As he only discussed his reasoning verbally, there are no data on his justifications for his classifications. He classified item CV as a 3, item RT as a 1, item EG as a 3, and item GP as a 1. Only the classification of item RT could be valid. The others are necessarily incorrect as the arguments in items CV and EG do not establish the validity of the claim, and the missing justification in item GP is too important for that argument to be considered a rigorous proof.

Exit re-interview.

Travis's exit interview was conducted several months after the conclusion of the study because an equipment malfunction had resulted in the loss of all audio data from this portion of the post-assessment.

Travis's confidence regarding writing proofs increased over the course of the study, and he attributed that increase to the cooperative group work in the seminar. He felt that seeing how other people reasoned and also how he reasoned and how their arguments made affected the people around him helped him learn how to better construct proofs. He also thought it would be beneficial to have more group work in his other classes. "I would like that for more of the math classes, just maybe a half hour where we just, I guess, talked to other people about what, other people as in your classmates, about what they're doing, how they're doing it." However, he did not think he was any better at reading and understanding arguments than he had been at the time of the pre-assessment.

Usher.

Overview.

Usher was by far the most advanced of the seminar participants in the implementation study. On both of the assessments he was able to prove the first two composition items as well as the forward direction of item C3, and he was also able to provide correct classifications and justifications for all four validation arguments. Additionally, he demonstrated all but one of the proof skills intentionally targeted by the composition items on both assessments. The only skill he did not demonstrate was the application of previously shown results.

Details.

Attitudes and beliefs.

Usher did not provide the researcher much insight into his attitudes or beliefs about proof and proving, due to his very short answers to all interview questions. The only change noticeable from pre-assessment to post-assessment was that while he first stated a preference for disproving over proving because of its ease, on the post-assessment Usher no longer stated a preference for either.

Composition.

Usher was able to produce valid proofs that met all clear and concise criteria for all items except for the backwards direction on item C3 ($8n+1$ a perfect square implies n triangular) on the pre-assessment. He did similarly well on the post-assessment and came close to proving the second direction on item C3 but did not recognize $4k^2 + 4k + 1$ as a perfect square despite having used that fact in his proof of the other direction.

On the pre-assessment, Usher changed from a direct proof attempt to contradiction and then to contrapositive on item C1. On the post-assessment, he was able to start with contrapositive and produce a valid proof. On item C2 on the pre-assessment, Usher started with induction, but on the post-assessment he worked briefly with direct and contrapositive arguments before settling into an induction proof. On both assessments, Usher switched methods three times on the backwards direction of item C3. He was able to prove the forward direction with a direct proof on both attempts without trying anything else first.

Proof validation.

On the pre-assessment, Usher classified item CV, "The Converse," as a 2, a rigorous proof of a different claim. He noticed that the author assumed the intended conclusion and worked with it to establish the hypothesis. It was clear to Usher that the author was not intending to prove something else but did. According to him, the argument proved the converse of the original claim, that "if n is a multiple of 3, then n^2 is a multiple of 3."

Usher thought the argument in item RT, "The Real Thing," was good until the final line. According to him, "they're proving it by contrapositive up until this and then they say it's a contradiction, but they assumed that to begin with, so there's no contradiction." Other than identifying a contradiction Usher didn't see, the author was seen as making no mistakes, so Usher classified the argument as a 3, a non-rigorous proof of the claim.

He determined that the argument in item EG, "Errors Galore," was not a proof because of "a lot of things." He commented in particular on the use of n to represent multiple quantities and the fact that the representation of n^2 as an odd integer was incorrect because $(3n + 1)^2$ is not an odd integer for all values of n. He summed up his concerns by saying "they didn't kind of like show anything. They just said things that aren't true," and classified the argument as a 4, not meeting the standards of a proof.

On item GP, "The Gap," Usher was not very specific about his concerns, but he classified the argument as a 3 because he wasn't fully convinced by the jump from sentence [3] to sentence [4]. All four of his classifications and corresponding justifications were correct.

On the post-assessment, in addition to pointing out that the author of item CV had assumed the conclusion in sentence [2], Usher took issue with the conclusion because the author had not established the stated conclusion that m was a multiple of 3. He classified the argument as a 4. His analysis of item RT was very similar to his analysis of that item on the pre-assessment: he pointed out that the author did not contradict anything stated in the proof but that the validity claim was definitely established by the argument, and once again classified it at a 3.

On item EG, Usher was initially not sure whether to choose option 4 or 5; however, since he was able to identify what the author was trying to do, he decided that he did understand the argument, so he classified the argument as a 4. He cited the use of n to represent multiple quantities as well as the unjustified conclusions as rationale. "So like this $(3n)^2$, that's only true for $n = 0$. And $[n^2 = (3n + 1)^2]$, that's never true. They used the same n for everything. And they're supposed to say that n^2 is divisible by 3, and they just kind of jump and say that n is divisible by 3."

Usher classified item GP as a 1, a rigorous proof of the claim, but when asked if he wanted to say anything about it, Usher explained that it would depend on the level of understanding the reader was expected to have. He thought that even for a reader with a lower level of understanding, the argument was salvageable, but the jump from sentence [3] to sentence [4] would need to be explained. Usher's classifications and explanations of item CV, RT, and EG were all correct. While a classification of 1 for item GP would generally be considered incorrect, the qualification Usher placed upon his decision shows that he understood the lacking justification to be an issue, and his reasoning was still considered valid

Exit interview.

Because his confidence regarding writing proofs was high at the beginning of the study, Usher did not feel that his participation had lead to any change in his confidence. However, he did feel more confident in his ability to read and understand proofs. Usher attributed this change to talking to his group mates, but he found that needing to explain his own reasoning was what made it easier to read and critique the reasoning of others.

"If I have to explain it to others and I have to explain it to myself in a way that, I can say a lot of things to myself that make sense to me but will be not true, so being more clear about what I say, it helps clear up what it actually means." In addition to leading to increased confidence, Usher felt that working in a cooperative group had been beneficial to his learning because there were more ideas coming in which helped him better understand what was right. While he felt that the group processing had been beneficial in the beginning, Usher said that his group's conversations lacked the depth that would have been needed for the exercise to be very helpful.

Individual Analysis for Comparison Participants

1865.

Overview.

1865's composition skills were the strongest of all the participants in the implementation study; he provided valid arguments for all three items on both assessements. His validation skills were similarly strong, as he was able to correctly justify all of his classifications; however, he declined to classify item EG, "Errors Galore," on the pre-assessment because he found the argument so confusing.

Details.

Attitudes and beliefs.

1865 agreed with most of the seminar participants that disproving is easier than proving, but on both assessments, he stated a preference for proving because he felt he gained more from the challenge of needing to provide a proof. However, a high level of confidence in his ability to write proofs accompanied his preference for the more difficult task. On the pre-assessment, 1865 said that he enjoyed writing proofs because it was "a great way to learn the material." On the post-assessment, he added that writing proofs was enjoyable because he generally felt like he knew when he had done it correctly.

Composition.

On the pre-assessment, this participant was the strongest of all participants in the study. He produced a valid proof of item C1 by contrapositive after writing a contradiction proof as well. It is unclear why he crossed out his valid contradiction proof in favor of the contrapositive argument. His argument of item C2 was valid as well, but was lacking in formality (see Figure 5.4).

1865 was not only able to prove both directions of the biconditional in item C3, but he explicitly applied the results from item C1 in the process and demonstrated all three tested proof skills. The only thing lacking in his proof was a concluding statement.

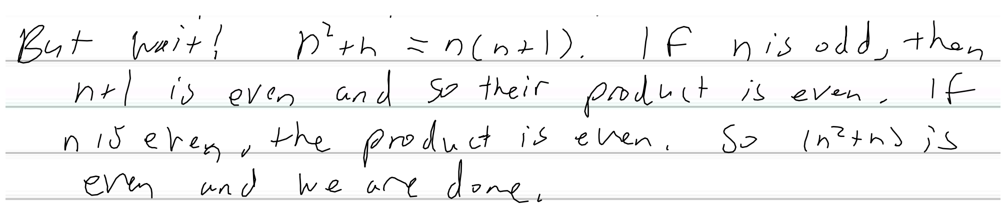

But wait! $n^2+n = n(n+1)$. If n is odd, then n+1 is even and so their product is even. If n is even, the product is even. So (n^2+n) is even and we are done.

Figure 5.4 - An Excerpt of 1865's Pre-Assessment Work on Item C2

There was not much room for improvement on the post-assessment, but 1865 did manage to do so. His proof of item C2 was no longer informal, and he included the missing concluding statements in item C3. On the post-assessment, he again wrote a contradiction argument for item C1, but he was apparently happier with it and did not produce a second argument as he had on the pre-assessment.

Proof validation.

On item CV, "The Converse," on the pre-assessment, 1865 noted "in the second line … they assume the conclusion to be true. So all they have proved is $p \Rightarrow p$ which is trivially true," and classified the argument as a 4, not meeting the standards of a proof. However, the argument does not actually prove $p \Rightarrow p$ because while the author states that n^2 is a multiple of 3, s/he does not work with that assumption.

1865 summarized the contrapositive argument in item RT, "The Real Thing," and classified it as a 1, a rigorous proof of the claim. "This is a good proof using the contrapositive. They assume $3 \nmid n$, which leaves only two cases to consider. Both cases imply $3 \nmid n^2$, so we can conclude that $3|n^2 \Rightarrow 3|n$."

On item EG, “Errors Galore,” he was confused and chose option 5, I cannot classify this, because I do not understand the argument. He stated “I can’t say what they’ve proven, because the argument is confusing. The English doesn’t correspond to the math,” but he did not elaborate. The missing justification in item GP, “The Gap,” was identified and provided by 1865 on the pre-assessment. “The structure is good, but they need to justify the step from ‘$nn = 3x$’ to ‘$3|n$’. It’s true because 3 is prime, but they should have said that.” He classified the argument as a 3 because of the gap.

While 1865’s classification of item CV was correct, because the author assumes the conclusion in sentence [2] and invalidates the argument, 1865’s claim that the argument proves $p \Rightarrow p$ is faulty. His choice of option 5 on item EG cannot be considered either correct or incorrect, but his other two classifications and justifications were valid.

On the post-assessment, 1865 again assigned a 4 to item CV. He pointed out that in the second line the conclusion was assumed, and said “the fourth line states that $n = 3m$, but that was stated without justification in line 2.” This justification is problematic because sentence [4] actually states that m is a multiple of 3 and says nothing about n.

On item RT, 1865 spotted the contrapositive argument again and classified the item as a 1, although he did state “the phrasing about contradictions was slightly confusing.” 1865 was not confused by item EG on this assessment and gave a lengthy description of the errors present, pointing out many of the errors (see Figure 5.5).

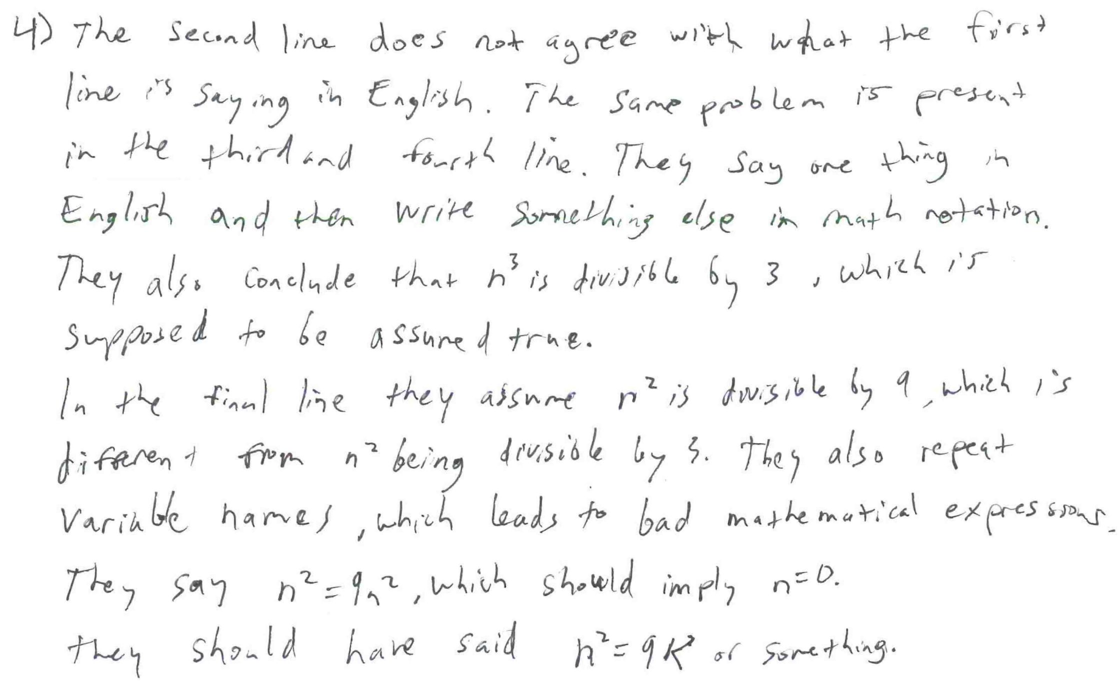

4) The second line does not agree with what the first line is saying in English. The same problem is present in the third and fourth line. They say one thing in English and then write something else in math notation. They also conclude that n^3 is divisible by 3, which is supposed to be assumed true.
In the final line they assume n^2 is divisible by 9, which is different from n^2 being divisible by 3. They also repeat variable names, which leads to bad mathematical expressions. They say $n^2 = 9n^2$, which should imply $n=0$.
They should have said $n^2 = 9k^2$ or something.

Figure 5.5 - 1865's Post-Assessment Work on Item EG

This participant once again pointed out the hole in the reasoning in item GP but decided it was more critical than he had judged it on the pre-assessment and classified the argument as a 4. "This is written in the correct format, but it doesn't have an argument. The sentence beginning 'Thus' requires some justification. It's a legitimate conclusion that follows from the previous lines, but it doesn't say why. For example, if 3 were not prime the conclusion would not be correct."

As on the pre-assessment, 1865 gave a valid classification with a problematic justification for item CV on the post-assessment. He also classified item RT correctly for the second time. No longer confused by item EG, 1865 provided a nice description of many of the errors present and gave a valid classification. His classification of item GP was not valid on the post-assessment, however, as his explanation describes a logically intact argument that is just missing justification, which the researcher described to the participants as a 3, a non-rigorous proof of the claim.

3099.

Overview.

3099 wasn't able to provide valid arguments for any of the composition items on the pre-assessment, but she was able to write a proof of one item on the post-assessment. Her validation results also showed improvement. While she was only able to correctly classify and justify two arguments on the pre-assessment, she correctly validated all four arguments on the post-assessment.

Details.

Attitudes and beliefs.

3099's statement about how convinced she was by rigorous proofs on the pre-assessment indicated that she held an authoritarian proof scheme. "Once I've seen a rigorous proof of a theorem, the level of confidence I feel as far as the theorem being true depends on where I saw it. If it's in a book, I'll believe it. If I just saw it on a blog or in elsewhere [*sic*], I wouldn't feel that confident that the theorem is true. It just depends on the source." She did not answer the question on the post-assessment.

3099 misunderstood the question about preference regarding proving and disproving either because she misread it or because she did not know what it means to disprove a claim. She equated disproving with proving by contradiction, and the same misinterpretation occurred on both assessments, indicating that she did not know what it meant to disprove a claim.

From the pre-assessment: “I don’t have a preference as far as proving or disproving a claim. It just depends on the claim. I definitely appreciate the fact that disproving can count as mathematical proof, thanks to truth tables/rules of logic. Sometimes it’s so much simpler to prove something by contradiction compared to proving by direct proof.”

From the post-assessment: “I prefer whatever method is faster ☺ [*sic*]. Sometimes it is faster to prove directly, and other times by contradiction. Thanks to the awesome rules of logic, both ways work when applied correctly to a particular proof. Proving $p \Rightarrow q$ is logically the same as proving not $p \Rightarrow$ not q.”

3099’s post-assessment statement also indicates that she does not understand the logical connection between contrapositive of an implication and the original implication.

Composition.

On the pre-assessment, 3099 first attempted to prove item C1 by induction but abandoned that work after establishing two base cases, scratching out her work. She then proceeded to produce a contrapositive argument, but there were logical and mathematical flaws. She called it a proof by contradiction but did not produce a contradiction; she incorrectly defined m as an even number and then provided an invalid justification for why m^2 was also even.

3099 turned to induction again on item C2, but her inductive argument was flawed and thus invalid. She correctly established a base case, n=0, but illustrated a misunderstanding of what the inductive hypothesis means and used a single example, n=1, in place of the inductive step (see Figure 5.6).

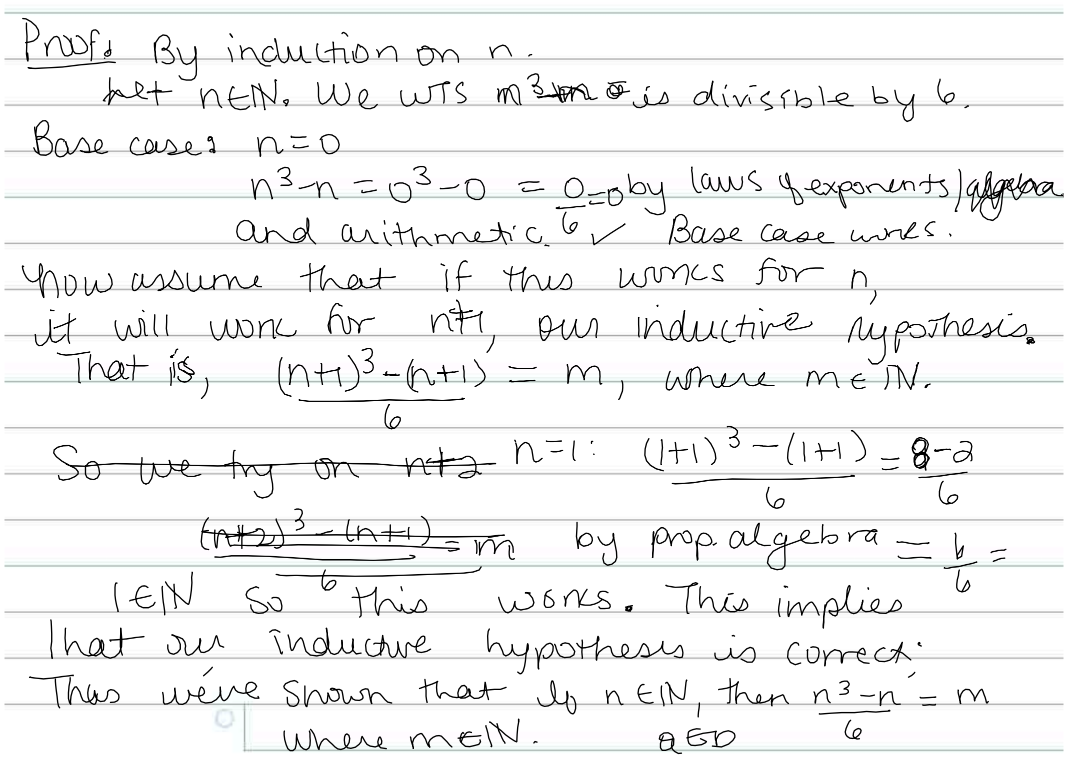

Proof: By induction on n.
~~Let~~ n∈ℕ. We WTS m^3 ~~-m~~ is divisible by 6.
Base case: n=0
$n^3-n=0^3-0=\frac{0}{6}=0$ by laws of exponents/algebra and arithmetic. ✓ Base case works.
Now assume that if this works for n, it will work for n+1, our inductive hypothesis.
That is, $\frac{(n+1)^3-(n+1)}{6}=m$, where m∈ℕ.
~~So we try on n+2~~ n=1: $\frac{(1+1)^3-(1+1)}{6}=\frac{8-2}{6}$
~~$\frac{(n+2)^3-(n+1)}{6}=m$~~ by prop algebra $=\frac{6}{6}=$
1∈ℕ So this works. This implies that our inductive hypothesis is correct.
Thus we've shown that if n∈ℕ, then $\frac{n^3-n}{6}=m$ where m∈ℕ. QED

Figure 5.6 - 3099's Pre-Assessment Work on Item C2

On item C3, 3099 attempted general arguments of both directions but was unsuccessful on both. She tried to use induction on m to prove that $8n+1=m^2$ implies n is triangular, which is inappropriate, and she was unable even to provide a base case because she chose to start with $m=4$, (which is impossible for n a natural number, let alone a triangular number). For the other direction, she was able to use the definition of triangular to form a basis for her argument, but she made a mathematical error when she dropped the $+1$ term from her computations and ended up with an expression that was not equal to a perfect square.

