


7-9-2009

Development of scoring rubrics and pre-service teachers ability to validate mathematical proofs

Timothy J. Middleton

Follow this and additional works at: https://digitalrepository.unm.edu/math_etds

 Part of the [Applied Mathematics Commons](#), [Mathematics Commons](#), and the [Statistics and Probability Commons](#)

Recommended Citation

Middleton, Timothy J. "Development of scoring rubrics and pre-service teachers ability to validate mathematical proofs." (2009). https://digitalrepository.unm.edu/math_etds/30

This Thesis is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Mathematics & Statistics ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

Timothy J. Middleton

Candidate

Mathematics

Department

This thesis is approved, and it is acceptable in quality and form for publication on microfilm:

Approved by the Thesis Committee:

Dr. Kristin Umland

, Chairperson

Dr. Richard Kitchen

Dr. James Ellison

Accepted:

Dean, Graduate School

Date

**DEVELOPMENT OF SCORING RUBRICS
AND PRE-SERVICE TEACHERS' ABILITY
TO VALIDATE MATHEMATICAL PROOFS**

BY

TIMOTHY J. MIDDLETON

**B.S., EDUCATIONAL CURRICULUM AND INSTRUCTION,
TEXAS A&M UNIVERSITY, 1990**

THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

**Master of Science
Mathematics**

The University of New Mexico
Albuquerque, New Mexico

May, 2009

ACKNOWLEDGEMENTS

My first thanks must go to Dr. Kristin Umland, my advisor and thesis chair. Without her persistence and patience, I could not have imagined completing this endeavor. Her passion for meaningful work in mathematics education continually inspired me and was invaluable in helping me plan, implement, and evaluate my study.

I would also like to thank my other committee members, Dr. Richard Kitchen and Dr. James Ellison. The idea for my thesis resulted from a paper I wrote in a course with Dr. Kitchen, and his insights in education provided me with direction. Dr. Ellison's instruction expanded my horizons in mathematics and gave me a more thorough framework for continuing the work I love – teaching high school calculus.

I appreciate the efforts of all my professors in the Department of Mathematics. Each helped me explore a little more of the world of math. In particular, I am grateful for the time and support provided by Dr. Cristina Pereyra and Dr. Alexandru Buium.

Finally, I cannot say “Thank you” enough to Mrs. Ann Landry and Dr. Morris Gelber – two former teachers who became my dearest friends and who have always believed in me. Their love and support carried me through this project.

**DEVELOPMENT OF SCORING RUBRICS
AND PRE-SERVICE TEACHERS' ABILITY
TO VALIDATE MATHEMATICAL PROOFS**

BY

TIMOTHY J. MIDDLETON

ABSTRACT OF THESIS

Submitted in Partial Fulfillment of the
Requirements for the Degree of

**Master of Science
Mathematics**

The University of New Mexico
Albuquerque, New Mexico

May, 2009

**DEVELOPMENT OF SCORING RUBRICS AND PRE-SERVICE TEACHERS'
ABILITY TO VALIDATE MATHEMATICAL PROOFS**

by

Timothy J. Middleton

B.S., Educational Curriculum and Instruction, Texas A&M University, 1990

M.S., Mathematics, University of New Mexico, 2009

ABSTRACT

The basic aim of this exploratory research study was to determine if a specific instructional strategy, that of developing scoring rubrics within a collaborative classroom setting, could be used to improve pre-service teachers' facility with proofs. During the study, which occurred in a course for secondary mathematics teachers, the primary focus was on creating and implementing a scoring rubric, rather than on direct instruction about proofs. In general, the study had very mixed results. Statistically, the quantitative data indicated no significant improvement occurred in participants' ability to validate proofs. However, the qualitative results and the considerable improvement by some participants warrant further investigation of the attempted instructional technique.

TABLE OF CONTENTS

LIST OF FIGURES.....	viii
CHAPTER 1: INTRODUCTION	1
Contextual Background	1
Research Questions	3
Structure of the Study	4
Overview of Results	5
CHAPTER 2: REVIEW OF LITERATURE	7
Proofs for Understanding	7
Proofs in Mathematics Education	9
Research Investigating the Experiences of Pre-College Students With Proofs	10
Research Investigating the Experiences of Undergraduates With Proofs	13
Pedagogical Aspects of Proof	16
Innovations in the Teaching of Proofs	20
Conclusion	23
CHAPTER 3: METHODS	25
Action Research and Mixed Methods Design	25
Participants	28
Instruments, Data Sources, and Artifacts	29
Procedures	32
Timeline for the Study	33
Scoring and Coding of the Pre-Tests and Post-Tests	36

Analyzing the Qualitative Data	38
Validity	38
CHAPTER 4: RESULTS	41
Quantitative Results	41
Qualitative Results	47
Initial confusion about rubrics	48
Grade issues	51
Proof Conceptions	54
CHAPTER 5: DISCUSSION	57
Complex Interactions	57
Confounding Factors	63
Further Work	66
Conclusion	68
APPENDICES	69
Appendix A: Pre-Test	70
Appendix B: Post-Test	77
Appendix C: Consent to Participate in Research Form	84
Appendix D: Rubric Drafts	87
Appendix E: Example Rubrics	89
Appendix F: Arguments Presented to Participants During the Study	91
Appendix G: Writing Prompts	96
REFERENCES	98

LIST OF FIGURES

<i>Figure 1.</i> Number of correct multiple-choice responses by item on Pre-Test versus Post-Test.	42
<i>Figure 2.</i> Number of correct open-ended responses by item on Pre-Test versus Post-Test.	43
<i>Figure 3.</i> Number of participants with a given score on the multiple-choice items of the Pre-Test.	44
<i>Figure 4.</i> Number of participants with a given score on the multiple-choice items of the Post-Test.	44
<i>Figure 5.</i> Number of participants with a given score on the open-ended items of the Pre-Test.	45
<i>Figure 6.</i> Number of participants with a given score on the open-ended items of the Post-Test.	45

Chapter 1

Introduction

Several research studies conducted over the past three decades have demonstrated the various problems encountered by students and teachers when trying to understand and write formal mathematical proofs (e.g., Bell, 1976; Harel & Sowder, 1998; Knuth, 2002a, 2002b; Selden, A., & Selden, J., 2003; Senk, 1985). However, few researchers have tested new instructional strategies in this area (e.g., see Goff, 2002). The basic aim of my research was to determine if a specific instructional strategy, that of developing scoring rubrics within a collaborative classroom setting, could be used to improve pre-service teachers' facility with proofs. In particular, the study tried to ascertain how the development and utilization of rubrics for scoring mathematical arguments, as well as the associated classroom discussions, affected pre-service teachers' ability to validate a proof, to recognize the logical structure required for a proof, and to identify the specific errors in an invalid proof. The research reported in this thesis represents an exploratory investigation of an indirect approach to teaching proof validation. The primary focus in class was on creating and implementing a scoring rubric, rather than on direct instruction about proofs.

Contextual Background

My study arose in the context of questions posed by other mathematicians and researchers. What do students believe makes a proof valid? How do students define or identify correct mathematical reasoning? What are the difficulties students have with validating proofs? In the context of my work, "validating a proof" means determining if a

specified mathematical argument in fact represents a valid proof of the statement given.

Many researchers have documented the fact that high school and undergraduate students, as well as high school teachers, are not proficient with proofs. In particular, studies by Bell (1976), Harel & Sowder (1998), Herbst (2002), Knuth (2002a, 2002b), Moore (1994), Selden & Selden (1995, 2003), Senk (1985), and Thompson (1996) found the following:

- Pre-college students, undergraduates, and secondary teachers tend to focus on the superficial details of an argument, such as the format or language used in the presentation, rather than the mathematical reasoning of the argument. For example, subjects in these studies would determine that a given argument represented a correct proof simply because it appeared to use correct mathematical notation, when in fact it proved the converse.
- Students cannot determine the logical structure of mathematical statements so as to know how a correct proof should be structured. Also, students generally are unable to recite definitions and do not see how the definitions dictate the possible structure for a proof.
- Students possess little or no intuitive understanding of the mathematical concepts they are working with and cannot create their own examples to gain a better understanding of a mathematical statement. Moreover, students misunderstand the use of examples and counterexamples in establishing the veracity of a statement.
- Even undergraduates in transition or bridge courses that teach logic and the writing of proofs cannot validate proofs. This is also true of pre-service and in-service teachers.

- Secondary teachers have little understanding of the role of proofs within the mathematical community or within secondary education.

Reviewing the literature related to proofs, I began to ask other questions. What can mathematics educators do about the difficulty students have with proofs? Can particular instructional methods improve students' facility with proof? What aspects of my course can lead to improved mathematical reasoning? Can one course be a catalyst for change? Teaching high school calculus for 11 years, I had seen that many students entered the course with insufficient algebra skills. Rather than focus directly on reteaching algebra skills, I learned that a more successful approach was to simply incorporate algebra review in the context of solving the more complicated problems that arise in calculus. My experience with contextual learning, therefore, generated the idea of not teaching proof methods directly, but in the context of another focus. Moreover, in my secondary classes, I had frequently used scoring rubrics to assess complicated student work, such as creative projects. Therefore, when I learned that I would be teaching Math 338: Mathematics for Secondary Teachers at the University of New Mexico, I had already planned to introduce the students to the use of scoring rubrics in math. A synthesis of my experiences with contextual learning, my growing interest in student difficulties with proofs, and my plan to introduce pre-service teachers to scoring rubrics led to the creation of the study reported in this paper.

Research Questions

The focus of the study thus became the use of an indirect instructional technique to improve participants' facility with proofs. Specifically, the research addressed the question "Does the development and utilization of rubrics for scoring mathematical

arguments, as well as the associated classroom discussions, improve pre-service teachers' ability to validate a proof, to recognize the logical structure required for a proof, and to identify the specific errors in an invalid proof?" This was not a definitive study, but an exploratory investigation of this question. Other supporting questions drove my research design. What is the nature of the changes my student-participants experience with regard to analyzing mathematical arguments? Which aspects of the instructional approach seem to help the most? Which aspects seem to interfere? What are the correct ways to measure these changes?

Structure of the Study

Since my questions aimed to reveal potentially complicated interactions, the study employed *action research* techniques and used a *mixed methods* design, combining both quantitative and qualitative data. The study did not involve a statistically significant number of participants, so purely quantitative methods would not have been appropriate. Also, qualitative methods were more helpful in clarifying the connection between the intervention technique and any improved performance by participants. Quantitative data sources included the results of the Pre-Test and the Post-Test (Appendices A and B). These tests included both multiple-choice and open-ended items. Participants were allowed to take as much time as they wanted to complete each test. Qualitative data came from a variety of sources that arose as normal instructional activities in class, such as prompted responses written by student-participants and preliminary scorings collected from participants as the class developed a rubric. My own field notes as the teacher-researcher also became a source of qualitative data.

Subjects in the study were students enrolled in Math 338: Mathematics for Secondary Teachers at the University of New Mexico during the Fall 2003 semester. This class met one evening per week, so all references to a specified week involve a single class period. The study lasted a total of 14 weeks. After an introduction to the study, student-participants took the Pre-Test. Then, over the course of six weeks, the participants and I developed a scoring rubric for mathematical arguments, tested and refined the rubric by using it to score proof attempts that I presented to the class, and discussed our experiences both verbally and in writing. At the end of the study, participants took the Post-Test. A more specific timeline for the study can be found in Chapter 3: Methods.