On the post assessment, 3099 produced a valid, error-free proof by contrapositive of item C1. Her induction argument of item C2 was better in that she demonstrated a better understanding of the principle of mathematical induction, but she inappropriately switched variables mid-argument and made an unjustified claim in the last line when she stated that $6k + 3n^2 + 3n$ was clearly divisible by 6, which needed justification.

On item C3, 3099 turned to induction for both directions, which is inappropriate in both cases. She attempted to induct on n to establish n triangular implies $8n + 1$ a perfect square, but in the inductive step that meant working with $n + 1$, which is never triangular when n is. For the reverse direction, she produced an unusual argument (see Figure 5.7) involving using a single example to establish the contrapositive and the claim that one can induct on k when no k is present.

Proof validation.

On item CV, "The Converse," on the pre-assessment, 3099 used her own lack of confidence in writing proofs to question her judgment when validating. "I'm no one to judge, because I can't write proofs to save my life, but ['$3m$ times $3m$'] doesn't clearly show that m is a multiple of 3." She went on to classify the argument as a 3, a non-rigorous proof of the claim, stating that it needed more "validation" and asking what properties were used. From the context, it is clear that by "validation" she meant justification.

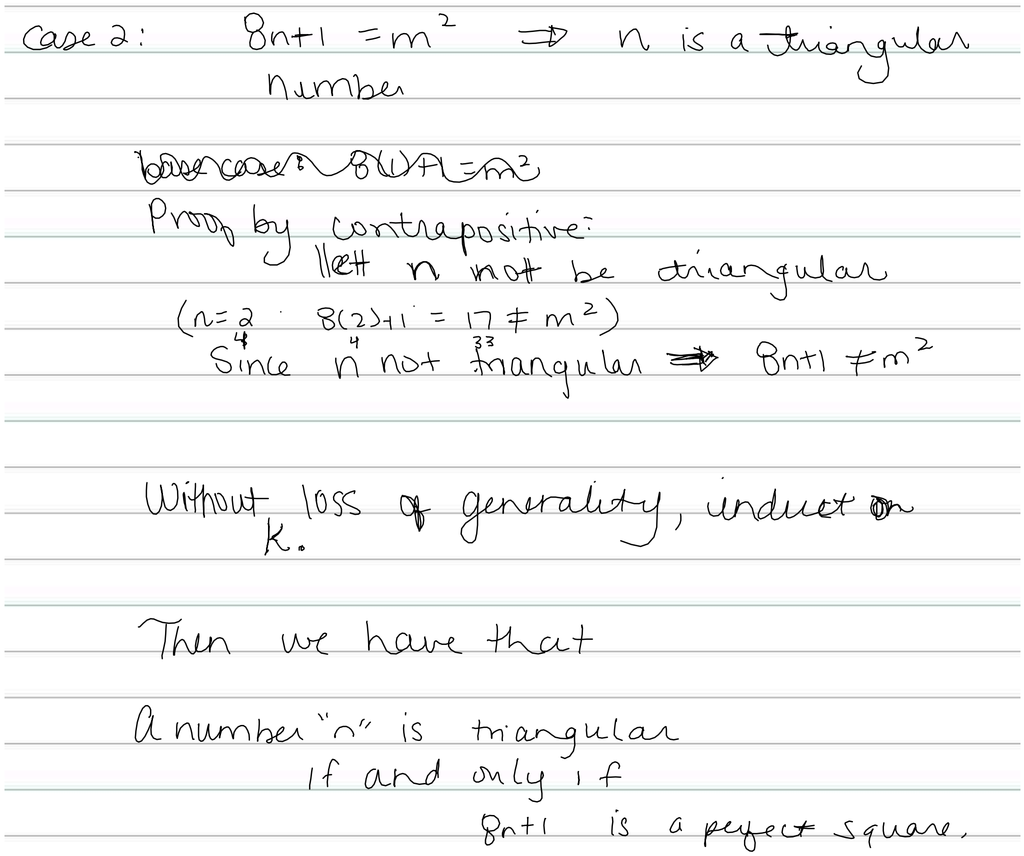


Figure 5.7 - Excerpt from 3099's Post-Assessment Work on Item C3

On item RT, "The Real Thing," she was again concerned about the lack of validation and the fact that the properties used weren't explicitly stated. That said, she still considered it a "great proof" but one of a different claim. She pointed out "a proof of a different claim was made: that $\neg q \Rightarrow \neg p$. This person should've written an additional sentence saying $(\neg q \Rightarrow \neg p)$ implies $(p \Rightarrow q)$," and classified the argument as a 2, a rigorous proof of a different claim. So while she recognized the equivalence of the contrapositive to the original implication, she thought the lack of a concluding statement linking the two meant that the argument did not prove the original claim.

3099 also classified item EG as a 2, but she did not explicitly state what claim she thought was being proved. She asked for more validation again, “maybe a sentence saying ‘$3n(3n)$ is a multiple of 3 because …,’” (ellipses are hers) and asked why factoring $9n^2$ as $3n(3n)$ implies that n is a multiple of 3 as stated in the argument. However, she did like the fact that the author attempted to prove by cases.

Despite the fact that the researcher encouraged participants to ask any clarification questions that arose, 3099 did not ask about the divisibility notation in item GP. She stated “I don’t even know what this means.” Because of that, she was concerned with sentence [4] not actually being a sentence, and she pointed out that proofs are supposed to be grammatically correct. She then classified the argument as a 4, not meeting the standards of a proof.

Item CV cannot be considered a proof, rigorous or not, of the claim since the conclusion is assumed in sentence [2], so her classification is incorrect. However, her classification of item RT is valid and fairly well reasoned. It is interesting that she knew that the contrapositive implied the original and that the argument proved the contrapositive but still did not consider the argument as a proof of the original claim. As item EG proves nothing, her classification is incorrect and the fact that she didn’t state an alternate claim is problematic. Since she didn’t understand the notation in GP and found some statements to be nonsensical, a classification of 4 is reasonable, but choosing 5, “I cannot classify this, because I do not understand the argument” would have been more appropriate.

3099’s post-assessment validations were much better than they had been on the pre-assessment. On item CV, she recognized that the author assumed the conclusion in sentence [2] without proving it and classified the argument as a 4. She wanted to see a concluding statement in item RT, "the only thing I would add is a final statement indicating that this contradiction implies that $n^2 = 3m \Rightarrow n = 3m$," and classified the argument as a 1, a rigorous proof of the claim. However, her provided conclusion is problematic, because she used m to stand for two distinct quantities. On item EG she recognized that implicit assumptions were being made about n when the author had claimed to be working with n^2, and she was unclear about sentences [2] and [3]. She classified the argument as a 4. Between pre- and post-assessment, 3099 learned the meaning of the divisibility notation used in item GP and was only concerned with the reasoning gap between sentences [3] and [4], noting that the author "didn't prove anything; didn't show steps but arguments are correct." With the exception of using m to stand for distinct quantities, these classifications are all valid and well justified.

5105.

Overview.

5105 performed similarly on the first two items of the composition portion of the assessments, but she was unable to prove the forward direction of item C3 on the post-assessment as she had done on the pre-assessment. Her validation results did not indicate any regression, however. The two items she correctly validated on the pre-assessment were correctly validated on the post assessment as well, and she was able to correctly classify item EG, "Errors Galore," on the post-assessment, which she had not done on the pre-assessment.

Details.

Attitudes and beliefs.

On the pre-assessment, 5105 expressed many reasons for writing proofs of claims that are already known to be true: to be able to focus on strategy, to practice proving skills, and "simply for the beauty of a classic, canonical proof." On the post-assessment she said that the practice encourages perseverance. On both assessments she said that she could be confident in the truth of a proven claim provided that the proof did not skip steps that were not obvious to her. Like 1865, she found the practice of proving to be more rewarding than that of disproving and said she preferred it.

Composition.

5105 provided stream-of-consciousness written work for her compositions, and her work shows great persistence. On item C1, after doing scratch work and part of a proof by contradiction followed by more side work, 5105 produced a valid contradiction proof. Her work on item C2 was much more difficult to interpret; she did some side work that involved factoring $n^3 - n$ in a product of monomials and included an incorrect statement about divisibility, "if a # divides a #, then it divides one of the factors." Then she moved on to an attempted proof by induction. She correctly defined her base case as $n = 1$ and set up her induction hypothesis, but she got stuck at $(n+1)^3 - (n+1) = n(3n+3) + 6\tilde{k}$ where $6\tilde{k} = n^3 - n$ by the induction hypothesis. She made several attempts at the induction proof but got stuck in that spot each time and concluded "I know it should be obvious I'm sure its here somewhere!"

On item C3, she was able to use the definition of triangular numbers to form a basis for a proof that n triangular implies $8n + 1$ is a perfect square, but apart from writing down the assumption that $8n + 1$ is a perfect square, she was unable to make progress on the reverse direction of the biconditional.

On the post-assessment, 5105 again wrote a very nice contradiction proof of item C1 but without as much scratch work as on the pre-assessment. She attempted proof by induction of item C2 again but got stuck in the same place. 5105 did not do as well on item C3 on the post-assessment as she had on the pre-assessment. She attempted an argument for $8n + 1$ a perfect square implies n triangular, but it was a flawed and incomplete argument. She was also unable to produce an argument for the other direction despite having done so on the pre-assessment. A critical flaw in her arguments for both directions was the conflation of n and k, which are very different quantities in the problem. At one point, she even explicitly stated that they were equal (see Figure 5.8).

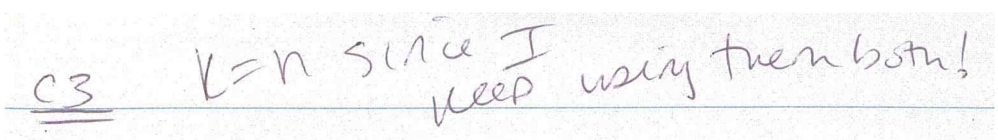


Figure 5.8 - Excerpt from 5105's Post-Assessment Work on Item C3

Proof validation.

On item CV, "The Converse," 5105 did not notice the author was assuming the conclusion, but she did get confused between sentences [3] and [4] and did some computations of her own to clear up her confusion. Ultimately, she classified the arguments as a 4, not meeting the standards of a proof, “since the claim was not proved.”

On item RT, “The Real Thing,” 5105 wanted the author to explicitly state the assumption that n^2 was a multiple of 3 “to demonstrate more clearly,” but she had no other qualms and classified the argument as a 1, a rigorous proof of the claim.

5105 chose option 5 on item EG, “Errors Galore,” stating “I cannot classify this, because I do not understand the argument,” because she found “the word ‘odd’ disturbing” and noted that $3n(n+2)+1$ is not divisible by 3. This participant did not find any issue in item GP, “The Gap,” and rated it as a 1. “Nice, it seems clear and concise. Am I missing something?” 5105’s first three classifications are valid and well justified, but as item GP cannot be considered a rigorous proof of the claim, that classification is incorrect.

On the post-assessment, 5105 noticed that the author of item CV assumed the conclusion and from that assumption ended up proving the converse. Thus, she classified the argument as a 2, a rigorous proof of a different claim, “since this student did show that if n is a mult of 3, then n^2 is a mult. [*sic*] of 3.”

She was again happy with item RT, writing “this pf seems rigorous, nicely written, and easy to understand,” and classified it as a 1. The number of errors in item EG confused 5105, but she also knew that it was incorrect, and she classified the argument as a 4 while also choosing option 5.

On item GP, 5105 noticed the reasoning gap and did some work to convince herself that sentence [3] does imply sentence [4]. She determined that it did, but thought that it needed “maybe one extra line of justification,” but ultimately decided it was unnecessary and classified the argument as a 1. Again, 5105’s classifications of items CV, RT, and EG were valid and well justified, and her classification of item GP was incorrect; however, her reasoning on item GP was much improved since she attended to the missing justification.

5635.

Overview.

On the pre-assessment, 5635 provided empirical arguments for all the composition items. While she was able to provide a valid deductive argument for one of the items on the post-assessment, she wasn’t able to construct any argument for the other two items. 5635 also struggled greatly with the validation portion of the assessment, correctly validating two arguments on the pre-assessment and only one on the post-assessment.

Details.

Attitudes and beliefs.

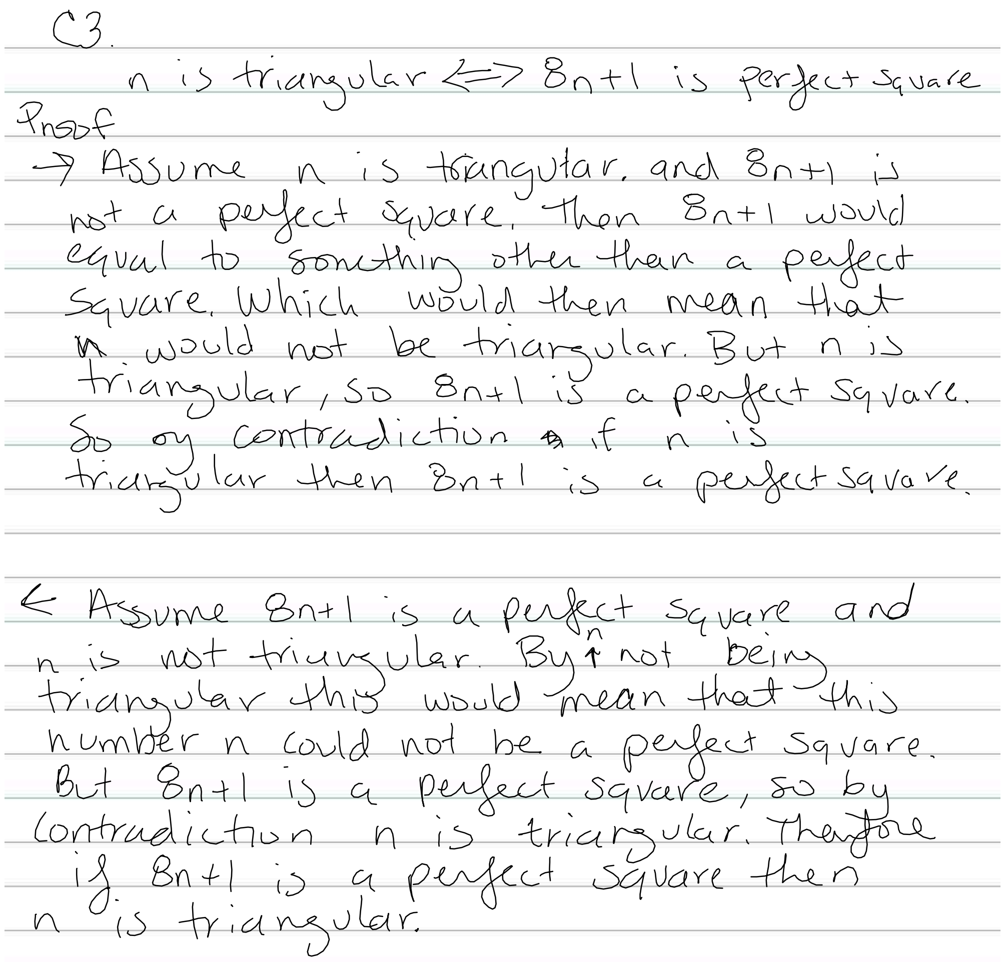
Of all the participants in the implementation study, 5635 was the least convinced by proofs written by others. On the pre-assessment she said that she needed to validate the “truthness [sic] of the theorem” for herself before she could feel fully confident that the theorem was true. In fact, she saw the purpose of writing proofs of theorems already known to be true as verification. “The purpose of writing proofs of theorems that are already known to be true is so that one personally can know how ‘true’ this theorem really is. Because as skeptical as mathematicians are, one may never know if a theorem is true or not." On the post-assessment she said that she was confident she could determine whether or not a theorem was true, which indicates that she still felt the need to verify the truth of the theorem for herself.

On the pre-assessment, 5635 stated that she did not have a preference for either proving or disproving because she did not think she was very good at either; however, on the post-assessment she stated that she didn't have a preference but added that if she had to chose she "would prefer to prove a claim because … it's simpler."

Composition.

On the pre-assessment, 5635 provided empirical arguments for all items. On item C1, she produced a single example of the converse, for $m = 3$, m^2 is odd, and concluded that "for all odd integers m^2, then m is odd." On item C2, she chose $n = \frac{1}{2}$ and showed $n^3 - n$ was not divisible by 6. She then concluded that "by contradiction, if n is a natural number, then $n^3 - n$ is divisible by 6." She gave three examples for item C3 but made no generalizations.

5635 seemed to have abandoned her empirical proof scheme by the post assessment. She was able to produce an informal, yet valid, contradiction argument for item C1. On items C2 and C3, 5635 interpreted claims and reworded them informally, but she did not provide any examples or arguments (see Figure 5.9).



C3.

n is triangular ⟺ 8n+1 is perfect square

Proof

→ Assume n is triangular, and 8n+1 is not a perfect square. Then 8n+1 would equal to something other than a perfect square. Which would then mean that n would not be triangular. But n is triangular, so 8n+1 is a perfect square. So by contradiction if n is triangular then 8n+1 is a perfect square.

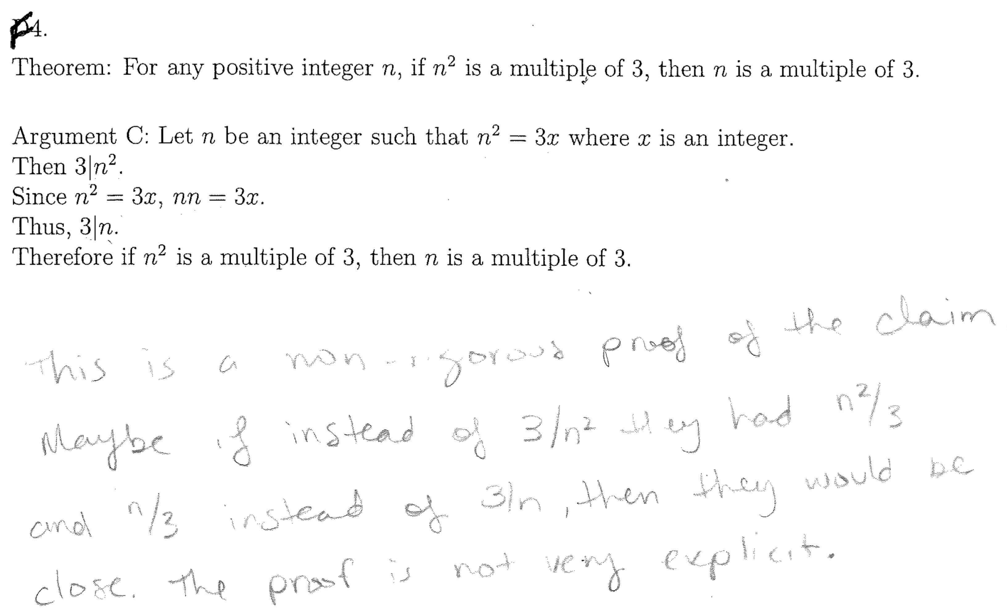
← Assume 8n+1 is a perfect square and n is not triangular. By n not being triangular this would mean that this number n could not be a perfect square. But 8n+1 is a perfect square, so by contradiction n is triangular. Therefore if 8n+1 is a perfect square then n is triangular.

Figure 5.9 - 5635's Post-Assessment Work on Item C3

Proof validation.

5635 classified item CV as a 1, a rigorous proof of the claim, on the pre-assessment and did not provide further discussion or explanation. She also classified item RT as a 1 without explanation. She decided item EG was "a rigorous proof of a different claim. because it is trying to prove something other than the actual proof. Its trying to prove if n^2 is odd, then n is odd." It is not clear that 5635 understood the notation used in item

GP. She classified it as a 3, a non-rigorous proof of the claim, but her proposed improvements are nonsensical (see Figure 5.10). Her classification of item RT was valid,


4.

Theorem: For any positive integer n, if n^2 is a multiple of 3, then n is a multiple of 3.

Argument C: Let n be an integer such that $n^2 = 3x$ where x is an integer.
Then $3|n^2$.
Since $n^2 = 3x$, $nn = 3x$.
Thus, $3|n$.
Therefore if n^2 is a multiple of 3, then n is a multiple of 3.

this is a non-rigorous proof of the claim. Maybe if instead of $3/n^2$ they had $n^2/3$ and $n/3$ instead of $3/n$, then they would be close. the proof is not very explicit.

Figure 5.10 - 5635's Pre-Assessment Work on Item GP

but her classifications of items CV and EG were not. While item GP can be considered a non-rigorous proof of the claim as 5635 stated, her justification for that classification is problematic since replacing $3|n^2$ and $3|n$ with $\frac{n^2}{3}$ and $\frac{n}{3}$, as proposed, would destroy the grammar of the argument and render it incomprehensible.

On the post-assessment, 5635 again classified item CV as a 1, providing this justification: "This is a rigorous proof of the claim because it states step by step each claim of their proof and one can clearly understand each step and its motive." However, she no longer considered item RT to be a valid proof the claim. Instead, she classified it as a 2, a rigorous proof of a different claim, writing that it "proved that if n is a multiple of 3, then n^2 is a multiple of 3."

She classified item EG as a 1 and wrote that the argument clearly stated "in various ways how the assumption leads to the conclusion." Her classification of item GP remained the same, but her justification was improved, if imprecise: "this is a non-rigorous proof of the claim because it is not clearly understood what is exactly trying to be stated. There isn't enough explanation and the explanation is very vague."

Her classifications of the first three items are all incorrect since items CV and EG do not establish the validity of the claims, and the alternate claim she determined item RT proved is not supported by that argument. Her classification of item GP is valid, but her justification is lacking in detail.

6293.

Overview.

6293 was one of the few participants to mention the benefits of using examples during the proving process, but he did not attempt to use many examples during his own composition attempts, which were all unsuccessful. The only argument he was able to correctly validate was item RT, "The Real Thing," which was the only valid argument in the set.

Details.

Attitudes and beliefs.

Like most of the other comparison participants, 6293 said on both assessments that he preferred proving a claim to disproving one saying that proving "is more interesting." He also said on both assessments that proving results that are already known to be true improves one's mathematical reasoning ability. He had little previous experience writing proofs, but he knew empirical evidence was insufficient for establishing the truth of a claim, although he felt that such evidence can improve someone's understanding of the subject matter.

Composition.

6293 started his work on item C1 with a graph, presumably of $y = x^2$, on the pre-assessment. He crossed out all written work on that page, but did not cross out the graph itself. He then moved to an attempt at a general argument but started with assumptions about m and reached conclusions about $m - 1$ that were not logically equivalent to a validation of the claim. On item C2, he attempted to work directly, but made little progress with that argument. He then split the claim into subclaims and attempted to construct subarguments. He was able to justify divisibility by 2, but he made a faulty claim about divisibility by 3, stating "if n is even then one of the terms $(n - 1)$ or $(n + 1)$ is divisible by 3." So he failed to prove divisibility by 3 in the case where n is even and did not attempt an argument in the odd case. 6293 was able to connect the hint provided with the definition of triangular number, but he was unable to provide an argument for either direction of the biconditional, and there is no evidence that he understood there were two directions that needed to be established.

The post-assessment showed some improvement as well as some regression for this participant. His argument for item C1 again started with assumptions about m and was unfinished. He established the subclaims appropriate for item C2 and started attempting to produce subarguments by cases. However, he could not complete an argument for the even case and did not begin an argument for the odd case, thus ultimately making less progress than he had on the pre-assessment. 6293 was almost successful at proving the forward direction of item C3, but he wrote that he was assuming k to be triangular instead of n, which invalidated his argument. There is no argument provided for the reverse direction, and again there is no indication that he knew there needed to be one.

Proof validation

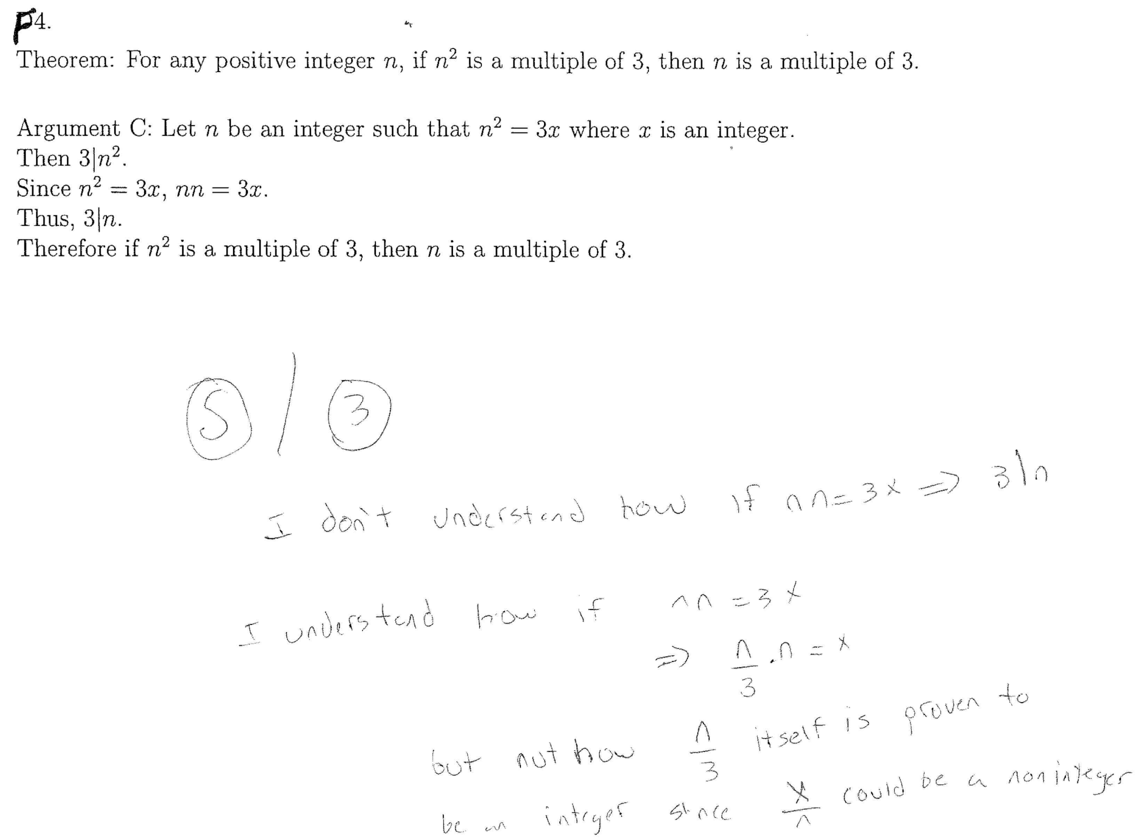
6293 had not taken a proof-based course prior to enrolling in the study and struggled with making sense of the arguments he was asked to validate. On item CV, "The Converse," he classified the argument as both a 3, a non-rigorous proof of the claim, and a 4, not meeting the standards of a proof. He wrote, "It looks like they proved it assuming the argument was true and then worked backwards. Not sure what rigorous means exactly since I haven't taken a proof-based course this argument doesn't seem to prove m is a multiple of 3. It only seems to say $n = 3m$."

He classified item RT, "The Real Thing," as a 3, but his description is more consistent with a classification of a 2, a rigorous proof of a different claim: "They only proved that if n is not a multiple of 3, then its square is not a multiple of 3, I don't think this makes the theorem true."

On item EG, "Errors Galore," 6293 was confused by the author's use of n to represent different quantities, writing "shouldn't you substitute another letter in," and chose option 5 – I cannot classify this, because I do not understand the argument. He also chose option 5 on item GP, "The Gap," because he wasn't sure how the author could conclude sentence [4] from the points that had been stated.

It is interesting that despite his lack of experience working with proofs, 6293 was able to identify errors missed by more experienced participants. His classification and justification for item CV is correct, and his reasoning on item RT is valid; it is understandable that he would not know that a contrapositive statement is equivalent to its corresponding implication. He also picked up on a major flaw in item EG that most other participants did not attend to and was concerned by the reasoning gap in item GP.

6293's post-assessment classification for item CV does not make sense with his justification. He classified the argument as a 3, but he wrote that the argument "only seems to prove that if n^2 is a multiple of 9, then n is a multiple of 3," which would mean that it was a proof of a different claim. However, he also pointed out that the conclusion is assumed in the second line.



F4.

Theorem: For any positive integer n, if n^2 is a multiple of 3, then n is a multiple of 3.

Argument C: Let n be an integer such that $n^2 = 3x$ where x is an integer.
Then $3|n^2$.
Since $n^2 = 3x$, $nn = 3x$.
Thus, $3|n$.
Therefore if n^2 is a multiple of 3, then n is a multiple of 3.

Figure 5.11 - 6293's Post-Assessment Work on Item GP

On item RT, he classified the argument as a 2 and specified that it was proving "that if n is not a multiple of 3, its square is not a multiple of 3." He went on to say that he does not believe proving a claim this way is logically sound, which again is understandable, given his background. On item EG, he was again troubled by the use of n to represent two distinct quantities and he furthermore noticed that sentence [3] was not supported by sentence [2]. He classified the argument as a 3.

6293 noticed the justification gap in item GP and classified the argument as a 3 while also choosing option 5 (see Figure 5.11). Given the severity of the errors pointed out by 6293, his classifications of 3 for items CV and EG were surprising; his classifications for the other two items were valid and well reasoned, however. Some of the dissonance between his classifications and justifications can be explained by his lack of experience working with proofs outside of this study.

Summary

With the exception of Usher, who was the strongest of all seminar participants with regards to both composition and proof validation, all the seminar participants demonstrated improved proof composition skills on the post assessment. Each of those participants were able to improve the main argument code on at least one of their arguments and to demonstrate on the post-assessment at least one of the intentionally targeted proof skills they had not demonstrated on the pre-assessment. While these data cannot conclusively determine the reason for the improvements, the seminar participants responded very positively to the experience of working in cooperative groups during the study seminar, and the comparison participants did not demonstrate much improvement. The conjunction of these results is consistent with the researcher's hypothesis that working in a Cooperative Learning environment on carefully designed materials may be beneficial to the acquisition and development of proof skills.

Chapter 6 - Discussion

Overview

Research questions.

This study examined the following questions:

Is there evidence that after working on proof-based problems in a Cooperative Learning environment there are measurable differences in

- an individual's attitudes about mathematical proof?
- an individual's proof composition skills?
- an individual's proof validation skills?

Many students graduating with bachelor's degrees in mathematics struggle with producing and understanding mathematical proofs at the time of graduation, yet members of the mathematical community deem these skills essential to the study of mathematics. The research presented in this dissertation was undertaken in order to begin to understand the relationship between Cooperative Learning experiences and the transition to proof. The results of this research and further research motivated by the studies presented here could have a profound impact on how transition to proof and other proof classes are taught.

Findings overview.

There was very little change in the general attitudes and beliefs about mathematical proof among the participants. This is unsurprising, as all participants were interested in studying mathematics and had generally positive attitudes about mathematics and proof at the beginning of each study. There was also little measurable change in the validation skills of both the seminar and the comparison participants; perhaps, this is more surprising since most of the seminar participants demonstrated stronger composition skills on the post-assessment than they had on the pre-assessment, and composition and validation skills are thought to be linked (Selden & Selden, 2007). Because of the design of the studies discussed in this dissertation, no definite conclusions can be drawn about what facilitated the improvement of the seminar participants' composition skills, but these results motivate the need for further study into the relationship between Cooperative Learning experiences and the transition to proof. They also bring up questions about how validation and composition skills are related, and whether they need to be taught explicitly. In addition to demonstrating improved composition skills on the post-assessment, most seminar participants changed proof methods more frequently on the post-assessment than they had on the pre-assessment. While this increased tendency to switch methods is similar to the findings of E. Hart (1994), the relationship between this flexibility and proving expertise is complicated and also warrants more study. Investigating those questions could have a profound impact on how transition to proof and other proof-based courses are taught in the future.

The existence and identification of transferrable, content-independent proof skills also merits further investigation. The researcher of the studies presented here designed the composition tasks to test specific skills she believed to be transferrable across content areas. In the literature, such content-independent skills are implicitly assumed to exist (e.g. Blanton & Stylianou, 2003; Selden & Selden, 2003; Sowder & Harel, 2003; Weber, 2003), but that has not been explicitly studied. Since most seminar participants demonstrated at least one proof skill on the post-assessment that they had not demonstrated on the pre-assessment, and the content areas of the assessments and the seminar were distinct, this research lends support to the notion that such skills exist, but more study needs to be done on this.

Cooperative Learning Strategies

While the primary research questions did not concern the effective use of Cooperative Learning Strategies, it is important to discuss what the researcher learned about working as a CL instructor. During the pilot study, the cooperative groups did not work well together as teams. Participants in both groups preferred to work independently on the problem sets, share their work with their group mates, and then come to a group conclusion about what the recorder should write as the group's solution. Additionally, once the recorders started writing, they were frequently left alone by their group mates who started to work on other problems. Responses from the pilot study to the cooperative group work were largely negative on the post-assessments as most participants did not feel that they had benefitted from working as members of a cooperative group.

In an attempt to combat the participants' inclination to work independently, the researcher gave the groups in the implementation study only one copy of each problem set. During that study, participants sat closer together and engaged in more discussion and debate about the problem sets than their pilot study counterparts had.

To further foster positive group dynamics, the researcher adjusted the assigned roles and group processing questions. Because she felt the role of Explainer could not be adequately filled due to the constraints of the study design, the researcher abandoned that role after the pilot study and introduced the Yes-Man role in its place. Instead of bearing the responsibility of explaining and summarizing, the Yes-Man's goal was to further discussions one step at a time by saying "yes, and …" and helping his/her group mates determine what other conclusions could be made. The participants in the pilot study were often not particularly cognizant of the roles that needed to be filled, so the researcher added questions to the group processing exercise addressing that issue. During their post-assessment interviews, implementation study participants responded much more positively to the group work and the majority felt that working cooperatively had been beneficial to their learning.

Attitudes/Beliefs

There was very little change from pre-assessment to post-assessment in the implementation participants' reported attitudes and beliefs about mathematics and mathematical proof. This is not surprising since all participants volunteered to take part in the study and had fairly positive relationships with mathematics before enrolling. It is unlikely that a student who did not enjoy mathematics or want to learn more about proof would have volunteered to take part.

Abandoning the Likert-type scale survey of the pilot study and replacing it with open ended questions gave the researcher a deeper, though still limited, understanding of the implementation participants and their attitudes. Even though the data from the implementation survey did not capture changes in the participants' attitudes, they were richer than the data from the pilot study survey; participants were able to provide explanations for their responses, which was not an opportunity shared by their pilot study counterparts

Composition

In both the pilot and implementation studies, almost all of the seminar participants were more successful at writing proofs on the post-assessment than they had been on the pre-assessment. These results support the researcher's hypothesis that working in structured, cooperative groups on carefully crafted problem sets may facilitate the acquisition and development of proof skills.

Participants' increased flexibility with proving methods on the post-assessment is an outcome that is consistent with the intentional design of the seminar including the group work structure and the problem sets. During the seminar sessions, when groups were working together on the problem sets, group members brought diverse backgrounds, perspectives and skill sets to the discussion. This diversity allowed the groups to overcome impasses and produce complete arguments for all claims with minimal intervention from the researcher. Even though participants often were not focused on fulfilling their assigned roles, the groups generally paid attention to the need for explanation and skepticism, and many participants both questioned their groups' work and supported it with explanations. Within their groups, the seminar participants speculated, conjectured and used trial and error to engage with the material on the problem sets, which is in stark contrast with how students generally experience instruction in proof-based mathematics (Almeida, 2000) and more closely aligned with the activities employed by professional mathematicians and expert provers (e.g. Maclane, 1994). Instead of being stuck on their own when they reached impasses, participants had the opportunity to ask questions and see their progress from alternative perspectives. Even in the pilot study this cooperation occurred because participants would stop working independently to engage with their group mates when someone was at an impasse.