As the study progressed, I realized that student-participants had very diverse conceptions and experiences related to rubrics. One of the first steps, therefore, was to define locally what type of rubric I and the participants would be developing collaboratively. We explored examples of three different styles of rubrics, which I classified as holistic style, matrix style, and outline-and-point-value style. Further explanations of these three styles and references to the examples used in the study can be found in Chapter 4: Results. I specified that the outline-and-point-value style would be the type of rubric we would develop for our work with scoring mathematical arguments. The question that then drove the next few weeks of work was “What do we need in a rubric for proofs?”

Overview of Results

In general, the study had very mixed results. Statistically, the quantitative data indicated no significant improvement occurred in participant responses from the Pre-Test

to the Post-Test. Tracking the progress of several student-participants through the qualitative data sources revealed several potentially complex interactions occurred. Some participants became concerned about how the score from the rubric would be translated into a grade for students in their classes. The rubric therefore seemed to interfere with their focus on analyzing arguments. For other participants, the rubric seemed to confuse them by adding another layer of complexity. However, eight participants did show improvement from the Pre-Test to the Post-Test. In fact, two participants more than doubled their number of correct responses. Investigating their work with the rubric indicated that the development of the rubric did in fact help them be more attentive to the important aspects of a mathematical argument, including the overall logical flow or structure of an argument. Due to the small sample size, no inferences can be drawn from the quantitative data. However, the qualitative results and the considerable improvement by some participants warrant further investigation of the attempted instructional technique.

Chapter 2

Review of Literature

Many recent research studies have investigated the experiences that students have with writing and analyzing mathematical proofs and that teachers have with teaching proofs. The associated literature clearly indicates the need for improved instruction when teaching proof both to high school students and to undergraduates, particularly those who aspire to become secondary mathematics teachers. To put the purpose and results of my study in context, the following review of literature includes articles by several authors regarding the role of proof in mathematics and more specifically in mathematics education. Based on this exploration of previous research, I became intrigued by the idea of devising an approach to presenting proofs in a mathematics course for pre-service teachers that would facilitate their understanding of proofs.

Proofs for Understanding

Upon first thinking about proofs, one may believe that proofs are absolute and very little doubt surrounds the purpose, validity, and meaning of proof. However, mathematical philosophers throughout the ages have held different, often conflicting beliefs concerning the need for proofs and the formality and absoluteness thereof (Hanna, 1995; Harel & Sowder, 1998; Hersh, 1993; Kleiner & Movshovitz-Hadar, 1990; MacKernan, 1996). For instance, in the era shortly after World War II, mathematicians often took one of two differing views regarding proofs: that of formalism (classicism) or that of constructivism. Formalists generally accepted the laws of formal logic and built proofs based on a set of accepted axioms, statements that are considered to be

unprovable. Constructivists rejected the law of the excluded middle with regard to infinite sets and were skeptical of the axiom of choice (Hersh, 1993). By the end of the 20th century, however, little remained of the controversy surrounding these assumptions and most proofs implicitly accept them as true. In a similar way, mathematicians currently engage in debates about whether or not to accept new forms of proof that have arisen as a result of technology, such as computer-generated proofs and graphical representations of mathematical concepts. A now classic example is Appel and Haken's proof of the four-color theorem. Some mathematicians are concerned about the fallibility of electronic processes, particularly those originating from human errors in the production of software, whereas others object to the fact that computer-generated proofs hide the nature of the proof and therefore do not provide insight as to why a theorem is true (Hersh, 1993).

Consequently, the evolution of ideas about what constitutes a proof points out that perhaps more central to the work of mathematics educators are the *roles* that proof plays within the mathematical community rather than the actual methods of proof. Lately, researchers and philosophers have generally agreed upon several main purposes of proof (Bell, 1976; Hanna, 2000; Knuth, 2002a). Indisputably, proofs serve to verify the truth of a mathematical statement, and it is this role of proof that is most familiar to students of mathematics. However, proof also serves as a form of communication between mathematicians and as such represents the product of social interactions. Additionally, proofs can lead to the discovery or creation of new mathematics, as was the case in the development of non-Euclidean geometries. A fourth role of proof is the systemization of mathematical results into a coherent structure of axioms, definitions, and theorems.

Finally, proofs can demonstrate why a particular mathematical statement is true by illuminating the underlying concepts. In her often-cited work "Proofs That Prove and Proofs That Explain", Hanna (1989) considers this role of proof as the most important for educators. She advocates the use of proof as a tool for promoting a deeper understanding of mathematical definitions and theorems in the classroom—a theme that surfaces in other articles as well (de Villiers, 1995; Hanna, 1995, 2000; Hersh, 1993; Knuth, 2002a, 2002b).

Proofs in Mathematics Education

Since proof plays such complex and varied roles within the mathematical community, mathematics educators warn that a curriculum devoid of meaningful excursions into proofs, as is often the case in elementary and secondary mathematics classrooms, does not reflect the true nature of mathematics. Moreover, with very little previous exposure to formal proof, students in the U.S. are unreasonably expected to master the complex aspects of proof in a single course— high school geometry. Consequently, NCTM (2000) and others (Sowder & Harel, 1998; Thompson, 1996) have advocated increased attention to proof and deductive reasoning throughout the K-12 curriculum. Also, NCTM states that "reasoning and proof are not special activities reserved for special times or special topics in the curriculum but should be a natural, ongoing part of classroom discussions, no matter what topic is being studied" (p. 342), rather than treating proof simply for the sake of proof.

Some authors, however, believe that many educators may be exaggerating the need for proofs in pre-college mathematics. Wheeler (1990) feels that attempts within the classroom to reproduce the social nature of mathematical proof are inherently artificial.

He also asserts that the complex nature of proofs and the incredible diversity among the forms of acceptable proofs make the learning of proofs, particularly in high school, an incredibly difficult goal. Articles by Herbst (2002) and Slomson (1996) agree, stating that the teaching of proofs places conflicting demands on secondary teachers. According to Almeida (1996), teachers should understand that students' informal arguments are an important component in the development of mathematical reasoning. Furthermore, Fischbein (1982) and Hewitt (1996) advocate that students must gain an intuitive familiarity with mathematical concepts before they can hope to produce formal proofs. MacKernan (1996) states that in fact much of mathematics is the result of induction and intuition rather than formal proof.

Regardless of their views about the amount of emphasis formal proofs should receive, researchers agree that the current methods of dealing with proofs in high school are misleading, impractical, insufficient, and even counter-productive (Epp, 1994; Herbst, 2002; Senk, 1985; Slomson, 1996; Sowder & Harel, 1998; Thompson, 1996). Typical high school curricula only include proofs in geometry classes and even then only teach proofs as a formula to follow (two-column proofs being the classic form). Because students are seldom challenged to prove statements that they do not already perceive as "obviously" true, proofs produce little enlightenment for students and even encourage the notion that mathematics is a static subject in which all the answers are already known.

Research Investigating the Experiences of Pre-College Students With Proofs

The difficulty pre-college students have with proofs has been well documented over the last 30 years. One of the first research studies in this area was completed by Bell (1976), who found that almost half of the 14- to 15-year-old students he surveyed could

not even read and understand given mathematical propositions involving simple number and geometry concepts, much less begin correct explanations for the validity of the propositions. Bell also discovered that the students completely avoided the use of algebra in proofs. Perhaps more importantly, his work brought to the forefront many of the topics that would later be explored more extensively by other mathematics educators, including (a) the various roles that proof plays in mathematics and the importance of the role of explanation in mathematics classrooms, (b) the discrepancy between formal proofs and the type of evidence that students find convincing, (c) the schemes that students use to attempt a proof, and (d) the difficulties students have with even beginning proofs due to a lack of conceptual understanding of the statement or an inability to ascertain the logical structure of the statement.

High school students' inability to start a proof also surfaced in other studies. A comprehensive study by Senk (1985) of 1520 students from 11 schools in 5 states reported that as much as 47% of students scored 0 out of 4 on some of the requested proofs, meaning that students wrote nothing, or only incorrect or useless deductions. Surprisingly, all the students had completed a year-long geometry course that included proofs and the proofs used in the study were similar to those given in standard geometry textbooks. Investigating proofs by contradiction, Thompson (1996) found that many students could not write the correct negation of the original statement in question, which represents the first step in an indirect proof.

Another important facet of students' ability to construct proofs is their belief about the nature of proofs. Chazan (1993) found that students are unclear about what actually constitutes a proof. About one-fourth of the subjects in his study believed that empirical

evidence, such as measuring a few examples, provided a "proof" of a general statement in geometry. Many other students wavered between this belief and one that asserted deductive arguments are necessary for a proof. A more troublesome finding for Chazan was that once a theorem was proven deductively in class, many students still did not believe the theorem. That is, students did not understand or accept the generality of a proof.

In contrast to studies such as those above that explore only cognitive aspects of proof, Hoyles (1997) argues that mathematics education researchers must also investigate the systemic influences of school and curriculum organization on students' facility with proofs. For example, Healy and Hoyles (2000) found evidence that the National Curriculum mandated by England and Wales in 1995 had impacted student perceptions about proof. Students in their nationwide study generally used empirical arguments for their own proofs, but realized that these would not receive the highest marks from teachers and that a valid proof must be general. Students found arguments containing algebraic expressions difficult to follow and instead preferred narrative arguments. In fact, students experienced more success when constructing a proof using words and found these types of proofs to be more explanatory than those involving algebraic manipulations. Healy and Hoyles attribute these findings in part to the structure of the National Curriculum, which includes proofs in the Attainment Target associated with mathematical investigations. Unintentionally, this structure implied that proofs are the final stage in a developmental sequence characterized by data collection and informal testing of concepts with empirical examples. Moreover, the structure of the Curriculum implied that proofs should be independent of algebraic and geometric contexts. Their

study also found evidence of social influences on student performance, noting a smaller variation across schools in the performance of females than the variation for males, and discovered influences related to the organization of the school, such as the amount of time allotted for mathematics instruction or the grouping of students by ability level.

Research Investigating the Experiences of Undergraduates With Proofs

Unfortunately, the problems faced by high school students with proofs are not often solved in college. Mathematics education researchers have completed numerous studies that examine the ability of undergraduates to read and write proofs. These investigations have involved students in transition or "bridge" courses that explicitly teach proofs (Goff, 2002; Moore, 1994; Selden, A., & Selden, J., 2003; Selden, J., & Selden, A., 1995) as well as students in advanced undergraduate college mathematics courses that require the writing of proofs (Harel & Sowder, 1998). More specifically, some studies have involved pre-service teachers, that is, students in a college math course explicitly designed to prepare them for teaching mathematics at the elementary or secondary level (Even, 1993; Jones, 2000). However, these latter studies will be discussed in the next section regarding pedagogical aspects of proofs. This section focuses on the difficulties undergraduates have in mastering proofs. The research shows that although undergraduates may have a better grasp of proofs than do high school students, several deficiencies remain, as does the tendency to complete proofs without understanding the mathematical concepts involved.

Moore (1994) notes that students are expected to write proofs in real analysis, linear algebra, abstract algebra, and other upper-division courses even though (a) students often have only seen proofs in a high school geometry course and therefore have no real

context for working with proofs; (b) precise definitions and proofs, such as $\varepsilon - \delta$ proofs for limits, have been almost completely eliminated from the basic calculus courses preceding upper-level undergraduate math classes; and (c) many colleges do not offer a course that introduces proof writing and thereby attempts to ease the transition to upper-level mathematics courses. He conducted a study in one of these transition courses designed to teach students how to read and write proofs and to acquaint students with some of the pervasive ideas in mathematics. Moore found that students could not even begin a proof because they lacked basic preliminary tools. First of all, most students in the course possessed little or no intuitive understanding of the mathematical concepts they were working with and could not create their own examples to gain a better understanding of a mathematical statement. Subjects in the study also could not understand or use mathematical notation correctly. Moreover, students generally were unable to recite definitions and did not see how the definitions dictated the possible structure for a proof.