Successful cooperative work cannot be accomplished if the work isn't engaging (e.g. Heller & Hollabaugh, 1991). The problem sets designed and adapted for this study encouraged conjecture and discovery by asking participants to explore examples and form their own conjectures. They also asked participants to "prove or disprove" claims to engage them more deeply with the material and their own uncertainty. Additionally, many of the problems asked participants to outline arguments utilizing several different proving methods in order to help them become more familiar with different methods, more comfortable at identifying productive proving pathways, and more flexible with switching proving methods. The seminar students and the researcher also engaged in multiple discussions about when to employ the differing proof strategies and how to determine whether or not such an outlined strategy might be effective or ineffective.

Hart's (1994) study of abstract algebra proof skills showed that as students' proof skills increased, they were more likely to change proof methods until they reached the final stage of development when their tendency to change methods decreased. As discussed in chapters 4 and 5, most seminar participants exhibited greater comfort and flexibility with proving methods on the post-assessments, and these results are consistent with Hart's findings. However, whether the increased flexibility was responsible for the improvement, the participants became more flexible because their skill level increased, or the skill level and flexibility increased simultaneously but independently is not addressed here. While the study was not designed to test the relationship between proving expertise and flexibility, it raises important questions for future research.

It is still an interesting question that merits further investigation since it is not the case that expert provers always exhibit a low tendency to switch proof methods. Especially when working in an unfamiliar area, even research mathematicians will sometimes switch proof methods multiple times while trying to prove a single theorem. The author of this dissertation hypothesizes that both comfort level with the mathematical context and proving expertise are needed for a prover to be able to consistently choose a productive proof method at the beginning of a proof attempt. However, for a novice prover, unfamiliarity with alternative proof methods and/or an unwillingness to change proof methods when at an impasse generally results in an inability to consistently prove theorems. In this study, we saw that increased flexibility was generally paired with improved performance, although that flexibility itself didn't necessarily indicate a student's position along a proving expertise continuum.

Content-Independent Proof Skills

When selecting composition items for the assessments, the researcher intentionally chose items that necessitated a variety of skills; the right-hand column of Table 6.1 indicates the specific context independent proof skill(s) each item intended to assess. She chose a mathematical context that would be accessible to a broad range of students and then designed and adapted problem sets with a different content focus from that of the assessment items. The purpose of assessing in one content area and working in another was to see if participants' skills increased despite receiving no instruction in content-specific proof methods.

Each of the specific skills tested by the composition items was addressed during the seminar, some more so than others. The data collected show that most seminar participants wrote better arguments or more complete partial arguments on the post-assessment than they had on the pre-assessment despite having received no instruction in number theory in the interim.

The participants were asked multiple times to write or outline arguments using indirect proving methods, since many of the problems the participants worked with explicitly asked them to outline the structure of direct, contradiction, and contrapositive proofs. The researcher also engaged in multiple discussions with the participants about the different proof methods and when and how to use them.

Assessment Item	Proof Skill(s) Tested
C1. Prove: If m^2 is odd, then m is odd.	A. Use of indirect proof methods. B. Avoidance of a more appealing but logically inequivalent converse argument.
C2. Prove: If n is a natural number, then n^3-n is divisible by 6.	A. Ability to identify pertinent subclaims and construct subarguments (divisibility by 2 and 3).
C3. A *triangular number* is defined as a natural number that can be written as the sum of consecutive integers, starting with 1. Prove: A number, n, is triangular if and only if $8n+1$ is a perfect square. (You may use the fact that $1+2+...+k = \frac{k(k+1)}{2}$.)	A. Use of the specifics of a definition to form a basis for a proof. B. Ability to identify the logical implications of "if and only if" statements. C. Use of previously established results (to prove $8n+1$ a perfect square implies that n is triangular, the result of item C1 needs to be applied).

Table 6.1 - Composition Items and Targeted Skills

The construction of subclaims and subarguments was not explicitly discussed during the seminar, but the participants worked with both on the problem sets. For example, in the problem set on the derivative-like function and fixed points (see Appendix 6c), participants were asked to prove that numbers of the form $4k+3$ cannot be written as the sum of two squares. They were guided, by the researcher in the pilot study and by the problem set in the implementation study, to prove the claim by contrapositive using three subarguments.

On several problem sets, the participants were provided with definitions and required to work with those definitions to form the basis of proofs; however, there were no questions that asked the participants to apply any results they had proved previously in the seminar. Very few participants demonstrated that skill (the ability to apply the a previously proved result) on the assessments, and it is possible more of them would have on the post-assessment if it had been addressed in the seminar.

All of the seminar participants from both studies, with the exceptions of Zach and Usher, demonstrated at least one proof skill on the post-assessment that they had not demonstrated on the pre-assessment. Zach did not put significant effort into his post-assessment, see Chapter 4 for discussion, and Usher demonstrated all but one proof skill on the pre-assessment. The only skill he did not demonstrate was the ability to apply previously proven results.

Much of the research literature on proof and proving discusses proof skills as if they were applicable across content areas (e.g. Blanton & Stylianou, 2003; Selden & Selden, 2003; Sowder & Harel, 2003; Weber, 2003), but whether such skills exist and are transferrable has not been studied. The researcher of this dissertation chose to look for the skills listed in Table 6.1 because they can be employed to prove claims in a variety of mathematical domains and because they are skills often addressed in transition to proof courses (e.g. Moore, 1994; Levine & Shanfelder, 2000; McLoughlin, 2010; Selden & Selden 2007). The fact that seminar participants improved on the composition portion of the assessment after working on argument generation in a mathematical context distinct from the context of the assessment is consistent with the hypothesis that content-independent, transferrable skills exist. The participants improved on specific skills such as identifying and utilizing different argument structures and applying the details of a definition to a particular argument. However, these studies did not test whether students could apply such skills across different content areas, which is an intended continuation of this research.

Proof Validation

Because students need to learn how to assess their own proofs for correctness as they are writing, is has been suggested that students should learn to read and check proofs other students have written for accuracy (Selden & Selden, 2007). However, this research shows that students can become better at writing proofs without a corresponding increase in their ability to validate student-generated arguments, which sheds some doubt on the necessity of developing validation skills for constructing proofs. This was true for the seminar participants despite the fact that they engaged in explicit validation activities during the seminar sessions. This researcher speculates that the skills required for validating one's own arguments are distinct from the skills required to identify errors in someone else's. This speculation is supported by 6772's written proof of a claim's converse and simultaneous rejection of the converse argument presented in the validation portion of the assessment.

This is not to say that the ability to read and reliably check others' arguments for accuracy is unrelated to the ability to read and check one's own arguments, but how the two abilities are related has not been rigorously studied.

While the validation tool was created using tools that had previously been used to explore students' validation skills (Selden & Selden, 2003; Weber, 2010), this researcher found the tool to be inadequate. The presented arguments did not fit neatly into the classification categories provided, and the researcher was unable to determine if errors the participants didn't mention weren't noticed, were determined not to be errors, or simply weren't mentioned. Even if one is concerned solely with whether or not the arguments represent valid proof, it is unclear whether these four arguments provide a basis for such investigation, as is underscored by Weber's (2008) study on how mathematicians validate proofs.

In future research on validation skills, the researcher plans to use the argument assessment tool she adapted to assess the participants' compositions to have participants evaluate student-generated arguments. Such a tool will provide the participants with a more robust evaluation language for discussing the arguments. The researcher agrees with Selden and Selden that students need to assess student-generated arguments, but she would select different arguments in order to have an argument associated with each main code of the assessment tool.

Limitations of the Study

This study involved a very small number of volunteer participants. Both because of the small sample size and because the participants volunteered and were allowed to choose their participation level (seminar or comparison), no definitive conclusions can be drawn about what caused the seminar students' improvement nor the comparison students' relative stasis, and no generalizations beyond the study participants can be made. Given the extra commitment of coming to regular seminar sessions, it may be that the seminar groups appealed to more motivated students and that motivation and work ethic influenced the results. However, comparison student 5105 initially volunteered for the seminar group but the seminar conflicted with her other courses, and she could only participate in the comparison group, and her proof composition and validation did not improve. This is not enough to say that the motivation of students interested in the seminar group wasn't responsible for their improvement, but it is a detail that lends support to the researcher's hypothesis that carefully structured group work on intentionally designed problem sets could contribute to the acquisition and development of proof skills.

Since the pre-assessments and post-assessments were identical for all participants, the participants were already familiar with the content of the post-assessment because of their work on the pre-assessment. This change in relationship could have created a bias in the post-assessment results. Specifically, the composition portions of the both assessments were identical, so participants were more familiar with the claims on the post-assessment than they had been on the pre-assessment, and such familiarity may have led to the improved composition performances of the individual participants. However, since all participants took identical assessments, this does not explain why some participants did not demonstrate increased performance.

Furthermore, it is possible the seminar participants' relationships with the researcher affected their behavior and performance on the post-assessments. They worked closely with her during the seminar and may have persevered longer on their post-assessments because of the relationships they formed. Similarly, these relationships may have affected the participants' behavior during the seminar sessions. For example, Tammy stated,

> Because it was for your research study, I really tried to one, make sure that other people were understanding and two, make sure that I understood as opposed to maybe in a class structure where I would have grinned and nodded and moved along for the sake of time or the sake of whatever. And so, in that case I think [working as a member of a cooperative group] did increase my learning and was maybe the combination of the group and the setting. You know those things like if you're being watched you behave differently, and I know that in a class setting, I know I would have behaved differently.

So even if Cooperative Learning is at the root of the seminar participants' improvement, a course that employs CL may not lead to the same improvements.

The difference in assessment setting may have been problematic because the seminar participants were observed while working, and the comparison participants were not, even though the researcher was present during their assessments. Some research suggests that individuals behave differently when they know they are being observed, and this may have skewed the results.

Additionally, because of the different settings for assessments for seminar and comparison participants, the researcher was able to collect data on the seminar participants' use of different proving methods during the composition portion of the assessments, but was unable to do so for the comparison participants. Most of the seminar participants from both studies showed greater flexibility during their post-assessments than they had during the pre-assessments. They were more likely to attempt proofs by different methods and to put work on one item aside to work on a different item before returning to it. It is possible the comparison students had similar changes, but because of the study design, no data was collected indicating such behavior. In future studies, assessments will need to be conducted in similar settings for the different sets of participants.

Finally, one can't say for certain whether the changes made to the group work protocol were responsible for the improved cohesion in the groups in the implementation study. However, those groups worked better together than the groups in the previous study. There were almost no instances of participants working on their own without talking to their group mates, and there was much more vigorous conversation and debate.

Implications for Future Research

More questions.

This research has raised additional questions that are important to the understanding of student experience with proof.

- What was behind the seminar participants' improvement? Did the Cooperative Learning environment contribute significantly to that improvement? Was engaging with the problem sets responsible? Would working in a different Inquiry Based Learning or other active learning setting result in similar improvement?
- How are composition and validation skills related? Specifically, what skills are required for students to validate their own arguments, and how do they differ from the skills required to validate arguments produced by others?
- How are flexibility and proof expertise related? Is change in one necessarily tied to a change in the other? What attitudes/aptitudes is this linked to? What can educators do in the classroom to foster this flexibility? What role does specific content knowledge play in this mix?
- Are there identifiable content-independent proving skills? What are they? What implications will the determination and identification of these skills have on instruction of transition to proof classes?

Finding the answers to these questions could have a profound impact on classroom practices in proof-based courses and lead to better preparation of undergraduate math majors for advanced study of mathematics.

CL and proof composition.

Proof is an important part of advanced undergraduate mathematics curriculum, and most students are expected to compose proofs in the course of their studies. However, it has been established that many undergraduate students struggle to understand and produce proofs, even at the end of their undergraduate studies (Sowder & Harel, 2003). While

Inquiry Based Learning, Cooperative Learning, and other alternative classroom models for teaching proof have been proposed (e.g. Levine & Shanfelder, 2011; McLoughlin, 2010), little research has been undertaken on the general efficacy of such models. That the results presented here are consistent with the hypothesis that working on carefully designed problem sets in a Cooperative Learning environment could be beneficial to the development and acquisition of proof construction skills. This motivates the need for a large scale study into the causes of the improvement.

Large-scale study of CL environments and proof would likely entail studying transition to proof classes taught in a CL style and compared to classes taught with the same materials in a traditional lecture environment. In such studies, more data would need to be gathered about the group dynamics to ensure the goals of CL were being met, and the groups of students taught by different methods would need to be assessed in similar settings.

Flexibility and proof composition.

These studies were not designed to examine the relationship between students' tendency to switch proof methods and their proving expertise, but the researcher observed much greater flexibility in regards to proof method use on most of the seminar participants' post-assessments than had been demonstrated on the pre-assessments. This leads to questions about what caused the switching increase in the participants and how that flexibility was related to their ability to compose proofs.

Since it is not the case that working mathematicians always exhibit low flexibility, research into when and why mathematicians switch proof methods could shed light on how and when switching methods improves composition success. Additionally, the results from such research as well as from studying how advanced provers determine what proving methods are likely to be productive could inform instructors how to better guide novice provers to productive proving methods.

Proof validation and proof comprehension.

This researcher speculates that the current discussion of validation skills needs to be split into a conversation about the skills needed to validate one's own arguments and another about the skills needed to validate arguments produced by others. Research needs to be done to investigate the separate skill sets and test their interconnectivity. This researcher is currently exploring the usability of the Argument Assessment Tool presented in this dissertation (see Appendix 3b), as a way to assess compositions, as a classroom tool for student evaluation of each other's arguments.

6772's identification and concurrent use of the converse, as well as the difference in Bill's ability to compose arguments before and after engaging in the validation exercise, leads to questions regarding how peer grading could affect individuals' proof production, and answering those questions could also lead to a difference in classroom practices.

In studying validation skills, it would be beneficial for the researcher to ask students or study participants to identify all errors and concerns instead of focusing on classification of the arguments. This researcher also recommends that composition and validation not be studied simultaneously unless their relationship is being studied.

Transferrable, content-independent proof composition skills.

Content-independent proof composition skills are generally assumed to exist in the current research literature, but their existence and identification has not been established. This researcher plans to explore these questions as the continuation of the research presented here. Through interviews with mathematicians and math educators, she plans to determine a list of skills that are seen as common to a variety of content areas such as number theory, abstract algebra, real analysis, and topology. She will then develop claims in each of those content areas that necessitate the use of the determined skills and see if students are able to employ the skills across the different content areas. For example, participants might be asked to prove claims in each of the content areas that required constructing an indirect proof. Their content knowledge of each domain would be assessed separately in order to tease apart participants' ability to transfer the skills and their ability to prove in that content area in general. The researcher would then use an adapted version of the assessment model for proof comprehension developed by Mejia-Ramos et al. (2011) to assess whether or not the participants were able to identify the commonalities across content areas. Participants in such a study would most likely be advanced undergraduate math majors or graduate students in mathematics, as they would need to have prior experience with the different content areas.

Implications for Teaching

The results of these studies are consistent with the researcher's hypothesis that working in a Cooperative Learning environment on carefully selected and developed materials is beneficial to students' acquisition and development of proof skills; she will be using principles of Cooperative Learning to guide her classroom practices going forward. She is currently teaching a transition to proof class and has her students working in cooperative groups while she acts as a facilitator of learning. She is also using the Argument Assessment Tool presented here (see Appendix 3b) to have students review and critique each other's work to help foster their ability to validate and comprehend arguments written by their peers. As she continues to develop classes for her future teaching career, she will reflect on the results of this study and her future research to determine the best practices for her classroom.

Appendix 1: Background Questionnaire

1. Gender: | M | F | **2.** Cumulative GPA: ________

3. Classification: | Freshman | Sophomore | Junior | Senior | Other: ________

4. Major(s): ________

5. Minor(s): ________

6. Indicate the proof-based courses you have COMPLETED, and list the grade recieved:

- ◯ MATH 306 - College Geometry: ________
- ◯ MATH 319 - Theory of Numbers: ________
- ◯ MATH 322 - Modern Algebra I: ________
- ◯ MATH 327 - Intro to Mathematical Thinking and Discrete Structures: ________
- ◯ MATH 401 - Advanced Calculus I: ________
- ◯ MATH 402 - Advanced Calculus II: ________
- ◯ MATH 421 - Modern Algebra II: ________
- ◯ MATH 422 - Modern Algebra for Engineers: ________
- ◯ MATH 431 - Introduction to Topology: ________
- ◯ Other: ________

7. Indicate the proof-based courses you are CURRENTLY TAKING:

- ◯ MATH 306 - College Geometry
- ◯ MATH 319 - Theory of Numbers
- ◯ MATH 322 - Modern Algebra I
- ◯ MATH 327 - Intro to Mathematical Thinking and Discrete Structures
- ◯ MATH 401 - Advanced Calculus I
- ◯ MATH 402 - Advanced Calculus II
- ◯ MATH 421 - Modern Algebra II
- ◯ MATH 422 - Modern Algebra for Engineers
- ◯ MATH 431 - Introduction to Topology
- ◯ Other: ________

Appendix 2a: Pilot Study Attitudes/Beliefs Survey

Using the scale:
1- strongly disagree, 2- disagree, 3- neutral, 4- agree, 5- strongly agree
Please rate the following statements:

A1. Mathematics is computation.	1	2	3	4	5
A2. Mathematics problems given to students should be quickly solvable in a few steps.	1	2	3	4	5
A3. Mathematics is the dynamic searching for order and pattern in the learner's environment.	1	2	3	4	5
A4. Mathematics is a beautiful, creative and useful human endeavor that is both a way of knowing and a way of thinking.	1	2	3	4	5
A5. Right answers are much more important in mathematics than the ways in which you get them.	1	2	3	4	5

Using the scale:
1- strongly disagree, 2- disagree, 3- neutral, 4- agree, 5- strongly agree
Please rate the following statements:

B1. A proof in mathematics is different from other kinds of proof.	1	2	3	4	5
B2. A proof in mathematics both verifies and explains.	1	2	3	4	5
B3. Examples illustrating a result do not always help me understand why the result is true.	1	2	3	4	5
B4. Proof is essential in pure mathematics.	1	2	3	4	5
B5. In mathematics evidence from examples tells you what it true.	1	2	3	4	5
B6. I can't see the point of doing proofs: all the results I encounter have already been proved beyond doubt by famous mathematicians.	1	2	3	4	5
B7. Proofs sometimes involve strategies that are not at all obvious.	1	2	3	4	5
B8. I like doing proofs in mathematics.	1	2	3	4	5
B9. I am not confident in my ability to prove results for myself.	1	2	3	4	5
B10. Working through a proof of a result in a textbook helps me to understand why it is true.	1	2	3	4	5
B11. Different proofs of a theorem help me to understand it better.	1	2	3	4	5
B12. A proof in mathematics depends on other mathematical results.	1	2	3	4	5
B13. Even if a result in mathematics is proved, I can't be certain that it is true.	1	2	3	4	5
B14. It is harder to prove than to disprove.	1	2	3	4	5

Appendix 2b: Pilot Study Attitudes/Beliefs Survey Results

Seminar Participants

	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14
Bill - pre	1	1	4	5	2	4	4	3	5	2	1	5	4	3	4	4	5	2	5
Bill - post	2	2	4	5	2	4	4	2	5	2	2	4	3	3	4	4	5	2	5
Ingrid - pre	2	1	5	5	2	3	4	4	5	4	1	5	5	4	5	5	5	1	5
Ingrid - post	2	1	4	5	2	4	4	4	5	3	1	5	5	3	4	3	3	2	5
Ivan - pre	1	2	4	5	3	5	5	2	4	3	2	5	2	3	3	4	5	1	3
Ivan - post	1	1	5	5	2	5	5	2	5	5	1	5	2	1	4	4	5	1	5
Nathan - pre	1	3	5	5	3	5	4	2	5	3	1	5	5	5	5	5	5	1	5
Nathan - post	1	2	4	5	3	5	4	3	5	1	1	5	5	1	5	5	4	1	5
Omar - pre	3	5	3	5	3	4	3	1	5	2	1	5	3	4	3	4	4	3	3
Omar - post	2	1	4	5	2	5	5	4	5	1	1	5	4	4	5	4	4	4	3
Ursula - pre	4	2	5	4	3	4	3	4	5	2	5	4	3	5	3	4	3	4	5
Ursula - post	4	3	5	4	4	4	4	4	4	2	4	4	3	4	2	4	4	4	4
Zach - pre	4	2	5	5	2	5	5	2	5	4	1	4	4	2	5	5	5	1	2
Zach - post	4	1	5	5	1	4	2	4	5	4	1	4	2	2	4	2	4	1	2

1 - strongly disagree **2 - disagree** **3 - neutral** **4 - agree** **5 - strongly agree**

Comparison Participants

	A1	A2	A3	A4	A5	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14
6772 - pre	3	2	3	5	4	1	5	4	5	2	2	5	3	4	2	4	2	1	5
6772 - post	1	1	4	5	1	1	5	5	5	1	4	4	3	2	4	5	5	1	5
0296 - pre	1	1	3	5	2	5	5	3	5	1	1	3	5	2	5	4	3	1	2
0296 - post	1	1	3	5	2	5	5	3	5	1	1	3	5	2	5	4	3	1	2
4586 - pre	1	3	5	5	4	4	4	4	5	2	1	5	5	4	4	3	5	1	3
4586 - post	2	2	5	5	3	4	4	5	5	3	1	5	5	3	3	4	5	1	1

1 - strongly disagree **2 - disagree** **3 - neutral** **4 - agree** **5 - strongly agree**

Appendix 2c: Implementation Study Attitudes/Beliefs Survey Questions

1. How does mathematical proof differ from other kinds of proof?

2. What is the purpose of writing proofs of theorems that are already known to be true?

3. Once you have seen a rigorous proof of a theorem, how confident are you that the theorem is true?

4. Why does empirical evidence not count as proof?

5. Do you prefer proving or disproving claims? Why?

6. What do you like/dislike about writing proofs?

7. How confident are you in your ability to construct proofs?

8. What are the challenges you struggle with when constructing proofs?

Appendix 3a: Composition Items and Proofs

Item C1. Prove that if m^2 is odd, then m is odd.

Proof:
Assume that m is even.
Then $m = 2k$ for some $k \in \mathbf{Z}$.
Thus, $m^2 = (2k)^2 = 4k^2 = 2(2k^2)$.
Since $2k^2 \in \mathbf{Z}$, m^2 is also even.
Therefore, if m^2 is odd, m must also be odd.

Item C2. Prove that if n is a natural number, then $n^3 - n$ is divisible by 6.

(Direct Proof)
Let $n \in \mathbf{N}$. Then $n^3 - n = n(n-1)(n+1)$. To show this is divisible by 6, it suffices to show that it is divisible by 2 and by 3.

Divisibility by 2 :

Case 1 : n is even
Suppose $n = 2k$ for some $k \in \mathbf{N}$.
Then $n^3 - n = n(n-1)(n+1) = 2k(n-1)(n+1)$ which is divisible by 2.

Case 2 : n is odd
Suppose $n = 2k+1$ for some $k \in \mathbf{N}$.
Then $n + 1 = 2k + 1 + 1 = 2k + 2 = 2(k+1)$ which is even. Thus $n^3 - n = n(n-1)(n+1) = n(n-1)(2(k+1)) = 2(n(n-1)(k+1))$ which is divisible by 2.

Thus $n^3 - n$ is always divisible by 2.

Divisibility by 3 :

Case 1 : n is divisible by 3.
If $n = 3k$ for some $k \in \mathbf{N}$, then $n^3 - n = n(n-1)(n+1) = 3k(n-1)(n+1)$ which is also divisible by 3.

Case 2 : n is not divisible by 3.
If n is not divisible by 3, then either $n = 3k+1$, or $n = 3k+2$ for some $k \in \mathbf{N}$.

Case 2a : Suppose $n = 3k + 1$.
Then $n - 1 = 3k + 1 - 1 = 3k$, and
$n^3 - n = n(n-1)(n+1) = n(3k)(n+1) = 3(n * k(n+1))$ which is divisible by 3.

Case 2b : Suppose $n = 3k + 2$.
Then $n + 1 = 3k + 2 + 1 = 3k + 3 = 3(k+1)$, and
$n^3 - n = n(n-1)(n+1) = n(n-1)(3(k+1)) = 3(n(n-1)(k+1))$ which is also divisible by 3.

Therefore, whether or not n is divisible by 3, $n^3 - n$ always is.

Since $n^3 - n$ is always divisible by both 2 and 3, $n^3 - n$ is always divisible by 6.

(Proof by Induction)

Base Case : $n = 1$
$1^3 - 1 = 0$ which is divisible by 6.

Inductive Step :
Suppose $k^3 - k$ is divisible by 6 for some $k \in \mathbf{N}$.
That is $k^3 - k = 6m$ for some $m \in \mathbf{N}$.
$(k+1)^3 - (k+1) = k^3 + 3k^2 + 3k + 1 - k - 1 = (k^3 - k) + 3(k^2 + k) = 6m + 3(k^2 + k)$.

Case 1 : k is even.
Let $k = 2l$ for some $l \in \mathbf{N}$.
Then $k^2 + k = 4l^2 + 2l = 2(2l^2 + l)$, and $3(k^2 + k) = 3 * 2(2l^2 + l) = 6(2l^2 + l)$.
In this case, $(k+1)^3 - (k+1) = 6m + 6(2l^2 + l) = 6(m + (2l^2 + l))$ which is divisible by 6.

Case 2 : k is odd. Let $k = 2l + 1$ for some $l \in \mathbf{N}$.
Then $k^2 + k = 4l^2 + 4l + 1 + 2l + 1 = 2(2l^2 + 3l + l)$, and
$3(k^2 + k) = 3 * 2(2l^2 + 3l + l) = 6(2l^2 + 3l + l)$.
In this case, $(k+1)^3 - (k+1) = 6m + 6(2l^2 + 3l + l) = 6(m + (2l^2 + 3l + l))$ which is also divisible by 6.

Therefore, $(k+1)^3 - (k+1)$ is divisible by 6 whenever $k^3 - k$ is.

Thus, for any natural number, n, $n^3 - n$ is divisible by 6.

Item C3. A *triangular number* is defined as a natural number that can be written as the sum of consecutive integers, starting with 1 (pilot study definition).

A *triangular number* is defined as a natural number that can be written as the sum of all positive integers less than or equal to a given positive integer, k (implementation study definition).

Prove that a number, n, is triangular if and only if $8n+1$ is a perfect square.
(You may use the fact that $1+2+...+k=\frac{k(k+1)}{2}$.)

Proof:

$\rightarrow$:

Let n be a triangular number.
By the hint, $n=\frac{k(k+1)}{2}$ for some $k \in \mathbf{N}$.
Therefore, $8n+1=8\frac{k(k+1)}{2}+1=4(k(k+1))+1=4k^2+4k+1=(2k+1)^2$ which is a perfect square.

$\leftarrow$:

Suppose $8n+1$ is a perfect square, that is, let $8n+1=m^2$ where $m \in \mathbf{N}$.
$8n+1=2(4n)+1$, so $8n+1=m^2$ is odd.
We already know (from C1) that if m^2 is odd, then m is as well.
Thus, $\exists k \in \mathbf{N}$ such that $m=2k+1$, and $8n+1=(2k+1)^2$.
Solving for n, we get $n=\frac{(2k+1)^2-1}{8}=\frac{4k^2+4k}{8}=\frac{k(k+1)}{2}$ which is the form of a triangular number, so n is triangular.

Therefore, a number, n, is triangular if and only if $8n+1$ is a perfect square.

Appendix 3b: Argument Assessment Tool

<table>
<tr><th>Argument Codes</th><th>Code Details</th><th>Code Evidence</th></tr>
<tr><td>Incoherent or not addressing the stated problem (A0)</td><td>1. Solution shows a misunderstanding of the mathematical content.
2. Ignores the question completely.
3. Interprets claim, provides no argument.</td><td>• List A0 and either 1, 2, or 3.</td></tr>
<tr><td>Empirical (example based) (A1)</td><td>1. Examples are used to find a pattern, but a generalization is not reached.
2. Only examples are generated as a complete solution.</td><td>• List A1 and either 1 or 2</td></tr>
<tr><td>Unsuccessful attempt at a general argument (A2)</td><td>1. There is a major mathematical error
2. Illogical reasoning; several holes and or errors exist causing an unclear or inaccurate argument.
3. Reaches a generalization from examples, but does not justify why it is true for all cases.
4. Solution fails to covers all cases.
5. Solution is incomplete. Argument stops short of generalizing the stated claim.</td><td>• List A2 and match the bulleted number (1-5) in the middle column with the work in the solution.</td></tr>
<tr><td>Valid argument but not a proof (A3)</td><td>1. The solution assumes claims, in other words the solution exhibits a leap of faith before reaching a conclusion
2. The solution assumes a conjecture or lists a non-mathematical statement as a conjecture.
3. Argument is sound, but does not use mathematical notation and/or language - too informal</td><td>• List A3 and either 1, 2 or 3 & address each of the points below **</td></tr>
<tr><td>Proof (A4)</td><td></td><td>• List A4 and address each of the three clear and convincing points below. **</td></tr>
<tr><td colspan="3">** for use with A3 and A4.
(+/-) The flow of the argument is coherent since it is supported with a combination of pictures, diagrams, symbols, or language to help the reader make sense of the author's thinking. Diagrams are fine as long as they are accompanied by an explanation. Explanation of ideas or patterns.
(+/-) There are no irrelevant or distracting points. Variables and definitions are clearly defined and any terms introduced by the author are explained. Common understood language
(+/-) The conclusion is clearly stated.</td></tr>
</table>

Appendix 3c: Pilot Study Proof Composition Results

Seminar Participants

	PRE			POST		
Student	**Item**	**Subclaims**	**CODE**	**Item**	**Subclaims**	**CODE**
Bill	**C1**		**A2.1.A,B**	**C1**		**A4.+++.A,B**
	C2		**A2.5.A**	**C2**		**A2.5.A**
		Div by 2	A4.+++		Div by 2	A4.+++
		Div by 3	A0.2		Div by 3	A0.2
	C3		**A2.5.B**	**C3**		**A3.2.+++.A,B**
		T->S	A0.2		T->S	A4.+++
		S->T	A2.5		S->T	A3.2.+++
Ingrid	**C1**		**A2.1,2.B**	**C1**		**A4.+-+.A.B**
	C2		**A2.1,5.A**	**C2**		**A2.2,5.A**
		Div by 2	A2.3.5		Div by 2	
		Div by 3	A0.1,2		Div by 3	
	C3		**A2.5.A,B**	**C3**		**A3.2.++-.A,B**
		T->S	A0.2		T->S	A4.-+-
		S->T	A0.2		S->T	A3.2.++-
Ivan	**C1**		**A2.2.N**	**C1**		**A2.2.N**
	C2		**A4.+++.A**	**C2**		**A2.1.A**
		Div by 2	A4.+++		Div by 2	A2.1
		Div by 3	A4.+++		Div by 3	A3.+++
	C3		**A2.2,3.A**	**C3**		**A2.5.A,B**
		T->S	A0.2		T->S	A4.-++
		S->T	A0.2		S->T	A0.2
Nathan	**C1**		**A4.---.A,B**	**C1**		**A4.-++.A,B**
	C2		**A2.5.N**	**C2**		**A2.2,4.N**
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		**A2.5.A,B**	**C3**		**A4.-+-.A,B,C**
		T->S	A4.---		T->S	A4.---
		S->T	A2.5		S->T	A4.---
Omar	**C1**		**A2.2.N**	**C1**		**A4.+++.A,B**
	C2		**A0.1.N**	**C2**		**A2.2,5.N**
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		**A0.2.N**	**C3**		**A2.5.A**
		T->S	A0.2		T->S	A2.5
		S->T	A0.2		S->T	A0.2
Ursula	**C1**		**A2.2.A,B**	**C1**		**A4.-+-.A,B**
	C2		**A2.1,2,5.N**	**C2**		**A2.5.A**
		Div by 2			Div by 2	A0.2
		Div by 3			Div by 3	A3.3.-+-
	C3		**A2.1,2.A,B**	**C3**		**A2.2,5.A,B**
		T->S	A2.1,2		T->S	A2.1
		S->T	A0.2		S->T	A0.2
Zach	**C1**		**A4.++-.A,B**	**C1**		**A2.2.B**
	C2		**A2.1,5.A**	**C2**		**A2.5.A**
		Div by 2			Div by 2	A0.2
		Div by 3			Div by 3	A0.2
	C3		**A2.1,5.N**	**C3**		**A2.2,3.N**
		T->S	A0.2		T->S	A0.2
		S->T	A0.2		S->T	A0.2

Comparison Participants

	PRE			POST		
Student	Item	Subclaims	CODE	Item	Subclaims	CODE
0296	C1		A4.+++.A,B	C1		A4.++-.A,B
	C2		A2.5.N	C2		A2.5.N
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		A0.2.N	C3		A0.2.N
		T->S			T->S	
		S->T			S->T	
4586	C1		A4.+++.A	C1		A4.+++.A
	C2		A3.1.N	C2		A3.1.N
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		A2.5.A,B	C3		A2.5.A,B
		T->S	A4.+++		T->S	A4.+++
		S->T	A2.5		S->T	A2.5
6772	C1		A2.2.N	C1		A2.2.N
	C2		A2.1.N	C2		A2.1,5.N
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		A3.2+++.A,B	C3		A2.5.A,B
		T->S	A4.+++		T->S	A4.+++
		S->T	A3.2+++		S->T	A0.2

Appendix 3d: Implementation Study Proof Composition Results

Seminar Participants

	PRE			POST		
Student	Problem	Subclaims	CODE	Problem	Subclaims	CODE
Ethan	**C1**		**A2.2,5.N**	**C1**		**A4.-+-.A,B**
	C2		**A2.5.N**	**C2**		**A2.5.N**
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		**A1.1.N**	**C3**		**A1.1.N**
		T->S	A1.1		T->S	A1.1
		S->T	A0.2		S->T	A0.2
Greg	**C1**		**A2.1,2.N**	**C1**		**A4.+-+.A**
	C2		**A2.2,4.N**	**C2**		**A2.2.N**
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		**A2.5.A**	**C3**		**A3.2.-+-.A,B**
		T->S	A4.---		T->S	A4.-++
		S->T	A0.2		S->T	A3.2.-+-
Nadia	**C1**		**A3.3.--+.B**	**C1**		**A4.+++.A,B**
	C2		**A2.5.A**	**C2**		**A2.5.A**
		Div by 2	A3.3.-+-		Div by 2	A4.-+-
		Div by 3	A0.2		Div by 3	A0.2
	C3		**A2.1,5.A**	**C3**		**A2.5.A,B**
		T->S	A2.1,5		T->S	A4.-++
		S->T	A0.2		S->T	A2.5
Nick	**C1**		**A0.2.N**	**C1**		**A2.2,5.N**
	C2		**A0.2.N**	**C2**		**A0.2.N**
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		**A0.2.N**	**C3**		**A2.5.B**
		T->S	A0.2		T->S	A2.5
		S->T	A0.2		S->T	A2.5
Tammy	**C1**		**A0.2.A**	**C1**		**A4.-++.A**
	C2		**A0.2.N**	**C2**		**A2.5.N**
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		**A0.2.B**	**C3**		**A2.5.A,B**
		T->S	A0.2		T->S	A4.++-
		S->T	A0.2		S->T	A2.5
Travis	**C1**		**A2.2.N**	**C1**		**A4.+--.A**
	C2		**A2.2,5.A**	**C2**		**A2.1,2,5.N**
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		**A2.5.A,B**	**C3**		**A2.2,5.A,B**
		T->S	A4.-+-		T->S	A2.2
		S->T	A0.2		S->T	A2.5
Usher	**C1**		**A4.+++.A,B**	**C1**		**A4.+++.A,B**
	C2		**A4.+++.A**	**C2**		**A4.+++.A**
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		**A2.5.A,B**	**C3**		**A2.5.A,B**
		T->S	A4.+++		T->S	A4.+++
		S->T	A2.5		S->T	A2.5

Comparison Paticipants

	PRE			POST		
Student	Problem	Subclaims	CODE	Problem	Subclaims	CODE
1865	**C1**		**A4.+++.A,B**	**C1**		**A4.+++.A,B**
	C2		**A3.3.++-.N**	**C2**		**A4.+++.N**
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		**A4.++-.A,B,C**	**C3**		**A4.+++.A,B,C**
		T->S	A4.+++		T->S	A4.+++
		S->T	A4.++-		S->T	A4.+++
3099	**C1**		**A2.1,2.A,B**	**C1**		**A4.+++.A,B**
	C2		**A2.3.N**	**C2**		**A2.1,5.N**
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		**A2.1,2.A,B**	**C3**		**A2.1,2,3.B**
		T->S	A2.1		T->S	A2.1,2
		S->T	A2.2.		S->T	A2.2,3
5105	**C1**		**A4.+++.A,B**	**C1**		**A4.+++.A,B**
	C2		**A2.1,5.N**	**C2**		**A2.1,5.N**
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		**A2.5.A,B**	**C3**		**A2.1,2.B**
		T->S	A4.-++		T->S	A2.1,2
		S->T	A0.2		S->T	A2.2
5635	**C1**		**A1.2.N**	**C1**		**A3.3.A,B**
	C2		**A1.2.N**	**C2**		**A0.3N**
		Div by 2			Div by 2	
		Div by 3			Div by 3	
	C3		**A1.1.N**	**C3**		**A0.3N**
		T->S			T->S	
		S->T			S->T	
6293	**C1**		**A2.2.N**	**C1**		**A2.2.N**
	C2		**A2.1,5.A**	**C2**		**A2.1.A**
		Div by 2	A3.3		Div by 2	A2.5
		Div by 3	A2.1,5		Div by 3	A2.1
	C3		**A0.2.N**	**C3**		**A2.2,5A**
		T->S	A0.2		T->S	A2.2
		S->T	A0.2		S->T	A0.2

Appendix 4a: Validation Items

Item EG, "Errors Galore"
Theorem: For any positive integer n, if n^2 is a multiple of 3, then n is a multiple of 3.