Annie and John Selden (1995, 2003) have discovered similar deficiencies in students' ability to tackle proofs. They found that undergraduates had difficulties unraveling the logical structure of mathematical statements, particularly those given in an informal form typical of most textbooks (Selden, J. & Selden, A., 1995). For example, "differentiable functions are continuous" lacks an explicit "if-then" structure. Selden and Selden assert that students' inability to unpack the logical structure of a statement represents an important obstacle to students' ability to determine a correct *proof framework*, the overall organization of a proof. They also reported that students had other difficulties in validating proofs, that is, in determining if an argument represents a correct

proof (Selden, A. & Selden, J., 2003). Students often focused on the form of an argument, rather than the content, incorrectly judging an invalid argument to be a proof simply because it followed a typical structure students had seen previously. This was one of the several ways in which the researchers realized that students tended to look for local errors in the arguments but not global ones. For example, 50% of the subjects in their study initially believed a given argument of a particular statement represented a correct proof when in fact it was a proof of the converse. Also, students often judged the validity of proofs largely based on their ability to understand the mathematics involved or based on the stylistic clarity of the argument.

Whereas some mathematics educators focused on studying *what* difficulties students have with proofs, Harel and Sowder (1998) attempted to explore *why* students have trouble. They realized that proof has meant different things to different people throughout history and their study researched the basic conceptions undergraduates have about proofs. They found that students had differing ideas about what constitutes a convincing argument, thereby giving rise to different *proof schemes*. In the context of Harel and Sowder's study, proof schemes do not refer to the various methods of formal mathematical proof, such as proof by induction or proof by contradiction. Instead, a scheme refers to the internal and external concepts and processes students use to convince *themselves* that a statement is true or false. These schemes can sometimes involve formal proof, but as Harel and Sowder discovered, students often find other methods to be more convincing. The researchers suggest three main categories for proof schemes, each of which has two or more subcategories, for a total of more than a dozen separate schemes. The first category encompasses external conviction proof schemes. For

example, a student may believe a statement is true simply because a person of authority, such as the instructor, has said so. Similarly, the students may accept an argument as a convincing "proof" based solely on the format of the proof. Harel and Sowder attribute students' dependence on these schemes to the emphasis in schools on writing proofs simply for the sake of writing proofs, without stressing the underlying mathematical concepts. The other two categories involve a more internal processing of truth. Regarding the first of these, empirical proof schemes, the authors found that a significant portion of students use the process of induction to arrive at mathematical conclusions (not the formal method of proof by induction), but many misunderstood the use of examples and counterexamples in establishing the veracity of a statement. Although Harel and Sowder feel that examples and inductive reasoning provide students with mathematical insight, they express concern that students often do not move beyond these schemes. Finally, the researchers group other approaches as analytical proof schemes, including both transformational thinking and axiomatic methods, but found that few students master these methods. Even students that become proficient with true mathematical proofs often tend to do so only when considering familiar objects such as the set of reals. The work by Harel and Sowder suggests that since most students consider external and empirical schemes convincing they may see little need to improve their analytical proof skills, which perhaps explains the previously discussed outcomes of other research studies.

Pedagogical Aspects of Proof

The findings above raise an important concern above how the limited grasp of mathematical concepts affects the ability of college students to become effective secondary teachers after they graduate. Certainly, students who found little meaning in

the proofs they were asked to do as undergraduates will very likely become teachers who do not value proofs in secondary classrooms. Also, even if students graduated with excellent math skills, does having technical expertise with a concept necessarily imply an advanced facility in teaching that concept? Recent studies have investigated the interactions between content knowledge and teaching effectiveness, as well as revealing some of the difficulties teachers face in bringing mathematical ideas into the classroom.

Research by Ruhama Even (1993) explored the connections between a teacher's subject-matter knowledge and his or her *pedagogical content knowledge*, specifically with regard to the definition of a function. Her work found that many pre-service teachers did not possess a concept image of functions that included the essential ideas of arbitrariness and univalence, and these omissions were even more evident in the pedagogical approaches they would take with students. Instead, these prospective teachers had a tendency to provide explanations that viewed functions simply as an operational process or viewed functions as having certain (incorrect) properties, such as a "smooth" graph. Arbitrariness refers to the concept that functions are not necessarily represented by equations, formulas, smooth graphs, or a set of "known" functions. Half of the subjects indicated that equations were the dominant basis for their conception of a function. Many went so far as to say that all functions can be represented by an equation or formula. Additionally, when deciding if

$$f(x) = \begin{cases} x, & \text{if } x \text{ is a rational number} \\ 0, & \text{if } x \text{ is an irrational number} \end{cases}$$

describes a function, participants often were troubled by the fact that the graph was not continuous. Furthermore, whereas most participants correctly believed that an infinite number of functions exist passing through three given points, a significant portion also

intimated that these functions could only be chosen from certain sets or families of functions. Univalence refers to the concept that each element in the domain is mapped to only one element in the range. Even found that the prospective teachers generally incorporated the idea of univalence in their definition of a function or explanations to students, but very few could explain the reason for this requirement. Therefore, they favored the emphasis of procedural knowledge over understanding, such as using the vertical line test to determine if a relation represents a function rather than discussing the mathematical benefits of functions over other relations.

The dependence of teachers on certain procedures or formats was also noticed in other research studies. For example, Knuth (2002a, 2002b) discovered a surprising tendency to focus on the superficial details of an argument, such as the format or language used in the presentation, rather than the mathematical reasoning of the argument. The experienced teachers participating in his study often categorized proofs based on degrees of formality and they had difficulty recognizing non-proofs when the argument appeared to have an archetypal form such as proof by induction. These results are not surprising given the discussions above regarding undergraduates and proofs. Herbst (2002) goes further, suggesting that the format of proofs may actually be the reason instructors have difficulty teaching proofs. His analysis of a high school mathematics lesson indicates that presenting the traditional two-column format for proofs, usually in a geometry course, creates conflicting demands on the teacher. By exposing the limitations of the formal two-column proof, Herbst challenges mathematics educators to try alternate approaches that align more closely with current ideas about the nature and role of proof. In particular, he believes that the two-column format diminishes

the role of proof for understanding and instead reduces proofs to the memorization of a rote procedure.

Discerning teachers' beliefs about the role of proof in mathematics and in mathematics education represented another component of Knuth's study (2002a, 2002b). For example, whereas participants expressed the view that the main purpose of proof is to establish the truth of a statement, few exhibited trust in this aspect of proof. The teachers often tested mathematical statements empirically with atypical examples, even after agreeing that an argument succeeded in proving the statement for all cases. "A significant number of these same teachers seemed to believe that a proof is a fallible construct—that counterexamples or other contradictory evidence may exist—or they expressed some other measure of doubt about the generality of a proof" (Knuth, 2002a, p. 401). More importantly, Knuth found little evidence that the teachers viewed proof as a vehicle for promoting insight or understanding in the classroom. They realized that proofs give reasons why a statement is true, but did not seem to recognize that proofs can also reveal underlying mathematical principals and relationships. Hence, teachers questioned the centrality of proof in secondary mathematics classrooms. In direct opposition to the stance advocated by NCTM in *Principles and Standards for School Mathematics* (2000), participants generally relegated the teaching of proofs only to select students, such as honors students, or only in select upper-level classes, such as calculus.

Jones (2000) used concept maps to investigate the conceptions of proof held by student teachers in the UK. After creating a list of important terms, each student teacher used these terms to create a concept map representing his or her knowledge and beliefs about the interconnections among ideas regarding proof. The maps were scored and then

correlated with measures of each participant's mathematical knowledge and teaching competency. (Jones admitted that the educational community at that time was not in agreement about the scoring of concept maps.) The study found that having the most advanced technical expertise with mathematics did not necessarily translate into being the best math teacher. That is, a high level of subject matter knowledge does not guarantee the kind of knowledge needed for effective teaching, which instead requires a rich understanding of the interconnections among mathematical ideas.

Innovations in the Teaching of Proofs

The wealth of studies describing the abysmal state of proof in mathematics classrooms clearly indicates the need for new instructional approaches. Because of the student errors investigated in their two research studies, Selden and Selden (1995, 2003) believe students should be presented with opportunities for and explicit instruction in validating proofs as a means of improving their own ability to construct proofs. They also assert that logic should be taught in the context of actual proofs rather than as a separate unit preceding proofs. Among others, Bell (1976), Dean (1996), Harel and Sowder (1998), and Movshovitz-Hadar (1988) advocate the use of "proof-eliciting problems" to stimulate the type of classroom interactions and student participation that foster mathematical understanding and improved conceptions about proof.

The goal is to help students refine their own conception of what constitutes justification in mathematics: from a conception that is largely dominated by surface perceptions, symbol manipulation, and proof rituals, to a conception that is based on intuition, internal conviction, and necessity (Harel & Sowder, 1998, p. 237).

An important component of these activities is that students should not readily know the correct answer. Instead, students work individually or in groups to solve a novel problem. When they share their solutions with the class, other students may be concerned about the legitimacy of the steps taken and expect justifications or explanations, which can eventually lead to proofs. In this way, students appreciate and understand the mathematics better because they helped to create it. Moreover, students see a real need for proofs.

Despite research indicating that students often confuse empirical evidence with mathematical proof, some mathematics educators advocate the use of technology in teaching proofs (de Villiers, 1995; Touval, 1997). Graphing calculators, mathematics programs such as *Maple* and *Mathematica*, and dynamic software such as *Cabri-Géomètre* and *Geometer's Sketchpad* allow students to quickly generate numerous examples and therefore develop and test conjectures empirically. Since students find this type of evidence more convincing than a formal proof anyway, the role of proof in the classroom shifts from that of verification to that of explanation and systemization. In this way, de Villiers suggests that students gain a better appreciation for the true nature of proofs; the development of proofs in the classroom more closely resembles the development of proofs by mathematicians.

Other instructional strategies advocated by educators for improving student facility with proofs are cooperative learning and the inclusion of writing within the mathematics classroom. Goff (2002) found that using standard composition techniques, such as writing for peers, peer critiquing, and small group discussions, not only improved students' understanding of proofs but also improved overall classroom instruction. The

beliefs inherent in his study were that (a) "the ability to write mathematically is an important part of becoming a good mathematician" (p. 239); (b) that students are rarely taught how to write well mathematically, even in "bridge" courses that teach proof writing; (c) that cooperative learning or peer-assisted learning improves overall student performance and academic abilities; and (d) that using techniques typically employed in English courses for teaching composition would improve students' mathematical compositions as well. In particular, he analyzed the results of two exercises given in his Spring 2002 discrete mathematics course. In the first exercise, students individually wrote a proof for a problem in set theory. Divided into groups of four, they then critiqued the proofs of the other students in their group. Afterwards, each group presented a single collectively written proof to the instructor for evaluation. Goff deliberately chose a relatively simple problem so that students could focus on the writing of the proof rather than on the mathematical content of the problem. He conjectured that the errors he found in the final group-written proofs could have been attributed to mathematical misunderstandings of the students or to the difficulty for students of combining English phrases and mathematical symbols. In either case, he believes the errors prompted discussions that demonstrated to the class the value of peer-assisted learning.

In the second exercise, which occurred later in the course, Goff (2002) assigned each student within a group a different problem to prove, although all the problems were derived from related concepts. Thus, students critiquing another's proof would not be subconsciously inclined to fill in missing details since they had not attempted the same proof. Goff hoped this would encourage students to write a more persuasive and careful proof. Bringing two copies to class, students first turned in a copy to the instructor before

having the second copy critiqued by their peers. In this way, the instructor hopefully could better determine the effect of the peer critiques on a student's understanding of the proof by comparing the first copy of the proof turned in at the beginning of class with a new, possibly revised proof turned in after the review process. Students were more familiar with the peer critique method at this point; however, numerous common errors still showed up in the final proofs. Discovering these common errors allowed Goff to learn which concepts were more difficult for his students and adjust instruction accordingly. Also, he asserts that students enhanced their ability to write mathematically by having both to write proofs and to critique them.

Conclusion

Mathematics educators and philosophers have come to realize that proofs serve important and distinct roles within the discipline. Of course, proofs serve to verify that a statement is true and help mathematicians systemize results. But proofs are also created in a social context through the communication of mathematicians, for whom proofs serve to illuminate the nature of the mathematics underlying a theorem. It is in this context that proofs would best be utilized in the secondary classroom, that is, for demonstrating *why* a particular mathematical concept works the way it does. To teach proofs only in the context of geometry, to present only the two-column format for proofs, or to teach proofs only to a select few is remiss and untrue to the nature of mathematics in many ways.

However, the current methods of teaching proofs seem to be inadequate. Many studies have found that students and teachers emphasize form to the extent that they often cannot identify a non-proof if it follows a standard proof method. A recurrent theme is the discrepancy between what is considered a valid proof and what actually convinces

students or even teachers that a statement is true. A reasonable conclusion that arises from this research is that students are not taught to value understanding in mathematics. Instead, they are expected to *perform* mathematics well. Starting in elementary school and progressing through undergraduate mathematics, students memorize algorithms for solving mathematical problems, memorize theorems, and memorize acceptable formats for proofs but are rarely asked to understand why or how these work. Consequently, students in college mathematics programs often graduate with a limited grasp of proof. Especially with regard to proofs, they simply were not taught well. How, in turn, can those that become teachers be expected to teach well? Their experiences as undergraduates cause a focus on memorization and adherence to format, which set up barriers to understanding. Also, effective teaching requires a different kind of knowledge than subject matter knowledge, so even good mathematics students do not necessarily make good math teachers. Instead, teachers must have a well-developed concept image that includes a sophisticated network of connections among ideas. Most importantly, that concept image must include valuing proofs for understanding.

Providing future teachers with a pedagogically rich concept of proof will require attention in both undergraduate mathematics courses and teacher preparation courses (Jones, 2000; Knuth, 2002a). "In short, teachers need, *as students*, to experience proof as a meaningful tool for studying and learning mathematics" (Knuth, 2002a, p. 403). The intent of my study was to provide that experience in a teacher preparation course for secondary mathematics teachers.

Chapter 3

Methods

Because studying the development of mathematical understanding is quite complex, much of the research in mathematics education in a particular area begins with small, focused, qualitative studies. This study is no exception. Many research articles addressing issues of students' development of mathematical reasoning have used qualitative methods of study, sometimes mixed with a bit of quantitative methods (e.g., Even, 1993; Selden, A., & Selden, J., 2003). Most studies that are relevant to this work were quite specific, focusing on a particular topic (proofs) and usually restricting themselves to only one or two aspects of that topic, such as how well students unpack the logic of math statements (e.g., Selden, J., & Selden, A., 1995; Thompson, 1996). I discovered that initial formulations of this project were too ambitious, as much of the current work done in this area is the result of years of research, often starting with a Ph.D. dissertation and continuing from there. In particular, thinking about the complexities of students' difficulties with proofs is easy; however, studying those difficulties is hard. Therefore, this study represents only a small part of a much bigger picture. This project focused on exploring a particular instructional technique aimed specifically at pre-service and in-service secondary math teachers.

Action Research and Mixed Methods Design

It was important to me to do a project that took some of the theoretical results that have already been reported in the literature and applied it in a classroom setting. Thus, this study employed *action research* methods in order to determine if there is a

correlation between the development of scoring rubrics and improvements in students' ability to identify correct mathematical reasoning. Here, action research (also known as practitioner research) refers to the practice in education of investigating an instructional strategy by actually implementing it in the classroom and documenting the results. That is, action research is "learning by doing" (O'Brien, 1998, ¶ 3). However, as O'Brien states, "what separates this type of research from general professional practices, consulting, or daily problem-solving is the emphasis on scientific study, which is to say the researcher studies the problem systematically and ensures the intervention is informed by theoretical considerations" (1998, ¶ 5).

An important aspect of practitioner research is that the study leader has a first-hand and often tacit knowledge about the area of concern. True practitioner research develops from an instinctual professional frustration or dilemma personally connected to or perceived by the researcher. Although I have limited experience with teaching proofs directly, I can say that the vast majority of students I have worked with do not possess good mathematical reasoning skills and probably would be unable to write or validate proofs. Moreover, in the graduate math courses I took, many of the other graduate students did not exhibit the attention to detail necessary for proof construction and validation. I often found myself pointing out glaring gaps in their logic. Thus, although I do not have practitioner experience with the difficulties of teaching proofs, my other experiences align well with the research previously cited about the wide-spread problems students have constructing mathematical arguments.

The sample size for this project made purely quantitative methods inappropriate. Also, quantitative methods would not have been helpful in clarifying the connection

between my intervention techniques and any improved performance by students. Thus, a mixed methods design was chosen for the research, which combined both quantitative and qualitative procedures. The quantitative portion of a mixed methods design serves to form the skeleton of the research, while the qualitative portion adds “flesh” to the study. That is, the quantitative data addresses the basic question: “Does this instructional strategy help?” However, the qualitative methods provide a more in-depth picture of when and how it might help. Changes in the performance of individual students were tracked from the Pre-Test through the qualitative data measures to the Post-Test.

Utilizing the framework suggested by Creswell (2003), the following four components guided the mixed methods design:

- Theoretical Perspective

The study emanated from gaps and suggestions found in the math education literature, as well as my own observations, about difficulties students have with proofs. In particular, few studies had investigated any intervention techniques aimed to improve student understanding of proofs.

- Implementation Sequence

The data collection method was sequential, but alternating. That is, data were collected over a period of time rather than during a single event, and in general, the quantitative and qualitative portions of the study did not occur simultaneously.

- Priority

The strength of the study lay in the qualitative data measures because the sample was too small to make meaningful statistical inferences.

- Integration

The data were “mixed” during the analysis phase of the study. That is, the quantitative and qualitative data generally were not considered together until analyzing the results at the end of the study.

Participants

Subjects in the study were students enrolled in Math 338: Mathematics for Secondary Teachers at the University of New Mexico during the Fall 2003 semester. This course explores secondary mathematics topics from an advanced standpoint and is designed to meet the needs of pre-service and in-service teachers; in fact, the course is open only to prospective and in-service teachers of secondary mathematics. All students in the course were invited to participate in the study, and all students accepted the invitation by signing the Consent to Participate in Research form (Appendix C). Therefore, sixteen subjects participated in the research study (six women and ten men). The research, however, did not consider the variables of gender, race, or ethnicity.

The participants in the course had a wide variety of mathematical experiences and abilities. Some were “typical students” who were finishing their first degree and considering the teaching profession. Other study members had been in other careers for many years and had returned to school to become mathematics teachers. Some participants had completed calculus (three semesters), differential equations, and linear algebra. Others had not been in a math class for five or more years. Some had even taught math before, or were currently teaching math, but needed to complete requirements for certification in New Mexico. This information about the background of the participants was collected anecdotally through class conversations; it was not requested of the

participants as part of this study. In fact, in order to ensure confidentiality, no identifying information was collected from participants besides their name and gender. Therefore, only the student's current work in the course were considered and discussed in this study.

For collaborative work during the study, participants were not assigned to particular groups in class. Instead, they established informal fluid groups. That is, students were allowed to work in groups of their own choosing, which varied from week to week and ranged in size from two to six.

All participants reported having some previous experience with scoring rubrics. For nearly all of the students, this experience was limited to having some of their papers in English or history courses graded using some type of rubric. No subjects reported having developed and used a scoring rubric themselves. Prior to starting the study, I had scored a couple of homework assignments in Math 338 using a rubric, and participants were given copies of the rubrics used along with the graded assignments. In general though, participants had very little exposure to the processes by which scoring rubrics are developed and implemented.

Instruments, Data Sources, and Artifacts

Quantitative data in this study resulted from two researcher created instruments: the Pre-Test (Appendix A) and the Post-Test (Appendix B). These tests sought to measure a participant's ability to determine the correctness of proofs, to specify the logical structure required for a proof, and to point out the specific errors in an invalid proof. Most of the questions on the two tests presented a mathematical argument purporting to prove a given statement and then asked for two responses. First, participants were to choose which of four possible statements best described the

argument. Then, participants were to write an explanation for their choices. A few of the questions on each test had a slightly different format or only asked for one of these two types of responses, since some questions did not lend themselves to this format. In total, the Pre-Test contained ten open-ended items asking for examples or explanations and eight multiple-choice items. For simplicity, the one true-false item was counted as a multiple-choice item. Because of a change in the direction of the study (described in Chapter 5: Discussion), only 13 items from the Pre-Test also appeared on the Post-Test, including seven multiple-choice items and six open-ended items. The Post-Test then introduced four new multiple-choice or true-false items, and four new open-ended items. These new questions helped to ascertain if participants simply remembered the Pre-Test questions or could demonstrate achievement on new questions. On the Post-Test, the order of the questions from the Pre-Test was changed and intermixed with the new questions.

The qualitative data in this study came from a variety of sources that arose from instructional activities in class, as well as my own field observation notes as the teacher-researcher. Data items collected from student-participants included scoring exercises using the draft rubrics, participant notes taken in class, and responses to reflective writing prompts. Only photocopies of student work were used for the study, and I applied removable (Post-It®) tape to cover over any names prior to photocopying, replacing them instead with pseudonyms as described in the Validity section below. A complete list of all data items, artifacts, and instruments incorporated in the study is given below. The following data items were collected from participants:

1. Pre-Test [Week 5]
2.
 - a. Participant class notes [Week 8]
 - b. Initial scoring exercise for Arguments M and L in class [Week 8]
 - c. Response to Prompt 1 in class [Week 8]
3.
 - a. Participant class notes [Week 9]
 - b. Scoring exercise for Arguments E and L in class using Rubric Draft 1 [Week 9]
 - c. Response to Prompt 2 in class [Week 9]
4. Scoring exercise for Argument M using Rubric Draft 1 [assigned Week 9; collected Week 10]
5. Scoring exercise for Argument J in class using Rubric Draft 2 [Week 10]
6. Scoring exercise for Arguments G and K [assigned Week 10; collected Week 11]
7. Scoring exercise for Arguments P and Q in class [Week 13]
8. Response to Prompt 3 [assigned Week 13; collected Week 14]
9.
 - a. Post-Test [Week 14]
 - b. Response to Prompt 4 on Post-Test [Week 14]

The following instruments and data items were created and/or collected by me as the teacher-researcher:

1. Pre-Test [Week 5] (Appendix A)
2. Field Notes [ongoing]
3. Rubric Draft 1 [Week 9] and Rubric Draft 2 [Week 10] (Appendix D)
4. Examples of other rubrics [Week 9] (Appendix E)

5. Arguments L, M, E, J, G, K, Y, P, and Q, in that order (Appendix F)
6. Prompt 1 [Week 8] , Prompt 2 [Week 9], Prompt 3 [Week 13], and Prompt 4 [Week 14] (Appendix G)
7. Post-Test [Week 14] (Appendix B)

Procedures

Students in Math 338 were notified about the study on the first day of class. Initial approval for the study was granted by the IRB at the University of New Mexico prior to the first day of class. Students were informed both verbally and in writing about the purpose, procedures, and potential risks of the research, as well as the voluntary nature of participating in the study and the process for voluntarily ending their participation. All students in the course were invited to participate, and all students agreed to participate by completing a signed Consent to Participate in Research form (Appendix C). The study began during the fifth week of class with the Pre-Test. The Pre-Test was not timed, and participants took from 30 minutes to one hour to complete the Pre-Test. Three weeks later and continuing for the next five weeks, the teacher-researcher led the class in discussing, developing, experimenting with, and reevaluating a scoring rubric for mathematical arguments purporting to be proofs. The class met only once per week for 2-1/2 hours per session, and the discussions and work surrounding the rubric's development generally took 30 to 50 minutes of that time. The rubric's development involved researcher-led collaborative discussions involving all participants, individual writing assignments, opportunities to experiment with the rubric individually and collaboratively, and a few short homework assignments. On the 14th week of class, the Post-Test was administered and two final writing assignments were collected. Like the

Pre-Test, the Post-Test was not timed, and participants took from 30 minutes to an hour to complete it.