[1] Assume that n^2 is an odd positive integer that is divisible by 3.
[2] That is $n^2 = (3n+1)^2 = 9n^2 + 6n + 1 = 3n(n+2) + 1$.
[3] Therefore, n^2 is divisible by 3. Assume that n^2 is even and a multiple of 3.
[4] That is, $n^2 = (3n)^2 = 9n^2 = 3n(3n)$.
[5] Therefore, n^2 is a multiple of 3.
[6] If we factor $n^2 = 9n^2$, we get $3n(3n)$; which means that n is a multiple of 3.

Item RT, "The Real Thing"
Theorem: For any positive integer n, if n^2 is a multiple of 3, then n is a multiple of 3.

[1] Suppose to the contrary that n is not a multiple of 3.
[2] We will let $3k$ be a positive integer that is a multiple of 3, so that $3k+1$ and $3k+2$ are integers that are not multiples of 3.
[3] Now $n^2 = (3k+1)^2 = 9k^2 + 6k + 1 = 3(3k^2 + 2k) + 1$. [4] Since $3(3k^2 + 2k)$ is a multiple of 3, $3(3k^2 + 2k) + 1$ is not.
[5] Now we will do the other possibility, $3k+2$.
[6] So, $n^2 = (3k+2)^2 = 9k^2 + 12k + 4 = 3(3k^2 + 4k + 1) + 1$ is not a multiple of 3.
[7] Because n^2 is not a multiple of 3, we have a contradiction.

Item GP, "The Gap"
Theorem: For any positive integer n, if n^2 is a multiple of 3, then n is a multiple of 3.

[1] Then $3|n^2$.
[2] Since $n^2 = 3x$, $nn = 3x$.
[3] Thus, $3|n$.
[4] Therefore if n^2 is a multiple of 3, then n is a multiple of 3.

Item CV, "The Converse"
Theorem: For any positive integer n, if n^2 is a multiple of 3, then n is a multiple of 3.

[1] Let n be a positive integer such that n^2 is a multiple of 3.
[2] Then $n = 3m$ where $m \in \mathbf{Z}^+$.
[3] So $n^2 = (3m)^2 = 9m^2 = 3(3m^2)$.
[4] This breaks down into $3m$ times $3m$ which shows that m is a multiple of 3.

Appendix 4b: Pilot Study Validation Results

Seminar Participants

Numerical Classification

	Bill		Ingrid		Ivan		Nathan		Omar		Ursula		Zach	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Errors Galore	**4**	**4**	3	**4**	2	**4**	**4**	**4**	3	**4**	3	4	**4**	**4**
The Real Thing	**2**	**3**	**3**	**3**	**2**	**2**	2	**1**	**1**	**1**	**1**	5	**1**	**1**
The Gap	**3**	**4**	**3**	2	**3**	**3**	5	1	**4**	**3**	**4**	**4**	**3**	1
The Converse	**4**	**4**	3	4,5	**4**	4	**4**	**4**	5	1	**4**	**4**	**4**	1

Classifications in **BOLD** were determined to be correct.

Specific Error Identification

	Bill		Ingrid		Ivan		Nathan		Omar		Ursula		Zach	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
Errors Galore														
using n to stand for distinct values							Y	Y					Y	Y
erroneous definitions	Y	Y				Y				Y			Y	Y
working with *n* instead of *n*2								Y						
factoring			Y								Y			
concluding n^2 div by 3 which was assumed	Y	Y		Y			Y						Y	
conclusion n div by 3 not supported														
The Real Thing														
unstated assumption		Y												
contrap not contrad				Y									Y	Y
The Gap														
n positive														
missing justification	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y		Y	Y
The Converse														
assuming conclusion	Y	Y			Y		Y	Y					Y	
conclusion that m div by 3			Y						Y					

Comparison Participants
Numerical Classification

	0296		4586		6772	
	Pre	Post	Pre	Post	Pre	Post
Errors Galore	**4**	5	**4**	**4**	3	2
The Real Thing	**1**	**1**	**1**	4	2	**1**
The Gap	1	**4**	**4**	**3**	**3**	3
The Converse	**4**	**4**	**2**	**4**	**2**	**2**

Classifications in **BOLD** were determined to be correct.

Specific Error Identification

	Bill		Ingrid		Ivan	
	Pre	Post	Pre	Post	Pre	Post
Errors Galore						
using n to stand for distinct values				Y		
erroneous definitions	Y		Y	Y		
working with *n* instead of *n*2		Y				
factoring					Y	
concluding n^2 div by 3 which was assumed						
conclusion n div by 3 not supported		Y				
The Real Thing						
unstated assumption						
contrap not contrad	Y	Y				
The Gap						
n positive						
missing justification	Y	Y	Y	Y		Y
The Converse						
assuming conclusion		Y	Y	Y		Y
conclusion that m div by 3	Y					

Appendix 4c: Implementation Study Validation Results

Seminar Participants
Numerical Classification

	Ethan		Greg**		Nadia		Nick		Tammy		Travis**		Usher	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
The Converse	3	2	**4**	**3**	**2**	3	1	1	**4**	**4**	**4**	3	**2**	**4**
The Real Thing	**3**	**3**	**1**	**1**	**2**	**3**	**1**	**2**	**3**	**1,3**	**1**	**1**	**3**	**3**
Errors Galore Te	**4**	**4**	5	**4**	**4**	**4**	1	4	5	**4**	5	3	**4**	**4**
The Gap	**3**	**3**	**4**	**3**	4	**3**	**3**	**3**	**3**	**4**	**4**	1	**3**	**1**

Classifications in **BOLD** were determined to be correct.

**No audio results were recorded for the post-assessment validation exercise and no justifications were recorded.

Specific Error Identification

	Ethan		Greg		Nadia		Nick		Tammy		Travis		Usher	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
The Converse														
assuming conclusion	Y	Y			Y				Y	Y			Y	Y
conclusion that m div by 3			Y			Y		Y	Y		Y			Y
The Real Thing														
unstated assumption						Y			Y	Y				
contrap not contrad	Y	Y						Y					Y	
Errors Galore														
overuse of n			Y			Y							Y	Y
erroneous definitions	Y	Y			Y	Y		Y	Y	Y			Y	
using n instead of n^2														
factoring						Y								
concluding n^2 div by 3 which was assumed					Y									
conclusion n div by 3 not supported														Y
The Gap														
n positive										Y				
missing justification	Y	Y	Y		Y	Y	Y	Y	Y	Y	Y		Y	Y

Comparison Participants
Numerical Classification

	1865		3099		5105		5635		6293	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
The Converse	**4**	**4**	3	**4**	**4**	**2**	1	1	4,3	3
The Real Thing	**1**	**1**	**2**	**1**	**1**	**1**	**1**	2	**3**	**2**
Errors Galore	5	**4**	2	**4**	5	**4**	2	1	5	3
The Gap	**3**	**4**	**4**	**3**	1	1	**3**	**3**	5	5,3

Classifications in **BOLD** were determined to be correct.

Specific Error Identification

	1865		3099		5105		5635		6293	
	Pre	Post	Pre	Post	Pre	Post	Pre	Post	Pre	Post
The Converse										
assuming conclusion	Y	Y		Y		Y				Y
conclusion that m div by 3			Y						Y	
The Real Thing										
unstated assumption					Y					
contrap not contrad	Y	Y	Y							
Errors Galore										
overuse of n		Y				Y			Y	Y
erroneous definitions		Y			Y					Y
using n instead of n^2				Y						
factoring										
concluding n^2 div by 3 which was assumed		Y								
conclusion n div by 3 not supported										
The Gap										
n positive										
missing justification	Y	Y		Y		Y	Y	Y	Y	Y

Appendix 4d: Nadia's Post-Assessment Validation Transcript

PI: You have 4 attempted proofs of the one theorem. So they're all trying to establish the same thing, and then as you read each argument, it's up to you to decide whether the argument is a rigorous proof of the claim, a rigorous proof of a different claim, and if you choose that, you have to tell me what claim it is the author is proving. Ok, your other options are that the argument is a non-rigorous proof of the claim meaning that any errors are minor enough that they don't affect the validity of the argument, but there may be assumptions or justifications missing, or that it doesn't meet the standards of a proof. If you don't understand what the author is trying to do, you can choose 5 and not classify it.

NADIA: Ok. So the theorem is for any positive integer n, if n squared is a multiple of 3, then n is a multiple of 3. So let n be a positive integer. This is an implication. If n squared is a multiple of 3, then n is a multiple of 3. Let n be a positive integer such that n squared is a multiple of 3. Cool. So that is like our p implies thingy. Then n equals 3m where m belongs to our positive integers. Uh, yes. I can see that. Because if uh n is a positive integer, then m has to be a positive integer as well. So m squared equals 3m squared uh huh. I can see that totally. Equals 9 m squared. Uh huh. Equals 3 times 3 m squared. I am following this logic. This breaks down into 3m times m. 3m times 3m sorry, which shows that m is a multiple of 3. Breaks down into 3m times 3m which shows that m is a multiple of 3. (pause) Hm. (pause) I'm just not understanding that last sentence. I was totally following until right there. This breaks down into 3m times 3m. So, are they talking about 3 times 3m times 3m? Which shows that m is a multiple of 3. I don't see how that shows m is a multiple of 3. (pause) And that's not what we were trying to show. We're trying to show that n is a multiple of 3. So it's either a rigorous proof of a different claim, or a non-rigorous proof. So let's see. Breaks down into 3m times 3m shows that m is a multiple of 3. I guess the question is does that show that m is a multiple of 3, because if it does, then it's a proof of a different claim. And if it doesn't, it's a non-rigorous proof. So (pause) I mean that would show that well, we already knew that n was 3m. I'm going to go with non-rigorous proof.

PI: Ok

NADIA: And my justification for that classification is that ... would be that while I totally followed the logic up to here like I said, all of this made complete sense and was like, seemed like a logical progression, and then at this point, first of all that's not very good wording because he's kind of vague, but breaks down into 3m times 3m, and I only saw that because like, ok because well, oh no, ok. 3m squared, hold up now. 3 times 3 times m times m. Ok, so I could see that being 3m times 3m um, so I can even follow up to this point, which shows that m is a multiple of 3, but unless we know what m equals on the other side, I'm not seeing how m; just because something is 3 times m doesn't mean m is 3 times something.

PI: Ok

NADIA: That's where that breaks down for me.

PI: Ok. Thank you. Alright, uh, next one.

NADIA: Uh for any positive integer n, if n squared is a multiple of 3, then - so it's the same oh yeah. You told me already about that

PI: It's okay

NADIA: Same implication. Ok, suppose to the contrary, ok so this is a proof by contradiction, that n is not a multiple of 3. Ok. They should probably assume. Let 3k be a positive integer that is a multiple of 3. 3k positive integer, ok that's kind of obvious, yeah. 3k is a multiple of 3, so that 3k plus 1 and 3k plus 2 are integers that are not multiples of 3. Cool. Now n squared equals, they didn't say anything about n, so that might be a number 3 as well. Let's see. (pause) All I know is that n is not a multiple of 3. Ok, let's see. So now n squared equals 3k plus 1 squared. Why? (pause) uh, ok, they did two cases it looks like and n squared equals 3k plus 1 and 3k plus 2 where it's not a multiple of 3. Ok. Equals 9 k squared plus 6k plus 1. Equals, then they factored out a 3 from this part. Uh huh. Ok. Since 3 times 3 k squared plus 2k is a multiple of 3, that is not. Totally ok. Now they'll do the other possibility, 3k plus 2. So n squared equals, 3k plus 2 squared equals 9 k squared plus 12k plus 6 times 6 that's 12. k plus 4 equals - they factored out a 3 and got, they got 3 k squared plus 4k plus 1 and there's the other 1 is not a multiple of 3. Uh huh. Because n squared is not a multiple of 3, we have a contradiction. So here's (pause) where it seems inadequate. Because n squared is not a multiple of 3, that's not a contradiction, well here's a question. Is this theorem, am I allowed to know if it's true?

PI: It is true.

NADIA: Um, so they said, yeah, seems like a non-rigorous proof because we didn't say at the beginning that we're assuming that n squared is a multiple of 3 and n is not. (pause) um, so I don't really know, it's like I don't have that to tell me ok, it's saying that n squared is not a multiple of 3 even thought we assumed it is, so that's a contradiction. It's saying that n squared is not a multiple of 3 even though we assumed it was, so it's not a contradiction because we haven't assumed anything about n squared.

PI: Ok

NADIA: I would say 3 then.

PI: Alright. Next one.

NADIA: Ok, for any positive integer, same thing, assume that n squared is an odd positive integer (pause) Ok. I think they did it for a case in which it's odd and a case in which it's even. That is divisible by 3. That is n squared equals 3n plus 1 squared. Uh, is an integer divisible by 3. Ok. Equals 9 n squared plus 6n plus 1. Equals 3n times um, that doesn't seem right. 3n times n plus 2 plus. I don't think that arithmetic is right. That would be 3n squared plus 6n plus 1. Ok. So I don't think that is right. Um, therefore n squared is divisible by 3. (pause) Is divisible by 3, which doesn't really follow from that either because if n squared is equal to 3 times something plus 1, then it's not divisible by 3. Let's see n squared is even and a multiple of 3, we'll put a little x to say it did not follow. I didn't follow that either. See if n squared is even and a multiple of 3, that is n squared is 3n squared equals 9n squared is that right? Would that be? Is that the only possible case where it's even and a multiple of 3? I think they're making an assumption, for example this, is this necessarily odd? No, I well see, if you're squaring it, that's not even going to be odd. Yeah, that's not ok. Um, I would say 4. Do I have to finish going through it, or can I say from halfway through it?

PI: Unless you think they might be able to recover in the last two lines.

NADIA: I don't think so because I can kind of see what they were trying to do like if you have an odd, if you have an even, but that's not even what they did

PI: Ok

NADIA: And so it seems like that's not even what happened here. I keep going for the heck of it. Therefore n squared is a multiple of 3. So yeah, then they used n in here too. Oh I didn't even notice that, that's not okay either. They should use a different variable. Now if it was, like a k or something, just to mess with it. Then, k squared, 3k times 3k, yeah, so n or n squared is a multiple of 3, yeah. We factor n squared we get, yeah, which means that, see k, we know k is a multiple of 3, but we don't know that n is. Yeah, I'd say 4.

PI: Ok

NADIA: Is that, I mean for number 5, I don't understand the argument, so I kind of understand what they were trying to do, so

PI: Yeah

NADIA: Ok, cool.

PI: If you can follow what they're trying to do, you understand their attempted argument.

NADIA: Ok … Positive, yeah. Again. Let n be an integer such that n squared equals 3x where x is an integer as well. Ok. Then 3, what's that bar?

PI: Divides. Then 3 divides n squared. It's just another way of saying n squared is a multiple of 3

NADIA: Ok. Then since n squared equals 3x. Uh huh. N times n equals 3x. Thus n divides 3x. I mean 3 rather. I guess so. No wait, no not necessarily. Because in order to say that, you have to say n is of the form 3x over n. Well, I guess so because x over n then is just some number, and 3, do we know that's an integer though?

PI: When we say divisible by and multiple of, we're talking all integers.

NADIA: Yeah. So (pause) Oh and well x is an integer, and n is an integer, but we don't know if x over n is an integer. Right? Uh. Yeah, thus 3 divides n. 3 divides n, which means n is a multiple of 3?

PI: Uh huh

NADIA: Ok so n is a multiple of 3. I would say 3 because I'm not entirely convinced that this spot's correct. That n times n equals 3 x means that n is equal to 3 times an integer.

PI: Ok. Thank you.

Appendix 5: Exit Interview Questions

1. How has your confidence level about constructing proofs changed as a result of your participation in this research?

2. What are the factors you think most contributed to that change?

3. How has your confidence level about reading and understanding proofs changed as a result of your participation in this research?

4. What are the factors you think most contributed to that change?

5. How did working as a member of a cooperative group affect your learning?

6. Did you feel responsible at all for the learning of the other members of your group?

7. Did you feel accountable for your own work?

8. Was the group processing beneficial to the functioning of your group?

9. Do you feel better able to work cooperatively as a result of your participation in this research?

10.* What does proof mean to you?

11.* How has that changed as a result of your participation in this research?

* Questions 10 and 11 were only asked during the pilot study.

Appendix 6a: Problem Set on Inverses and Inverse Images

pilot study only
Inverses

1. Given $f : A \to B$, define the map $\mathcal{H} : B \to \{C \mid C \subset A\}$ by
$\mathcal{H}(b) = \{a \in A \mid f(a) = b\}$.
For each of the following functions f_i, describe $\mathcal{H}_i$ and determine whether or not $\mathcal{H}_i$ naturally gives rise to a function $\mathcal{F}_i : B \to A$ such that $\mathcal{F}_i \circ f(a) = a$.

a. f_1 : {UNM students} $\to$ {active banner ID numbers} where $f_1(x)$=x's banner ID number.

b. f_2 : {locations on the surface of the earth} $\to [-10916,\ 8850]^*$ where $f_2(x) = x$'s elevation in meters.
(* The highest point on Earth is the peak of Mt. Everest which lies 8,850m above sea level, and the lowest point on Earth is at the bottom of The Mariana Trench at a depth of 10,916m below sea level. http://geology.com)

c. f_3 : {convex planar polygons} $\to \mathbf{R}$ where $f_3(x) =$ the perimeter of x in cm.

d. f_4 : {convex planar polygons} $\to (0, \infty)$ where $f_4(x) =$ the area of x in in^2.

e. $f_5 : \{p \mid p$ is a polynomial with real coefficients$\} \to \mathbf{R}$ where $f(p) = p(0)$.

f. $f_6 : \mathbf{R}^3 \to \{ax^2 + bx + c \mid a, b, c \in \mathbf{R}\}$ where $f_6(a, b, c) = ax^2 + bx + c$

2. For the examples in Problem 1 in which $\mathcal{H}_i$ does NOT give us a function $\mathcal{F}_i : B \to A$ such that $\mathcal{F}_i \circ f(a) = a$, explain why not. Is it possible to adjust B so that we DO find such a $\mathcal{F}$? Why or why not?

3. Let $f : (0, \infty) \to$ {squares drawn in $\mathbf{R}^2$ centered at (0,0)} where $f(x) =$ the square with side length $\frac{x}{4}$ whose sides are parallel to the coordinate axes, and let g : {squares depicted in $\mathbf{R}^2$ centered at (0,0)} $\to (0, \infty)$ where $g(s) =$ the perimeter of s.
Show that:

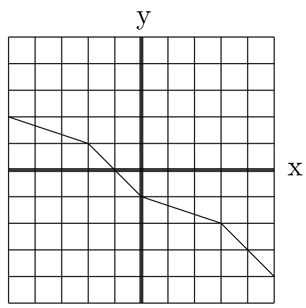
a. $g \circ f(x) = x$, and

b. find a suitable square, s, such that $f \circ g(s) \neq s$.

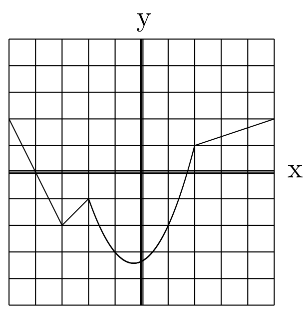
4. What needs to be true about f in order for $\mathcal{H}$ to lead us to a function $\mathcal{F} : B \to A$ such that $\mathcal{F} \circ f(a) = a$? Formalize this statement, and then prove it.

5. Given the graph of $f_i(x)$ below, plot $\mathcal{G} = \{(b, a) \mid b = f(a)\}$.

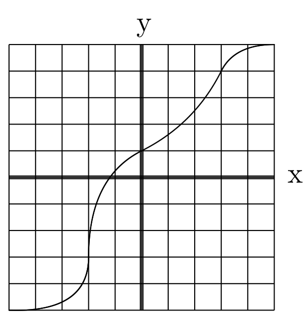
a.



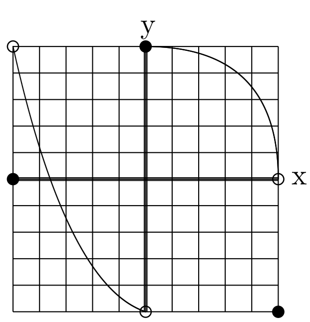
b.



c.



d.



e. How is the graph of f related to $\mathcal{G}$ in each of the above examples?

f. For which graphs is $\mathcal{G}$ the graph of a function on $[-5, 5]$?

g. What has to be true about about the graph of f for $\mathcal{G}$ to be the graph of a function on $[-5, 5]$?

6. When we can find $\mathcal{F}$, we call $\mathcal{F}$ the *inverse* of f. Prove that $f \circ \mathcal{F}(x) = x = \mathcal{F} \circ f(x)$ when $\mathcal{F}$ is the inverse of f

Appendix 6b: Problem set on the Gaussian Integers

pilot study
Pythagoras, Gauss and Norm

1. For $a, b \in \mathbf{Z}$, define $N(a+bi) = (a+bi)(a-bi)$.

- **a.** Explain why N satisfies the definition of a function, and restate its domain in set notation.
- **b.** Explicitly state an appropriate codomain of N.
- **c.** Why do we talk about "the" domain and "a" codomain?

2. Is N a one-to-one function? Prove it is or provide an example to demonstrate that it is not.

3. Prove or disprove that N is onto for the codomain you described in problem 1.

4. Show that any natural number of the form $4k+3$, where $k \in \mathbf{N}$, cannot be written as the sum of two squares.

5. Find all $a, b \in \mathbf{Z}$ such that

a.) $N(a+bi) = 1$ **b.**) $N(a+bi) = 5$ **c.**) $N(a+bi) = 13$

d.) $N(a+bi) = 65$ **e.**) $N(a+bi) = 6$ **f.**) $N(a+bi) = 35$

6. For $a, b, c, d \in \mathbf{Z}$ such that $N(a+bi) = 5$ and $N(c+di) = 13$, what do you notice about $N((a+bi)(c+di))$? Repeat this for $N(a+bi) = 5$ and $N(c+di) = 10$ (You do not have to test every combination of a,b,c, and d, but you should test at least four possibilities.)

7. Based on your work from Problem 5, make a conjecture about $N((a+bi)(c+di))$, and prove or disprove it. Ultimately, the goal is to come up with a true conjecture and proof.

8. How is $N((a+bi)^2)$ related to $N(a+bi)$?

9. Prove that if $n \in \mathbf{Z}$ and $n = (a+bi)(c+di)$, then $(a+bi) = \alpha\overline{(c+di)}$, where $\alpha \in \mathbf{R}$ and $\overline{(c+di)}$ is the complex conjugate of $(c+di)$.

6b: implementation study
Gaussian Integers and the Norm - Definitions

- **Z**: the *integers*, $\{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$.
- **i**: the square root of -1, so $i = \sqrt{-1}$, and $i^2 = -1$.
- **C**: the *complex numbers*, $\{a + bi \mid a, b \in \mathbf{R}\}$.
- $\overline{(a+bi)}$: the *complex conjugate of* $(a + bi)$, $\overline{(a+bi)} = (a - bi)$.
- $\mathbf{Z}[i]$: the *Gaussian Integers.* This is a subset of the complex numbers where a, b are both integers.
- One-to-one Functions: A function, $f : A \to B$, is called $one - to - one$ if every image of f comes from exactly one input, that is $a_1 \neq a_2 \Rightarrow f(a_1) \neq f(a_2)$, and $f(a_1) = f(a_2) \Rightarrow a_1 = a_2$. *one-to-one functions* are also called *injective functions.*
- Onto Function: A function, $f : A \to B$, is *onto* if every element of B is an image of the function, that is $\forall b \in B, \exists a \in A$ such that $f(a) = b$. *onto functions* are also called *surjective functions.*

6b: implementation study
Gaussian Integers and the Norm

1. For $a, b \in \mathbf{Z}$, define $N(a+bi) = (a+bi)(a-bi)$.

- **a.** Explain why N satisfies the definition of a function. Restate its domain in set notation.
- **b.** Explicitly state an appropriate codomain of N.
- **c.** Why do we talk about "the" domain and "a" codomain?

2. Is N a one-to-one function? Prove that it is or provide an example to demonstrate that it is not.

3. Prove or disprove that N is onto for the codomain you described in problem 1.

4. Claim: Any natural number of the form $4k+3$, where $k \in \mathbf{N}$, cannot be written as the sum of two squares.

- **a.** Show that $3, 7$ and 11 cannot be written as the sum of two squares.
- **b.** Show that the sum of two even squares is divisible by 4 and that the sum of two odd squares is even but NOT divisible by 4.
- **c.** Determine what other case needs to be checked to establish the claim, and write a complete proof of the claim.

5. Find all $a, b \in \mathbf{Z}$ such that

a.) $N(a+bi) = 1$ **b.**) $N(a+bi) = 5$ **c.**) $N(a+bi) = 13$

d.) $N(a+bi) = 65$ **e.**) $N(a+bi) = 6$ **f.**) $N(a+bi) = 10$

6. For $a, b, c, d \in \mathbf{Z}$ such that $N(a+bi) = 5$ and $N(c+di) = 13$, what do you notice about $N((a+bi)(c+di))$? Repeat this for $N(a+bi) = 5$ and $N(c+di) = 10$. (You do not have to test every combination of a,b,c, and d, but you should test at least four possibilities.)

7. Based on your work from Problem 6, make a conjecture about $N((a+bi)(c+di))$, and prove it is true.

9. Prove that if $(a+bi)(c+di) \in \mathbf{Z}$, then $(a+bi) = \alpha\overline{(c+di)}$, where $\alpha \in \mathbf{R}$.

Appendix 6c: Problem Set on Fixed Points and a Derivative-Like Function

pilot study
Derivatives and Fixed Points

For this problem set, let
$\mathcal{F} = \{f \mid f$ is a function with domain and codomain equal to $\mathbf{R}\}$.
$\mathcal{P} = \{p \in \mathcal{F} \mid p$ is a polynomial function with real coefficients$\}$, and
$D : \mathcal{P} \to \mathcal{F}$ where $D(p) = p'$, the derivative of p.

1. Use the definition of function to show that D is a function.

2. Evaluate $D(p_i)$ for the following p_i.

- **a.** $p_1(x) = 5$
- **b.** $p_2(x) = x^4 - 4x^2 + 4$
- **c.** $p_3(x) = \frac{x^5}{5!} + \frac{x^4}{4!} + \frac{x^3}{3!} + \frac{x^2}{2!} + x + 1$

3. Let $\mathcal{P}_d = \{p \in \mathcal{P} \mid \text{degree of } p \leq d\}$. Show that $\mathcal{D}$ maps $\mathcal{P}_d$ into $\mathcal{P}_{d-1}$.

- **a.** To prove the above statement, explain what the structures of the arguments would be if you were to prove it directly, by contradiction, and by contrapositive.
- **b.** Choose an argument, and prove the statement.

4. Categorize $\{p \in \mathcal{P} \mid D(p) = 0\}$. Prove you have a complete list.

For any function $f : A \to A$, x is a *fixed point* of f if $f(x) = x$.

5. For the following functions, f_i, find all the fixed points, or justify that the function has none.

- **a.** $f_2 : \mathbf{R} \to \mathbf{R}$ with $f(x) = x^3$
- **b.** $f_1 : \mathbf{R} \to \mathbf{R}$ with $f(x) = x^2 + 2$

6. Let $f : \mathbf{R} \to \mathbf{R}$ with $f(x) = x^4 + \frac{2}{3}x^3 - 3x^2 + \frac{4}{3}$. Determine how many fixed points f must have. You do not have to find the fixed points.

7. Let $f : \mathbf{R}^2 \to \mathbf{R}^2$ where $f(x, y) = (\frac{-3}{4}y, \frac{3}{4}x)$.

- **a.** Describe what the action of f is geometrically.
- **b.** Explain in geometric terms why f can't have more than one fixed point.
- **c.** Find the fixed point of f algebraically.
- **d.** What is the relationship between the distance from (x, y) to (α, β) and the distance from $f(x, y)$ to $f(\alpha, \beta)$?
- **e.** Use 7d to prove that f cannot have two fixed points.

8. $D : \mathcal{P} \to \mathcal{P}$, has only one fixed point.

- **a.** To prove the above statement, explain what the structures of the arguments would be if you were to prove it directly, by contradiction, and by contrapositive.
- **b.** Choose an argument, and prove the statement.

6c: implementation study
A Polynomial Function and Fixed Points - Definitions

- $\mathcal{P}$: the set of *real polynomials*, $\{p \mid p$ is a polynomial expression in x with real coefficients$\}$. Note that $p \in \mathcal{P}$ has the form $a_n x^n + ... + a_1 x^1 + a_0$, where $a_i \in \mathbf{R}$.

- $K : \mathcal{P} \to \mathcal{P}$ where $K(p) = K(a_n x^n + ... + a_1 x^1 + a_0) = n a_n x^{n-1} + ... + a_1 x^0 + 0 a_0$.

- Fixed Point: a *fixed point* of a function, $f : A \to A$, x is an element of A that doesn't change when f acts upon it. That is, x is a *fixed point* of f if $f(x) = x$.

6c: implementation study
A Polynomial Function and Fixed Points

1. Evaluate $K(p_i)$ for the following polynomials p_i.

a. $p_1 = 5$

b. $p_2 = x^4 - 4x^2 + 4$

c. $p_3 = \frac{x^5}{5!} + \frac{x^4}{4!} + \frac{x^3}{3!} + \frac{x^2}{2!} + x + 1$

3. Let $\mathcal{P}_d = \{p \in \mathcal{P} \mid \text{degree of } p \leq d\}$. Claim: $\mathcal{K}$ maps $\mathcal{P}_d$ into $\mathcal{P}_{d-1}$.

a. To prove the above claim, translate it into an implication statement,

b. explain what the structures of the arguments would be if you were to prove it directly, by contradiction, and by contrapositive, and

c. prove the statement using one of the arguments you outlined in b..

4. Find all $p \in \mathcal{P}$ such that $K(p) = 0$. Prove you have a complete list. [First write a biconditional statement, then prove both directions.]

5. For the following functions, f_i, find all the fixed points, or justify that the function has none.

a. $f_1 : \mathbf{R} \to \mathbf{R}$ with $f(x) = x^3$

b. $f_2 : \mathbf{R} \to \mathbf{R}$ with $f(x) = x^2 + 2$

6. Let $f : \mathbf{R} \to \mathbf{R}$ be the function given by $f(x) = x^4 + \frac{2}{3}x^3 - 3x^2 + \frac{4}{3}$. Determine how many fixed points f must have. You do not have to find the fixed points.

7. Let $f : \mathbf{R}^2 \to \mathbf{R}^2$ be the function given by $f(x, y) = (\frac{-3}{4}y, \frac{3}{4}x)$.

a. Describe the action of f on the plane geometrically.

b. Explain in geometric terms why f can't have more than one fixed point.

c. Find the fixed point of f algebraically.

d. What is the relationship between the distance from (x, y) to (α, β) and the distance from $f(x, y)$ to $f(\alpha, \beta)$?

e. Use 7d to prove that f cannot have two fixed points.

8. The function $K : \mathcal{P} \to \mathcal{P}$ has only one fixed point.

a. Translate the claim into a biconditional statement,

b. explain what the structure of the arguments would be if you were to prove it directly, by contradiction, and by contrapositive, and

c. prove the statement using one of the arguments you outlined in b.

Appendix 6d: Problem Set on Injectivity and Surjectivity

pilot study
-jectivity

1. For a subset $S \subset \mathbf{R}^2$, define $\pi_1 : S \to \mathbf{R}$ by $\pi_1((x, y)) = x$ and $\pi_2 : S_i \to \mathbf{R}$ by $\pi_2((x, y)) = y$. For the following subsets $S \subset \mathbf{R^2}$, determine whether or not π_1 and π_2 are injective or surjective.

a. $S_1 = \{(x, y) \mid x \in \mathbf{R}, y \in \mathbf{R}\}$

b. $S_2 = \{(x, y) \mid x \in \mathbf{Q}, y \in \mathbf{R}\}$

c. $S_3 = \{(x, y) \mid x \in \mathbf{R}/\{\mathbf{0}\}, y = \frac{1}{x}\}$

d. $S_4 = \{(x, y) \mid y = \frac{x}{2}\}$

2. For a function $f : A \to B$, let $S = \{(x, y) \mid x \in A, y = f(x)\}$, and consider $\pi_1 : S \to A$ and $\pi_2 : S \to B$. Answer the following statements and justify your answers.

a. Is π_1 is necessarily injective?

b. If f is injective, then what can you conclude about π_2?

c. If f is surjective, then what can you conclude about π_2?

3. Let $A,\ B,\ C$ be sets and $f,\ g,\ h$ be functions.

a. Given $g : A \to B$, and $f : B \to C$, find $f \circ g$.

i. $A = B = C = \mathbf{R}$, $g(x) = x$, $f(x) = x^2$.

ii. $A = B = C = \mathbf{R}$, $g(x) = x + 1$, $f(x) = x^3$.

iii. $A = \mathbf{R}^2,\ B = \mathbf{Z}[i],\ C = \mathbf{N}$, $g : A \to B$ is given by $g(\alpha, \beta) = \alpha + \beta i$, and $f : B \to C$ is given by $f(a + bi) = N(a + bi) = a^2 + b^2$.

iv. $A = \mathbf{R}^3,\ B = C = \{p \mid p \text{ is a polynomial}\}$, $g : A \to B$ is given by $g(\alpha, \beta, \gamma) = \alpha x^2 + \beta x + \gamma$, and $f : B \to C$ is given by $f(p) = D(p) = p'$.

b. For each of the f, g pairs in 3a, **if possible** find $h : A \to B$ such that $f \circ g = f \circ h$, but $g \neq h$.

c. Let $f : B \to C$ be a function. Complete and prove the following statement. If f is ________, then $\forall g : A \to B,\ h : A \to B$ functions, then $f \circ g = f \circ h \Rightarrow g = h$.

d. What is the converse of the statement in 3c? What would the structure of the argument be if you were to prove the converse directly? by contradiction? by contrapositive? Pick which approach you prefer and prove the converse.

4. Let $A,\ B,\ C$ be sets.

a. Show it is possible to find $A,\ B,\ C$, and functions, $f : A \to B$, $g : B \to C$, and $h : B \to C$, such that $g \circ f = h \circ f$, but $g \neq h$.

b. Let $f : A \to B$ be a function.
If f is __________, then $\forall g : B \to C,\ h : B \to C$ functions, then
$g \circ f = h \circ f \Rightarrow g = h$.

c. What is the converse of the statement in 4b? What would the argument be if you were to prove the converse directly? by contradiction? by contrapositive? Pick which approach you prefer and prove the converse.

5. Let $f : A \to B$ be a function. $\forall y \in B$, define $\mathcal{F}(y) = \{x \in A \mid f(x) = y\}$.

a. What is an appropriate codomain for $\mathcal{F}$?

b. Prove that if f is surjective, then $\mathcal{F}$ is injective.

c. What is the converse? Is it true? Prove or disprove it.

d. Can $\mathcal{F}$ be surjective? Prove or disprove it.

6d: implementation study
Injectivity, Surjectivity, and Function Composition - Definitions

- Injective Function: A function, $f : A \to B$, is called *injective* if every image of f comes from exactly one input, that is $a_1 \neq a_2 \Rightarrow f(a_1) \neq f(a_2)$, and $f(a_1) = f(a_2) \Rightarrow a_1 = a_2$.
- Surjective Function: A function, $f : A \to B$, is *surjective* if every element of B is an image of the function, that is $\forall b \in B, \exists a \in A$ such that $f(a) = b$.
- $\mathcal{G}(b)$: for a function, $g : A \to B$, and an element $b \in B$, $\mathcal{G}(b)$, the *pre-image of b under g*, is the set of all elements of A that get sent to b. That is, $\mathcal{G}(b) = \{a \in A \mid f(a) = b\}$.
- $\mathcal{P}$: the set of *r*eal polynomials, $\{p \mid p$ is a polynomial expression in x with real coefficients$\}$. Note that $p \in \mathcal{P}$ has the form $a_n x^n + ... + a_1 x^1 + a_0$, where $a_i \in \mathbf{R}$.
- $\lfloor x \rfloor$: the *floor function* or *greatest integer function.* $\lfloor x \rfloor : \mathbf{R} \to \mathbf{R}$ returns the value of the greatest integer less than or equal to x. For example, $\lfloor 2.99 \rfloor = \lfloor 2.01 \rfloor = \lfloor 2.5 \rfloor = 2$
- $\lceil x \rceil$: the *ceiling function* or *least integer function.* $\lceil x \rceil : \mathbf{R} \to \mathbf{R}$ returns the value of the least integer greater than or equal to x. For example, $\lceil 2.99 \rceil = \lceil 2.01 \rceil = \lceil 2.5 \rceil = 3$

Projection Maps:

- $\pi_1 : A\text{x}B \to A$ is called the *first projection map* and sends an ordered pair (a, b) to it's first component. That is, $\pi_1(a, b) = a$.
- $\pi_2 : A\text{x}B \to B$ is called the *second projection map* and sends an ordered pair (a, b) to it's second component. That is, $\pi_2(a, b) = b$.