Much of the preliminary work in developing the rubric centered around arguments attempting to prove the statement “For any positive integer n , if n^2 is a multiple of 3, then n is a multiple of 3.” This statement and Arguments J, G, and K considered in class were modified with permission from the work by Selden and Selden (2003). Using these arguments as examples, I then created Arguments L, M, and E, also for the statement “For any positive integer n , if n^2 is a multiple of 3, then n is a multiple of 3.” Finally, I created 2 additional statements and 3 associated arguments. Argument Y represented an attempted proof of the statement “If a number is not divisible by 2, then it is not divisible by 6.” Arguments P and Q represented proof attempts for the statement “In a triangle with side lengths a , b , and c , if $a^2 + b^2 = c^2$, then the triangle is a right triangle.” All arguments used while developing the rubric are given in Appendix F.

Timeline for the Study

The timeline below identifies the specific activities that occurred each week and the associated data items that were collected from each participant. The researcher’s field observation notes were ongoing and are therefore not included in the list. Artifacts such as Rubric Draft 2, Argument K, and Prompt 1 are contained in the appendices, as are the Pre-Test and Post-Test. References to the associated appendices are given above. The word “we” refers to the instructor-researcher and the student-participants collectively. Most activities occurred as typical instructional pieces of a normal collaboratively-constructed (learner-centered) math course, with dialogue from both the instructor and the students.

- Week 1: Students were informed about the study.
Item(s) Collected: None
- Week 2: Students were provided more information about the study.
Item(s) Collected: None
- Week 3: Students were given informed consent forms. Most were collected immediately.
Item(s) Collected: Consent to Participate in Research form
- Week 4: All students elected to participate in the study and returned an informed consent form.
Item(s) Collected: Consent to Participate in Research form
- Week 5: The study began, and the Pre-Test was administered without a time constraint.
Item(s) Collected: Pre-Test
- Week 8: The students were asked to score Arguments L and M using any method. We discussed their methods and results. Students completed written responses to Prompt 1, and we discussed their responses on a voluntary basis. We began discussing what components were desired in a scoring rubric.
Item(s) Collected: Participant notes, scoring exercise for Arguments L and M, response to Prompt 1
- Week 9: We discussed different types of scoring rubrics and looked at a few examples. I clarified the type that I intended for us to develop. We identified some of the components in a mathematical argument to be evaluated by a rubric. From this, we developed Rubric Draft 1. Participants used Rubric Draft 1 to score Argument E collaboratively. They scored Argument L individually, then

collaboratively. Participants completed responses to Prompt 2, and were assigned homework – re-score Argument M using Rubric Draft 1.

Item(s) Collected: Participant notes, scoring exercise for Arguments E and L, response to Prompt 2

Week 10: We discussed participant concerns about the use of rubrics to score mathematical arguments. I informed students that they are welcome to disagree with my perspective. Students turned in their scoring exercise for Argument M using Rubric Draft 1. Student work from the previous two sessions was returned, which prompted discussion leading to the development of Rubric Draft 2. Participants worked collaboratively to score Argument J using Rubric Draft 2. After discussing the results, participants were assigned homework – score Arguments G and K using Rubric Draft 2 and provide explanations for the scores.

Item(s) Collected: Scoring exercise for Argument M

Week 11: Students turned in their scoring exercise for Arguments J, G, and K. Further discussions were delayed until the next week.

Item(s) Collected: Scoring exercise for Arguments J, G, and K

Week 12: Student work from the previous two weeks was returned. Participants' questions and concerns were addressed, leading to further discussions about mathematical arguments and about rubrics. In particular, participants wanted feedback about how I would score the previous arguments. We had further discussions regarding the idea of overall logical structure and developed an

outline for a proof of the statement “Any point on the perpendicular bisector of line segment AB is equidistant from A and B.”

Item(s) Collected: None

Week 13: We discussed the use of the rubric for arguments with more complicated logical structures and considered Argument Y. Participants then worked collaboratively to score Argument P and individually to score Argument Q using Rubric Draft 2. We discussed the results of this scoring exercise and addressed further concerns about how to apply the rubric. Participants were assigned homework – complete a response to Prompt 3.

Item(s) Collected: Scoring exercise for Arguments P and Q

Week 14: Students turned in their responses to Prompt 3. The Post-Test was administered without a time constraint, and participants completed responses to Prompt 4 at the end of the test. The study concluded.

Item(s) Collected: Response to Prompt 3, Post-Test, response to Prompt 4

Scoring and Coding of the Pre-Tests and Post-Tests

The arguments appearing on the Pre-Test and Post-Test were crafted to contain only one basic flaw so that the multiple-choice items consisted of mutually exclusive choices regarding the validity of an argument. That is, only one correct answer existed for each multiple-choice item on the Pre-Test and Post-Test. My initial review of the tests therefore looked at simply the correctness of each participant’s multiple-choice responses. (Note: The Pre-Test also contained items pertaining to the original thesis involving examples and counterexamples, described in Chapter 5: Discussion. These

items were ignored for this study, which considered only the items pertaining to validating mathematical arguments.)

Once the multiple-choice responses were scored, I considered participants' open-ended responses. After a cursory look at these responses, a descriptive system was developed for coding the open-ended responses, which is given below. The Pre-Tests were scored and coded before any other aspects of the study began, such as developing the rubric in class. Open-ended responses that earned a code of I, II, or III were counted as "correct" and are reflected as such in Chapter 4: Results. The Post-Tests were scored and coded in the same way, without looking back at participants' results on the Pre-Tests. After coding responses on the Post-Tests, however, I then verified my coding by cross-referencing the responses and codes applied on the Pre-Test. The following codes were used:

- I. The multiple-choice answer is correct. The reasoning in the explanation is correct and clearly worded.
- II. The multiple-choice answer is correct. The reasoning in the explanation seems to be correct but the wording is less than clear.
- III. The multiple-choice answer is incorrect. However, the reasoning in the explanation is correct and clearly worded (possibly indicating a misunderstanding about the answer choices rather than the argument and its error).
- IV. The multiple-choice answer is correct. However, the reasoning in the explanation is incomplete, flawed, or irrelevant, or the wording is very unclear. (Thus, I could not tell if the participant had correctly identified the

error but had difficulty explaining it or if the participant simply guessed correctly.)

- V. The multiple-choice answer is incorrect. However, the explanation contains at least some correct reasoning or some correct and relevant knowledge statements.
- VI. The multiple-choice answer is incorrect. Also, the explanation contains mostly incorrect reasoning or irrelevant statements.
- VII. No explanation is given.

Analyzing the Qualitative Data

Qualitative data sources consisted of my field observation notes, as the instructor-researcher, and items collected weekly from student-participants, such as responses to reflective writing prompts, practice scoring exercises using the draft rubrics, and participant notes taken in class. The major analysis technique was identification of themes. I reviewed these data sources in a variety of ways. I identified themes in my researcher field notes. I also studied five particular cases, looking for themes within each case's work and across the five case studies. Finally, I combined the work from my field notes with that from the case studies.

Validity

Several measures were taken to increase the validity of the study. First and foremost were the steps taken to ensure confidential and voluntary participation in the study. From the first day of class, students in the class were informed that the study would take place throughout the semester. All students were invited to participate in the study, although the instructor repeatedly stressed both verbally and in writing that neither

participation nor lack of participation would have any effect on a student's grade in the course. All students in the course were required to complete the Pre-Test, prompt-response writing assignments, rubric scoring exercises, and Post-Test as normal instructional activities of the class. These un-graded assignments were collected as part of the course expectations without regard to a student's participation in the study. However, a student's work was only used in the research study if he or she allowed me to do so by signing and returning the Consent to Participate in Research form (Appendix C). Participating in the study neither increased nor decreased the amount of work students had to do in the course, nor did it improve or change their grades in any way. Participation simply meant that a student gave me permission to use his or her work in my research study. Fortunately, all students in the course agreed to participate.

Protecting the privacy and confidentiality of all participants was of utmost importance so as to increase participants' trust in the study and willingness to be open and honest in their work. No personal or identifying information was collected from participants except their names and gender. Once students had agreed to participate in the study, a hand-written list was generated that associated each participating student with a pseudonym, which was used throughout the study to analyze a participant's work over time. The pseudonyms were random generic student names, none of which were equivalent to the names of any students in the class. Any and all references in the study to participating students are pseudonyms. Only photocopies of student work were used for the study, and I applied removable (Post-It®) tape to cover over any names prior to photocopying, replacing them instead with pseudonyms. Thus, the sole connection between the data collected for the study and actual student names was the temporary list

generated at the beginning of the study correlating participants with a pseudonym. These procedures applied to all student work and my own field observation notes.

The study's design was also meant to increase its validity. Data was collected over a period of time, rather than during a single event, which added depth and reliability to the results. Moreover, the variety of data measures, both quantitative and qualitative, allowed for triangulation of results. In particular, the qualitative data added depth and understanding to the quantitative results. Furthermore, collecting the notes that participants took as we developed the rubric helped me verify my perceptions of the discussions. Also, participants were not given a time limit in which to complete the Pre-Test or Post-Test. Therefore, these tests were not a measure of speed and hopefully measured a participant's knowledge accurately

Finally, the methods used to score or code participant responses added to the validity of the study. For example, when coding the open-ended responses on the Pre-Test and Post-Test, I coded responses for the same question on all tests before moving to the next question so as to be more consistent in the coding. Also, I first coded the Post-Test independently of the Pre-Test, but then verified and aligned my coding by cross-referencing the responses and codes applied on the Pre-Test. The coding for all of the open-ended responses on the Pre-Test and Post-Test was checked at least three times in a variety of ways to ensure consistency and accuracy. The content analysis of the various data sources in the rubric developmental work was also checked a number of times, looking for themes both across activities and across individuals. Finally, the statistics describing the results of the Pre-Test and Post-Test were checked using a calculator as well as through the use of a spreadsheet.

Chapter 4

Results

As stated in Chapter 3: Methods, the framework guiding my mixed methods design initially treated the quantitative and qualitative data separately. Hence, they are treated separately in this chapter. In Chapter 5: Discussion, the data are combined to provide a more integrated analysis of the results.

Quantitative Results

As mentioned previously, the Pre-Test contained extraneous questions relating to the original, unused thesis involving examples and counterexamples. Of the remaining questions, seven were multiple-choice items (including the one true-false item), and six were open-ended items asking for clarification about the multiple-choice responses. All 13 of these items were used on both the Pre-Test and the Post-Test (Appendixes A and B, respectively). Figures 1 and 2 compare the number of correct responses, out of the 16 participants, on the Pre-Test versus the Post-Test for each of these items. Figure 1 considers the multiple-choice responses, and Figure 2 considers the open-ended responses.