Appendix 7: Group Folder Inserts - Pilot Study Version

ROLES and RESPONSIBILITIES

MANAGER

- keeps group on task
- organizes tasks into subtasks
- manages sequence of steps

- We also need to consider…
- We need to move on to the next step.
- Let's come back to that later.

SKEPTIC

- plays the role of the devil's advocate
- helps group avoid snap decisions and unreasoned agreement
- pushes members to explore all possibilities

- What else could we say about this?
- Are there other possibilities or cases?
- Why can we assume that?
- Why does that follow?

EXPLAINER

- explains and summarizes

- That follows because…
- So basically this proof says that…
- Intuitively, this means…

CHECKER/RECORDER

- checks for understanding and unanimous consent
- writes up the group's solutions, explanations and proofs
- presents group's work to class

- Can you explain how we got this?
- Does everyone agree?

ROLE ROTATION

	DAYS 1 & 5	DAYS 2 & 6	DAYS 3 & 7	DAYS 4 & 8
A	MANAGER	SKEPTIC	EXPLAINER	RECORDER
B	SKEPTIC	MANAGER	RECORDER	EXPLAINER
C	EXPLAINER	RECORDER	MANAGER	SKEPTIC
D	RECORDER	EXPLAINER	SKEPTIC	MANAGER

A - Ursula
B - Omar
C - Bill
D - Ingrid

GROUP PROCESSING

- What are three ways you worked well together today?
- What problems did you have interacting well as a group?
- What concrete steps could you take next time to interact as a group more effectively?

Write down your answers individually first, and then share your answers with your group. Be open and honest, but try to be considerate as well.

Group Folder Inserts - Implementation Study Version

ROLES and RESPONSIBILITIES

MANAGER

- keeps group on task and working together
- organizes tasks into subtasks
- manages sequence of steps

- We also need to consider…
- We need to move on to the next step.
- Let's come back to that later, right now we're working on ....

SKEPTIC

- plays the role of the devil's advocate
- helps group avoid snap decisions and unreasoned agreement
- pushes members to explore all possibilities

- What else could we say about this?
- Are there other possibilities or cases?
- Why can we assume that?
- Why does that follow?

YES-MAN

- confirms claims and assertions made by group mates
- checks for understanding and consensus
- continues conversation based on valid claims

- Yes, and …
- Does everyone agree that this is valid?
- Does everyone understand why?

RECORDER/Presenter

- writes up the group's work for the day
- presents group's work to class

- Could you please repeat that?
- Hang on just a moment while I write that down.
- Does everyone agree that this is what we want to present?

ROLE ROTATION

	Usher	Travis	Ethan	Nick
9/19	manager	skeptic	yes-man	recorder
9/26	skeptic	manager	recorder	yes-man
10/3	yes-man	recorder	manager	skeptic
10/17	recorder	yes-man	skeptic	manager
10/24	manager	skeptic	yes-man	recorder
10/31	skeptic	manager	recorder	yes-man
11/7	yes-man	recorder	manager	skeptic
11/14	recorder	yes-man	skeptic	manager

GROUP PROCESSING

- What are three ways you worked well together today
- What problems did you have working together as a group
- How well did you fulfill your assigned role? Explain.
- Did your group mates fulfill their roles well? Explain.
- What concrete steps could you take next time to interact as a group more effectively?

Write down your answers individually first, and then share your answers with your group. Be open and honest, but try to be considerate as well.

References

Alcock, L., & Inglis, M. (2008). Doctoral Students' Use of Examples in Evaluating and Proving Conjectures. *Educational Studies in Mathematics.* 69, 111-129. DOI 10.1007/s10649-008-9149-x

Alcock, L., & Inglis, M. (2009). Representation Systems and Undergraduate Proof Production: A Comment on Weber. *The Journal of Mathematical Behavior*. 28, 209-211

Almeida, D. (2000). A survey of mathematics undergraduates' interaction with proof: some implications for mathematics education. *International Journal of Mathematical Education in Science and Technology.* 31(6) 869-890.

Almeida, D. (2003). Engendering proof attitudes: can the genesis of mathematical knowledge teach us anything?. *International Journal of Mathematical Education in Science and Technology,* 34(4), 479-488.

Almeida, D. (1995). Mathematics Undergraduates' Perceptions of Proof. *Teaching Mathematics and Its Applications,* 14(4), 171-177.

Astin, A. W. (1993). What Matters in College? Implications for Cooperative Learning of a New National Study. In Cooper, J., Robinson, P., & Ball, D. (Eds.), *Small Group Instruction in Higher Education: Lessons from the Past, Visions of the Future.* Stillwater, OK: New Forums Press, Inc. 44-61

Bakó, M. (2002). Why we need to teach logic and how can we teach it? *International Journal for Mathematics Teaching and Learning*. Retrieved July 24, 2012 from www.cimt.plymouth.ac.uk/journal/bakom.pdf

Blanton, M., & Stylianou, D. (2002). Exploring sociocultural aspects of undergraduate students' transition to mathematical proof. In D. Mewborn, et al (Eds.), *Proceedings of the Twenty-fourth Annual Meeting for the North American Chapter of the International Group for the Psychology of Mathematics Education,* (Vol. 4, pp. 1673-1680). Athens, GA.

Blanton, M., Sylianou, D. A., David, M. M. (2009). Understanding Instructional Scaffolding in Classroom Discourse on Proof. In Stylianou, D. A., Blanton, M. L., & Knuth, E. J. *Teaching and Learning Across the Grades: A K-16 Perspective*. London, England: Taylor & Francis

Blanton, M. L., Stylianou, D. A. & David, M. M. (2003). The nature of scaffolding in undergraduate students' transition to mathematical proof. In the Proceedings of the 27th Annual Meeting for the International Group for the Psychology of Mathematics Education. Honolulu, Hawaii: Center for Research and Development Group, University of Hawaii. 2, 113-120.

Boyle, J. (2012). A Study of Prospective Secondary Mathematics Teachers' Evolving Understanding of Reasoning-and-Proving. (EdD), University of Pittsburgh.

CadwalladerOlsker, T. (2011) What Do We Mean by Mathematical Proof? *The Journal of Humanistic Mathematics.* 1 (1): 33-60.

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *The Wingspread Journal,* 9(2), special insert. (Reprinted from AAHE Bulletin, 39(7), 3–7.

Committee on the Undergraduate Program in Mathematics, (2001). Mathematics and the mathematical sciences in 2010: What should students know? CUPM preliminary report. New Orleans, LA: Author.

Cooper, J. (1994) What is Cooperative Learning? In Cooper, J., Robinson, P., & Ball, D. (Eds.), *Small Group Instruction in Higher Education: Lessons from the Past, Visions of the Future.* Stillwater, OK: New Forums Press, Inc. 3-5.

Cooper, J., & Robinson, P. (1994). FIPSE-Sponsored CL Research at Dominguez Hills and Community Colleges. In Cooper, J., Robinson, P., & Ball, D. (Eds.), *Small Group Instruction in Higher Education: Lessons from the Past, Visions of the Future.* Stillwater, OK: New Forums Press, Inc. 29-32.

Cuoco, A. & Rotman, J. (2013). *Learning Modern Algebra: From Early Attempts to Prove Fermat's Last Theorem*. MAA Textbooks Series.

Cuseo, J. (1992). Collaborative and Cooperative Learning in Higher Education: A Proposed Taxonomy. In Cooper, J., Robinson, P., & Ball, D. (Eds.), *Small Group Instruction in Higher Education: Lessons from the Past, Visions of the Future.* Stillwater, OK: New Forums Press, Inc. 18-26.

Cuseo, J. (1994). Critical Thinking and Cooperative Learning: A Natural Marriage. *In Cooper, J., Robinson, P., & Ball, D. (Eds.), Small Group Instruction in Higher Education: Lessons from the Past, Visions of the Future.* Stillwater, OK: New Forums Press, Inc. 63-74

Davidson, N. (1990). The Small-Group Discovery Method in Secondary- and College-Level Mathematics. In Dubinsky, E., Mathews, D., & Reynolds, B. E. (Eds.), *Readings in Cooperative Learning for Undergraduate Mathematics*. USA: The Mathematical Association of America. 237-247.

Deslauriers, L., Schelew, E., & Wieman,C. (2011). Improved Learning in a Large-Enrollment Physics Class. *Science.* 332(6031), 862-864.

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

Ericsson, K. A., Krampe, R. Th., & Tesch-Roemer, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review,* 100, 363-406.

Froyd, J. (2008). White paper on promising practices in undergraduate STEM education. Paper presented at the National Research Council's Workshop Linking Evidence to Promising Practices in STEM Undergraduate Education, Washington, DC. Available at http://www7.nationalacademies.org/bose/Froyd_Promising_Practices_CommissionedPaper.pdf.

Grassl, R. & Mingus, T. T. Y. (2004). Team teaching and cooperative groups in abstract algebra: nurturing a new generation of confident mathematics teachers. *International Journal of Mathematical Education in Science and Technology.* 38(5), 581-597.

Gruber, S. E., Habanek, D. V. (1996). College Classrooms' Lost Gold Mine: The Cooperative Base Group. In Cooper, J., Robinson, P., & Ball, D. (Eds.), *Small Group Instruction in Higher Education: Lessons from the Past, Visions of the Future.* Stillwater, OK: New Forums Press, Inc. 171-178.

Harel, G., & Sowder, L. (1998). Students' proof schemes. *Research on Collegiate Mathematics Education, Vol. III*. In E. Dubinsky, A. Schoenfeld, & J. Kaput (Eds.), AMS, 234-283.

Harskamp, E., Ding, N., & Suhre, C. (2008). Group composition and its effect on female and male problem-solving in science education. *Educational Research*, 50(4), 307-308.

Hart, E. (1994). A conceptual analysis of the proof-writing performance of expert and novice students in elementary group theory. In Kaput, J. & Dubinsky, E. (Eds.), *Research issues in undergraduate mathematics learning: Preliminary analyses and results*. Washington, DC: The Mathematical Association of America. 49-63

Heller, P., Hollabaugh, M. (1992). Teaching Problem Solving Through Cooperative Grouping Part 2: Designing Problems and Structuring Groups. *American Journal of Physics*, 60(7), 637-644.

Heller, P., Keith, R. & Anderson, S. (1992). Teaching problem solving through cooperative grouping. Part 1: Group versus individual problem solving. *American Journal of Physics*, 60(7), 627-636.

Inglis, M., & Simpson, A. (2009). Conditional inference and advanced mathematical study: Further evidence. *Educational Studies in Mathematics*, 72, 185-198.

Inglis, M. & Simpson, A. (2008). Conditional inference and advanced mathematical study. *Educational Studies in Mathematics*, 67 (3), pp. 187-204.

Johnson, D. W., Johnson, R. T. (1990). Social Skills for Successful Group Work. *Educational Leadership,* 48(7), 29-33.

Johnson, D. W., Johnson, R. T., Smith, K. (2007). The State of Cooperative Learning in Postsecondary and Professional Settings. *Educational Psychology* Review. 19(1), 15-29.

Jones, K. A., & Jones, J. L. (2008). A Descriptive Account of Cooperative-Learning based Practice in Teacher Education. *College Quarterly.* 11(1).

Leron, U. & Dubinsky, E. (1995). An abstract algebra story. *American Mathematical Monthly.* 102(3), 227-242.

Levine, A. and Shanfelder, B. (2000) The Transition to Advanced Mathematics. *PRIMUS*, 10(2), 97–110.

Marty, R. H. (1991). Getting to eureka! Higher order reasoning in math. *College Teaching,* 39(1), 3-6.

McLoughlin, P. (2010). Aspects of a Neoteric Approach to Advance Students' Ability to Conjecture, Prove, or Disprove. Paper presented at the annual meeting of the The Mathematical Association of America, MathFest, Omni William Penn, Pittsburgh, PA, 2011. from http://www.allacademic.com/meta/p435521_index.html

Mejia-Ramos,J. P., Fuller, E., Weber, K., Rhoads, K., Samkoff, A. (2012). An Assessment Model for Proof Comprehension in Undergraduate Mathematics. *Educational Studies in Mathematics.* 79, 3-18. doi: 10.1007/s10649-011-9349-7

Millis, B. (1992). How Cooperative Learning Can fulfill the Promises of the "Seven Principles." In Cooper, J., Robinson, P., & Ball, D. (Eds.), *Small Group Instruction in Higher Education: Lessons from the Past, Visions of the Future.* Stillwater, OK: New Forums Press, Inc. 39-43.

Moore, R. C. (1994). "Making the transition to formal proof. *Educational Studies in Mathematics,* 27, 249-266.

Piatek-Jimenez, K. (2010). Students' Interpretations of Mathematical Statements Involving Quantification. *Mathematics Education Research Journal,* 22(3), 41-56.

Prescott-Johnson, S. (1992). Getting Started with Cooperative Learning. In Cooper, J., Robinson, P., & Ball, D. (Eds.), *Small Group Instruction in Higher Education: Lessons from the Past, Visions of the Future.* Stillwater, OK: New Forums Press, Inc. 161-165.

Prescott-Johnson, S. (1992). More Tips for Getting Started in Cooperative Learning (CL). In Cooper, J., Robinson, P., & Ball, D. (Eds.), *Small Group Instruction in Higher Education: Lessons from the Past, Visions of the Future.* Stillwater, OK: New Forums Press, Inc. 155-160.

Schumacher, C. (2001) *Chapter Zero: Fundamental Notions of Abstract Mathematics (2nd Ed.).* Addison-Wesley.

Segal, J. (1999). Learning about mathematical proof: conviction and validity. *Journal of Mathematical Behavior,* 18, 191-210.

Selden, A. & Selden, J. (2007) Overcoming Students' Difficulties in Learning to Understand and Construct Proofs. In M. P. Carlson and C. Rasmussen (Eds.), *Making the Connection: Research and Teaching in Undergraduate Mathematics Education* (pp. 95-110), MAA Notes Volume No. 73, 2008

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

Selden, A. & Selden, J. (2003). Errors and misconceptions in college level theorem proving. In Novak, J. D. (Ed.) *Proceedings of the Second International Seminar on Misconceptions and Educational Strategies in Science and Mathematics*, Vol. III, Cornell University, July 1987, 457-470.

Selden, J. and Selden, A. (1995) Unpacking the logic of mathematical statements. *Educational Studies in Mathematics,* 29(2), 123-151.

Sowder, L., & Harel, G. (1998). Types of students' justifications. *Mathematics Teacher,* 91, 670–675.

Sowder L., Harel G. (2003) Case Studies of Mathematics Majors' Proof Understanding, Production, and Appreciation. *Canadian Journal of Science, Mathematics and Technology Education* 3(2)

Springer, L. (1998). Research on Cooperative Learning in College Science, Mathematics, Engineering, and Technology. In Cooper, J., Robinson, P., & Ball, D. (Eds.), *Small Group Instruction in Higher Education: Lessons from the Past, Visions of the Future.* Stillwater, OK: New Forums Press, Inc. 33-38.

Sriraman, B., VanSpronsen, H., and Haverhals, N. (2010). On Proof and Certainty -- Some Educational Implications. In *Theories of Mathematics Education: Seeking New Frontiers*. (Eds.) B. Sriraman and L. English. Berlin/Heidelberg: Springer Science. 369-380.

Stetson, N. E. (1996). Using Small-Group Development to Facilitate Cooperative Learning Base Groups. In Cooper, J., Robinson, P., & Ball, D. (Eds.), *Small Group Instruction in Higher Education: Lessons from the Past, Visions of the Future.* Stillwater, OK: New Forums Press, Inc. 166-170.

Tall, D. and Vinner, S. (1981) Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics,* Vol.12 (No.7). pp. 151-169.

Towns, M. (1997). Building Community One Brick at a Time. In Cooper, J., Robinson, P., & Ball, D. (Eds.), *Small Group Instruction in Higher Education: Lessons from the Past, Visions of the Future.* Stillwater, OK: New Forums Press, Inc. 201-206.

Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics,* 48(1), 101-119.

Weber, K. (2003). A procedural route toward understanding the concept of proof. In N. Pateman, B. Daugherty, and J. Zilliox (eds) *Proceedings of the 27th Conference of the International Group for the Psychology of Mathematics Education*. 4, 395-401. Honolulu, HI.

Weber, K. (2004). A framework for describing the processes that undergraduates use to construct proofs. In M. J. Hoines & A. B. Fuglestad (Eds.), *Proceedings of the Twenty-eighth Annual Meeting of the International Group for the Psychology of Mathematics Education.* 4 425–432. Bergen, Norway.

Weber, K. (2009). How Syntactic Reasoners Can Develop Understanding, Evaluate Conjectures, and Generate Counterexamples in Advanced Mathematics. *The Journal of Mathematical Behavior*. 28, 200-208.

Weber, K. (2010). Mathematics Majors' Perceptions of Conviction, Validity, and Proof. *Mathematical Thinking and Learning,* 12(4), 306-336. doi:10.1080/10986065.2010.495468

Weber, K., & Alcock, L. (2004). Semantic and syntactic proof productions, *Educational Studies in Mathematics,* 56, 209–234.

Weber, K. (2008). How mathematicians determine if an argument is a valid proof. *Journal for Research in Mathematics Education 39*, 431-459.

Weber, K., & Mejia-Ramos, J. P. (2009). An Alternative Framework to Evaluate Proof Productions: A Reply to Alcock and Inglis. *The Journal of Mathematical Behavior*. 28, 212-216.

White, A. L., Way, J., Perry, B., & Southwell, B. (2006). Mathematical Attitudes, Beliefs and Achievement in Primary Pre-service Mathematics Teacher Education. *Mathematics Teacher Education and Development.* 7. 33-52.

Young, A. N. (2012, January). Improv Techniques for the Mathematics Classroom. Paper presented at the Joint Mathematics Meeting, Boston.