The total number of correct multiple-choice responses increased from the Pre-Test to the Post-Test, although this increase was not statistically significant. Considering all 16 participants, 112 total correct answers were possible for the multiple-choice responses. On the Pre-Test, there were 46 correct multiple-choice responses. On the Post-Test, this amount increased minimally to 48, which is a 1.8% increase out of the total possible. The increase in the total number of correct open-ended responses was more pronounced.

Considering all 16 participants, 96 total correct answers were possible for the open-ended responses. On the Pre-Test, there were 34 correct open-ended responses, whereas on the Post-Test, this amount increased to 41. This represents a 7.3% increase out of the total possible.

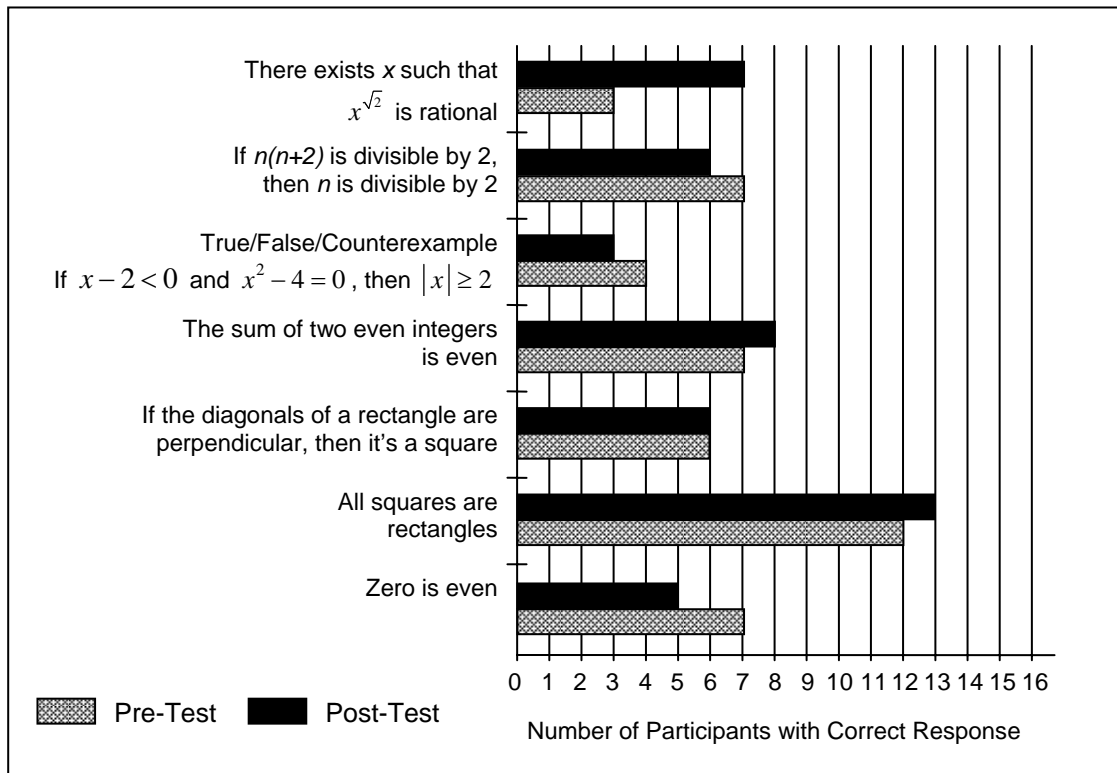


Figure 1. Number of correct multiple-choice responses by item on Pre-Test versus Post-Test.

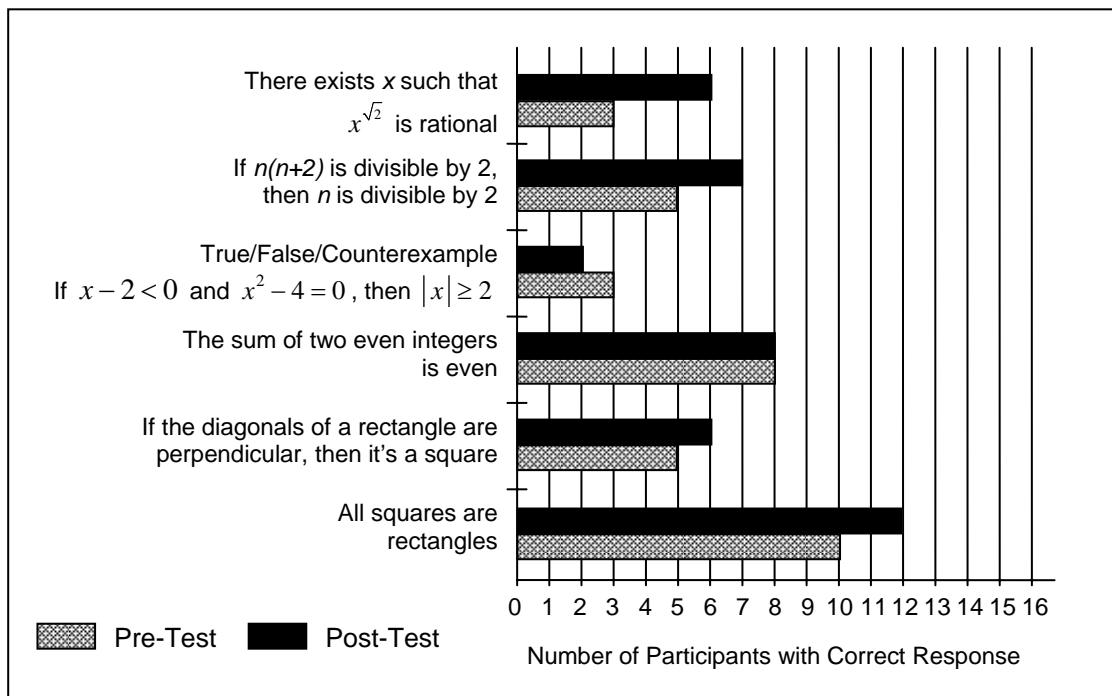


Figure 2. Number of correct open-ended responses by item on Pre-Test versus Post-Test.

Figures 3 and 4 display the number of participants earning a given score on the multiple-choice items of the Pre-Test and Post-Test, respectively. Figures 5 and 6 display the number of participants with a given score on the open-ended items for each test. Here, a participant's score simply means the total number of correct responses for the indicated portion of the test. As stated before, open-ended responses that received a code of I, II, or III were counted as correct.

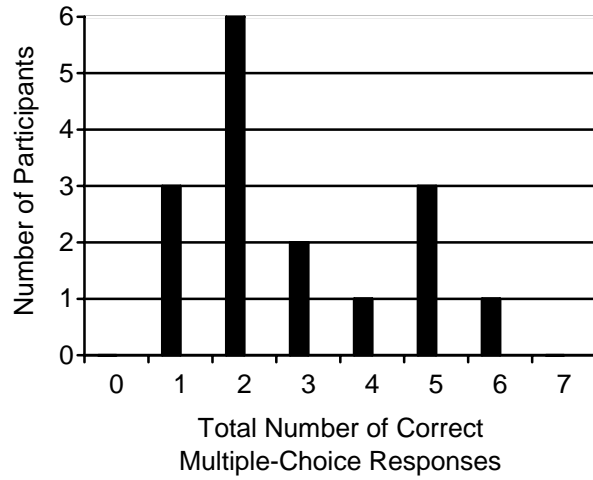


Figure 3. Number of participants with a given score on the multiple-choice items of the Pre-Test.

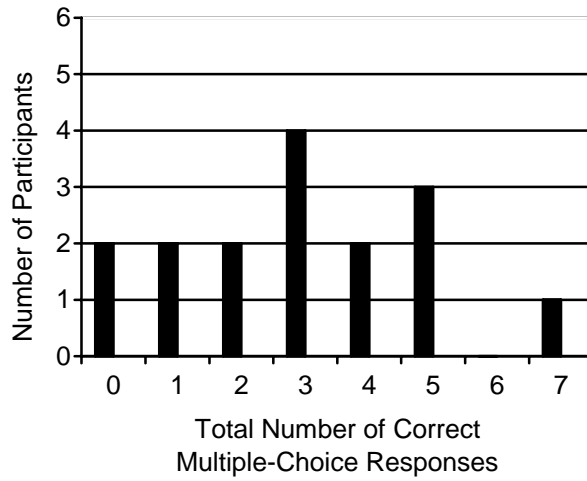


Figure 4. Number of participants with a given score on the multiple-choice items of the Post-Test.

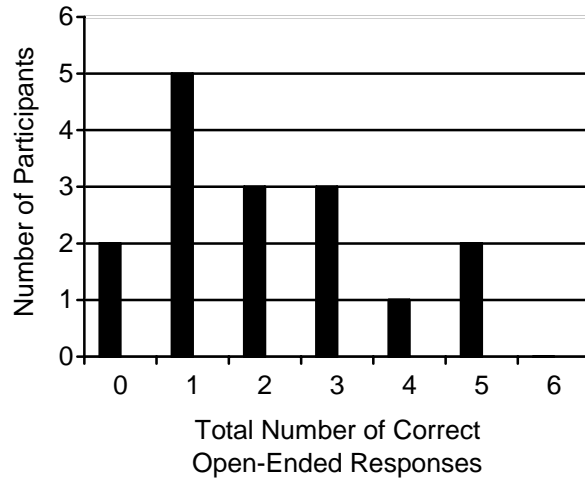


Figure 5. Number of participants with a given score on the open-ended items of the Pre-Test.

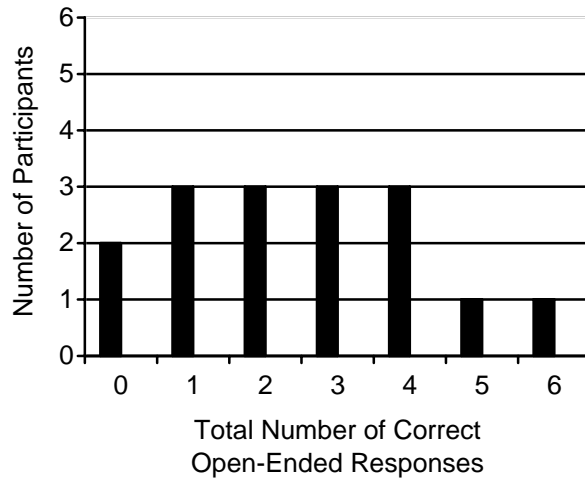


Figure 6. Number of participants with a given score on the open-ended items of the Post-Test.

Considering the multiple-choice responses, 8 student-participants improved their score from the Pre-Test to the Post-Test, 7 participants scored lower, and 1 participant made the same score. Considering the open-ended responses, 7 students improved from the Pre-Test to the Post-Test, 5 students scored lower, and 4 students made the same score. Overall, considering both multiple-choice and open-ended responses, 8 participants improved, 7 scored lower, and 1 stayed the same. One of the most notable results, however, is that no participant did better with multiple-choice responses and worse with open-ended responses, or vice versa. That is, if a student's score on the multiple-choice responses increased from the Pre-Test to the Post-Test, then his or her score on the open-ended responses stayed the same or also increased. This was also true of decreases.

Although student-participants did indicate an overall improvement on both the multiple-choice responses and the open-ended responses, paired *t*-tests indicated no significant difference existed between the results of the Pre-Test and the results of the Post-Test on the 13 common items. The *p*-value for the multiple-choice responses was 0.827, for the open-ended responses was 0.300, and for the two responses combined was 0.551, all of which were above the desired significance level of 0.05. Thus, the results for specific individuals are more notable in regards to this study. These are explored later.

Participants seemed to handle the new questions on the Post-Test a bit better than the questions that appeared on both tests. Of the multiple-choice items appearing on both tests, 41.1% of the responses were correct on the Pre-Test, and 42.9% were correct on the Post-Test. Of the open-ended items appearing on both tests, 35.4% of the responses were correct on the Pre-Test, and 42.7% were correct on the Post-Test. However, 48.4% of the

multiple-choice responses and 48.4% of the open-ended responses were correct for the eight new items appearing only on the Post-Test.

Most of the purported proofs or arguments presented on the Pre-Test and Post-Test were followed by both a multiple-choice item and a corresponding open-ended item. For those arguments appearing on both the Pre-Test and the Post-Test, the difference between the number of participants giving a correct multiple-choice response and the number of participants giving a correct open-ended response was no more than two. However, the discrepancy between multiple-choice responses and open-ended responses was greater for some of the new arguments presented on the Post-Test. For example, considering the argument in items #10 and #11 of the Post-Test (involving the statement “If triangle ABC is an isosceles triangle with congruent sides \overline{AB} and \overline{AC} , then the base angles ($\angle ABC$ and $\angle ACB$) are congruent.”) only 2 participants gave a correct multiple-choice response, but 7 gave a correct open-ended response explaining the flaw in the argument. Conversely, for the argument presented in item #19 of the Post-Test (involving the statement “If x is divisible by 5 and y is divisible by 3, then $x + y$ is divisible by 8.”) 15 of the 16 participants correctly identified the argument as not being a valid proof, but only 10 gave a correct explanation for the error.

Qualitative Results

Qualitative data sources consisted of my field observation notes, as the instructor-researcher, and items collected weekly from student-participants, such as responses to reflective writing prompts, practice scoring exercises using the draft rubrics, and participant notes taken in class. The primary analysis was reviewing the work for themes, which revealed that three particular challenges were evident during the study. First, the

participants and I had to develop a common understanding of the type of rubric we would create. Second, many participants focused on the rubric as a vehicle for assigning a grade rather than as a tool for measuring what a student understands mathematically. Third, some participants revealed incomplete proof conceptions and struggled to look beyond the superficial aspects of mathematical arguments, such as formatting and notation.

Initial confusion about rubrics. The first discussion regarding scoring rubrics occurred during the 8th week of class, three weeks after the Pre-Test. This initial discussion and participants' written prompt responses revealed that, whereas most participants had a basic grasp of the intent of rubrics, they had differing but minimal experiences with rubrics. Moreover, none had seen the use of scoring rubrics in a math classroom besides two homework assignments I had graded using a rubric early in the course. Thus, an initial obstacle to our work was developing a common understanding of the type of scoring rubric we would create.

In response to the question "What is a rubric?", the following verbal answers arose indicating an understanding of the intent of rubrics:

- a matrix style scoring device
- a standard set of checkpoints to be compared with students' answers
- an outline of a grading process
- a guideline or framework so the teacher grades consistently
- a set of clear expectations or criteria for a given score

A theme that clearly surfaced, therefore, was that most participants believed the purpose of a rubric was to reduce subjectivity and increase consistency. Also, participants felt strongly that rubrics can be very beneficial in letting students know what is expected of

them, provided the rubric is given to students before an assignment is completed. The participant Kim stated that a rubric “provides info on what areas are strengths and what areas need more work.”

However, participants generally did not have the same idea as I did about the format for a rubric. Most had only seen the holistic type of rubric that describes all the qualities necessary for a certain grade. Instead, I envisioned the type of rubric that delineates the specific components or aspects to be considered in scoring student work and assigns points along a specified scale for how well each aspect is evidenced in the work. Also, different aspects may have different scales or weights. I discussed three general aspects of student work that I try to assess when using scoring rubrics: accuracy of the mathematics involved, depth of understanding or mastery, and quality of the presentation.

Knowing that many different ideas existed about what we were attempting to do, I started our second week of work with the rubric (Week 9 of the study) by having more discussions about rubrics. I categorized three different styles of rubrics by way of examples. The three styles specified below are not technical labels, but simply the locally-defined labels used for the purposes of the discussion with the study participants.

1. Holistic: This style of rubric describes the general characteristics of the entire work necessary to attain a certain grade. Two examples I presented of this type were the 3-Point Rubric for Medium Constructed Response Items from the Colorado Department of Education (2003) and the Rubric for Grading Daily Work (Appendix E) that I had previously developed for my own high school courses.

2. Matrix: This style of rubric presents an array that separates the various general aspects to be considered when scoring a student's work and then describes the qualities of a response exhibiting differing degrees of accomplishment for each aspect. The weight of different aspects towards the final score can be varied. The aspects considered are often very general, such as conceptual understanding, procedures or strategies, and presentation or communication. The example I gave participants of this type was the Classic Math Rubric from Exemplars® (2003).
3. Outline-and-Point-Value: This style of rubric gives a detailed list of the specific components or aspects to be considered when scoring a student's work and then has a numerical scale along which each aspect is scored. Like the matrix style, the weight of the different aspects can vary. However, the aspects identified are quite specific to the assignment and not as general as those in a matrix style rubric. Because this style delineates the scoring items more fully, written descriptions of the reasons for a certain score are not given. The Bulletin Board Project scoring rubric (Appendix E) developed for my high school courses served as an example of this style of rubric.

I stated that we would be developing a rubric using the third or outline-and-point-value style. The explanation about the different styles, along with the examples presented, seemed to solidify the class' understanding of my goal. After the second week of discussions with examples, participants seemed to understand the type of rubric we were developing. They did not seem to dispute or struggle with the type of rubric, only how the rubric was being used and if a rubric was at all appropriate.

Grade issues. To start our discussions (Week 8) about using rubrics to score arguments, I asked participants to consider Arguments L and M (Appendix F). Specifically, I asked them, on their own, to decide if each argument represented a valid proof and then to determine a grade for each. Participants could use whatever scoring system they wanted, as long as they informed me of the scale and tried to give reasons for the scores they assigned. That is, they could give the argument a standard letter grade or assign it some number of points out of a total number of possible points (e.g. – the argument merits 16 out of 20 points). Initially using the word “grade” instead of “score” created an obstacle that was difficult to overcome for several participants. I did not make a distinction between these terms until later when the issue of using a rubric to assign grades arose. I definitely should have been more careful about this from the beginning. However, I could also argue that trying to make a distinction for this initial activity would have ruined its impact and spontaneity.

Throughout the following weeks’ discussions, the issue about grades became a recurring theme. Beginning with our second discussion (Week 9), most students assumed that the score from the rubric would be translated directly into a grade. For example, if an argument earned 10 out of 20 points on the rubric, then the student who wrote the argument would get a grade of 50%, or a failing grade. I stated that this did not have to be the case. Instead, the score could be scaled or augmented in some way to produce a grade more in line with teacher or class expectations. For example, a teacher could give a student a “completion grade” of 50% for doing the work and then add on points based on the score from the rubric. That is, if an argument earned 10 out of 20 points on the rubric, the author of the argument could receive a grade of $50 + 50 (10/20) = 75\%$, or a C. Over

the next two weeks, I repeated the point that grades are more of a local issue that teachers could deal with in their own way. Instead, we were trying to create a rubric for evaluating a student's understanding of proofs so that teachers could more easily assign the grades, as they see fit, for complex tasks such as writing mathematical arguments.

Despite my explanation on several occasions that grades were more of a subjective issue, David represents an example of a participant who struggled with the grade issue throughout most of the study. On the first attempt to score arguments, which was before developing the rubric, David understood that Argument M was better than Argument L, and even knew basically where the errors were in each. After our first development and use of a rubric during Week 9, David's conflict about grades came to light. In particular, he felt that his original score for Argument L was more accurate than that given by the rubric because he thought of the rubric's score as an exact grade to be given (see previous paragraph). His original grade was D-, whereas the rubric gave a "grade" of F-, which he thought was unfair. "I would not give an F- for this effort even though it is poor," David remarked. Despite his issue with grades, he succinctly states the main error in the argument: "doesn't understand induction." He was also quite consistent with his classmates in his scoring using the rubric.

In the next homework assignment, looking at Argument M again with the draft rubric, David continued to interpret the raw score from the rubric as the actual grade despite what I had said in class. In this assignment, he saw that dividing by zero would be a problem and even realized that the original statement negated this possibility. That is, he looked very carefully at the argument and had enough of a grasp of mathematics to identify a subtle point often missed by students. However, he overlooked the real error in

the argument and instead deducted points incorrectly for a wording issue, possibly indicating that his problem with the grade seemed to distract him from correctly analyzing the argument. During Week 10, I had another discussion about grades, but David continued to indicate a struggle with grades in the homework assigned afterwards. In his analysis of Argument G, he was distracted by the fact that his interpretation of the rubric led to penalizing a student twice for the same mistake, thereby giving a grade he felt was unfairly low.

There is a marked lack of discussion about grades in his subsequent scoring of Argument K. Perhaps this is due to the fact that the proof was basically correct and his notion of the appropriate grade coincided with the “grade” given by the rubric. By Week 13, David seemed to be doing better with the rubric. His scorings for Arguments P and Q were nearly identical to mine. There are at least two possible explanations: the arguments contained much neater errors, or David was actually becoming more comfortable with the rubric. If the latter, the results of his Post-Test indicate that his facility with the rubric perhaps occurred a bit too late.

Another important participant concern related to grades was the possibility that the rubric would cause an incorrect argument to be penalized twice for the same mistake. Participants, especially David, Sean, and Chris, became very engrossed in the discussion about where to count off for a particular mistake so as not to cause the “double whammy” effect. As Chris stated during Week 9, “That’s not fair to take off twice.” I acknowledged that this may occur sometimes, but our job was to delineate the components or aspects of an effective argument well enough in the rubric so as to minimize this occurrence. The on-going development of the rubric focused on clearly separating critical yet distinct

aspects of a mathematical argument. Serendipitously, the concerns about the double-whammy effect forced us to create a better rubric and forced participants to identify the errors in each argument more carefully. When group members disagreed about the score for a given argument, a common phrase was “Let’s start at the top of the rubric.”

Proof Conceptions. Even the initial development and use of a rubric made students aware of how difficult scoring proofs will be as a teacher. Participants understood that teachers must have a solid grasp of their expectations for student work in order to create an effective rubric. As Sean stated, “Assessing students’ understanding is harder than it sounds.” Students never explicitly made the connection that scoring arguments was difficult because they did not fully understand proofs themselves, but the rumblings of such a connection were apparent in participants’ self-reflection, such as those associated with Prompt 2 (Appendix G).

During Week 9, after clarifying the type of rubric to use, participants were asked to delineate the components or aspects we should look for in an argument. Nearly all began by listing superficial format aspects, such as “states the given,” “states the conclusion,” and “uses two columns to list steps and associated reasons.” These reflect the findings of other researchers who discovered that students focus on specific details rather than the overall picture (Knuth, 2002a, 2002b; Selden, A., & Selden, J., 2003). Moreover, very few student-participants recalled having seen a paragraph-style proof, and none had actually written an argument formatted in paragraph form. This suggests that participants have not needed to demonstrate any level of proficiency with proofs since their high school geometry class, where the two-column approach is often regarded as the only way to present an argument. This style starts by stating the given and ends by

formally restating the conclusion deduced. One column contains the mathematical statements that result from previous statements, and the other column provides the theorem or reason behind each step.

I did not get a formal assessment of students' proof conceptions. My sense was that they vaguely understood proofs to be absolute arguments, but how deeply they truly grasped this idea was unclear. The generality necessary in a valid proof was discussed on several occasions. Although students seemed to know superficially that proofs must be general, I was never quite convinced they truly understood how an argument establishes this trait. That is, they seem to embrace the idea that proofs need to be general, whether or not they could recognize when an argument accomplished this.

After some lengthy discussion, participants eventually were able to identify some of the deeper, more mathematical aspects of a valid proof. These aspects resulted from the need to delineate the components to include on the rubric.

- proves the general case
- each step is valid
- provides clear connections between steps or ideas
- "goes in the right direction"

However, several students continued to struggle with a faulty understanding of proofs. Jennifer, for example, clearly had exposure to proofs in college math courses. She wrote about the difficulty she was having with proofs in her real analysis course, and she was familiar with terms such as "conditional statement" and "contrapositive."

Throughout Weeks 9 to 13, though, Jennifer focused primarily on presentation aspects of arguments. When her explanations were content-oriented, they were often confusing or

off track. Rarely did she catch the most critical flaws of an argument, especially if they involved the overall structure, direction, or “flow” of an argument. Chris, John, and Eric were other participants who had difficulty with the overall logical structure of arguments, as well as some of the algebra and number sense required for understanding the arguments.

Despite these struggling students, however, the discussions had a different tone even after the first draft of the rubric. Rather than focusing on format, many participants began their analysis of an argument by looking at the first item on the rubric, overall logical structure.

Chapter 5

Discussion

In general, the study had very mixed results. Statistically, the quantitative data indicates no significant improvement occurred in participant responses from the Pre-Test to the Post-Test. Tracking the progress of several students through the qualitative data sources reveals several potentially complex interactions occurring for participants. For some students, the development of the rubric seemed to help them be more attentive to the important aspects of a mathematical argument, including the overall logical flow or structure of an argument. This was especially true for Maria and Julie, who more than doubled their number of correct responses on the items appearing on both tests. For other student-participants, such as Jennifer, the rubric seemed to confuse them by adding another layer of complexity. Still other participants, who seemed quite mathematically capable based on their comments and insights in class, seemed to get caught up in the grading aspect of the rubric. Of note here are David and Sean. However, several confounding factors in the study could be avoided in future work, suggesting that further research of the attempted instructional technique (development and use of scoring rubrics as a method to improve teachers' ability to validate proofs) may prove valuable.

Complex Interactions

Based on work collected from participants for the study, as well as other work not affiliated with the study, I had a tacit sense of each individual's mathematical ability. A satisfying aspect of the study was that the variety of abilities of participants made for a good sample. I was able to look at the impact of the instructional technique on a wide

range of mathematical competencies. One student had never seen proofs of any kind in his math classes. Another student was in a real analysis course that heavily involved proofs. Almost none of the members of the study, however, seemed to have been challenged to think carefully about mathematics, especially in the context of applications (word problems). Instead, they were more comfortable with an instructor working an example at the board that they would then simply mimic in homework. In-depth, critical, mathematically-rich, and innovative thinking was challenging for most of them.

Two of the more capable students, though, were Sean and David. The comments and insights they provided throughout the course demonstrated mathematical intuition and a capacity for creative thought in mathematics. As such, their performance on the Post-Test was surprisingly low. For them, the development of the scoring rubric seemed not to help because they were distracted by the issue of grades. Oddly, David's assessment of mathematical arguments became increasingly accurate and sophisticated towards the end, and was quite in line with my own judgments of the arguments. I therefore believe that the rubric was not necessarily the primary cause for his difficulties. Instead, simply my language during the study, especially at the beginning, created conflicts in his thinking. Had David and others been able to work another week or two with the rubric, I suspect they could have moved past the issue of grades and focused better on the mathematical flaws in each argument.

For some student-participants, such as Jennifer, the rubric seemed to confuse them by adding another layer of complexity. The reasoning that Jennifer presented, both on the practice scoring exercises and on the Post-Test, indicated that she had only a passing familiarity with proofs. Her analysis of many arguments focused primarily on

superficial details, such as the wording or format of the argument. When her explanations attempted to go deeper or cite specific mathematical flaws, they were often confusing or off track. Her ability did not seem to improve as the study progressed, despite the whole-class activities and interactions, such as developing the rubric, discussing the critical aspects of a valid argument, and working cooperatively to implement the rubric.

In contrast to David, Sean, and Jennifer, some students demonstrated a significant improvement in their facility with mathematical arguments. Among these were Maria and Julie. In fact, Maria's improvement (Post-Test versus Pre-Test results) correlates strongly with our work developing and implementing a scoring rubric for mathematical arguments.

Maria seems to have had at least some initial knowledge of proofs. For example, she used the word "contrapositive" on an early assignment during Week 8. She perhaps had also been introduced to formal logic, as indicated by her use of the symbol " \sim " for the negation of a statement (Week 9). This probably explains why she was one of the first student-participants to begin correctly identifying the flaws in the arguments discussed during the study. For Argument M (all arguments used in class are found in Appendix F), collected during Week 10, half of the participants in some way indicated that a flaw

existed with the portion of the argument in which $n = 3 \left(\frac{k}{n} \right)$ implied n was a multiple

of 3, but none explicitly said that $\frac{k}{n}$ may not be an integer. This is basically the only flaw

with this argument. Maria was the closest to understanding this flaw, saying that the

argument "should have stated that $\frac{k}{n}$ is an integer." Her wording on this seems to

indicate that she understood the necessity of $\frac{k}{n}$ being an integer for the argument to be a valid proof, but that she perhaps did not realize $\frac{k}{n}$ may not be an integer in all cases.

During our first discussion of Week 8, Maria also indicated some previous knowledge and insight about rubrics. She knew they were “a list of rules used for grading” that reduce guesswork when grading and would require some experience so that the teacher learns what he or she expects to see in student work. However, as we developed a rubric in class for scoring mathematical arguments, Maria initially struggled with implementing it. For our work during Week 9, she believed that Argument E was better than Argument L, yet her first attempt at using the draft rubric at that point gave Argument L a better score. This discrepancy concerned Maria. As mentioned above, Maria was the only student to specifically identify the problem in Argument M; however, she deducted points for the argument’s flow or connection between steps rather than for its mathematical correctness. That is, she basically understood the math and followed the argument well, but still was struggling with the use of the rubric. For Argument J, scored in groups during Week 10, Maria was very harsh. The most noticeable error for most students was the double use of the letter n , which appears in different contexts on both sides of the equal sign. Maria realized the argument was incorrect, but did not find the few merits of the argument. For example, Argument J clearly attempts to address the general case, though poorly. However, Maria gave zero points for this aspect on the rubric. Maria was aware of her difficulty with using the rubric. Throughout the first few weeks, she repeatedly mentioned that she was unsure about using the rubric correctly and that our work exploring arguments had even caused her to doubt her own grasp of proofs.

Nevertheless, Maria's later work demonstrated an excellent facility with the rubric and with identifying flaws in mathematical arguments. For example, prior to the rubric, she felt that Argument L was very poorly done. Using the rubric, however, she realized that the argument's principal error was not addressing the general case, but she also realized that it indicated some knowledge about the correct direction for a valid "if-then" proof. Her specificity about this aspect of the argument hints that the rubric caused her to think more carefully about the components of a valid proof. By Week 10, the rubric was reaching its final form. Student-participants used the rubric to score Arguments G and K individually for homework. Maria did better with these. She correctly identified the main problems in the arguments and deducted points in more appropriate places on the rubric. For example, her scoring of Argument G almost exactly matched my own. She correctly and specifically found the only error in Argument K, but was a bit harsh in its scoring. All in all, though, she seemed to be using the rubric more effectively by this point, and had little problem correctly specifying the errors in the arguments. By Week 13, Maria had really caught on to the rubric. She was fairly close to my scoring for Argument P, which participants scored collaboratively in small groups, and agreed exactly with my scoring for Argument Q, which participants completed individually. More importantly, her reasoning for the scores she gave was completely on target and indicated a sophisticated understanding of the nuances in each argument.

Maria noted that she preferred scoring arguments as a group, even though she seemed to do better when she graded them individually. Interestingly, she realized that as a teacher she will most likely be scoring her students' arguments on her own and comments that outside input in the scoring process would probably not be helpful since

only she would know exactly how she had taught her class to do proofs. Insights like these clearly indicated that Maria obtained an intuitive general understanding of rubrics, as well as competence with the specific rubric we developed for scoring mathematical arguments.

Maria's improved skill with rubrics was mirrored by her improved facility with mathematical arguments. On the Pre-Test, she had only 2 correct multiple-choice responses and 3 correct open-ended responses. On the Post-test, however, *all* of her answers and explanations were correct, except the one about isosceles triangles (item #10 on the Post-Test), which one could argue was a correct proof since it did not really use the fact that the third side was also called p . Her Pre-Test clearly shows that she was trying to think analytically about each of her responses, but she simply missed the mark on many of the questions. Her answers on the Post-Test, though, are not only correct, but her explanations were succinct and completely on target. In the end, Maria improved dramatically. She commented that the rubric helped her to look for the important aspects of a proof, such as the overall structure, and therefore she felt that she gained a "better facility with proofs." She also seemed to have a good idea about the usefulness of rubrics as a teacher, saying that they are beneficial for both students and teachers. She wrote, "They allow students to know what is expected in their work," and "They give teachers a good structure for grading."

Another surprising yet interesting result indicates that many student-participants were gaining a better facility with investigating mathematical arguments. Item #19 on the Post-Test (involving the statement "If x is divisible by 5 and y is divisible by 3, then $x + y$ is divisible by 8.") contained a typographical error. At least 7 participants

recognized this fact and stated that the problem contained a misprint, rather than citing the typographical error as the flaw in the argument. Thus, they were still focused on the formatting details of an argument, but their sophistication with analyzing arguments had improved to the point of recognizing an unintended flaw. This was foreshadowed by their scoring of Argument J, in which nearly every participant identified the redundant use of the variable.

Confounding Factors

Three primary factors affected the study negatively. The most important of these was a “false start.” Before the semester began, the original focus of the study was to be the development of scoring rubrics as a means to improve participants’ facility with mathematical arguments. Due to concerns about the complexity of this focus, I was advised to concentrate on the idea of examples and counterexamples in understanding mathematical arguments. However, most students did well with the items involving examples and counterexamples on the Pre-Test, which left little room for showing any improvement in this area. The test could have been changed to illuminate participants’ range of facility with examples and counterexamples better, but the timeline for the study did not allow for the design and proctoring of a new Pre-Test. Moreover, the creativity involved for participants to develop their own examples and counterexamples would have been difficult to measure and to teach. Fortunately, the Pre-Test did still include useful information about participants’ ability to identify correct mathematical reasoning. Therefore, the study then returned to its original focus of developing and implementing scoring rubrics as a means of improving student-participants’ facility with validating proofs. Appropriate IRB approval was obtained for each new manifestation of the study.

As a consequence of the false start, the Pre-Test contained questions that ultimately were irrelevant to the study about rubrics. Moreover, the rest of the study became rushed; activities were completed back to back. Ideally, the activities for this study should have been spaced throughout the semester better. A result of the rushed schedule was that I could not postpone the Post-Test, even though my years of teaching experience gave me a very definite sense that the weather was affecting the mood and involvement of the participants that day. Also, having the Post-Test close to the end of the semester may have interfered with participants' ability to concentrate since they potentially had finals and projects in other classes to consider.

A second factor confounding the study was that the Pre-Test and Post-Test were not piloted before being used in the study. Therefore, issues arose regarding the wording and clear presentation of some items. For example, the argument presented in item #8 of the Post-Test (involving the statement "For any integer n , if n^2 is odd, then n is odd.") was valid; however, it did not explicitly state that the contrapositive had been proved rather than the original statement. The reasoning provided by many participants on the associated open-ended item indicated that they could follow the logical flow and algebraic details of the argument, but that they did not quite see the contrapositive construction. In fact, a couple of other participants explicitly stated that the argument was not valid specifically because it did not say "by contrapositive." That is, they seemed to understand the argument completely, but the wording of the item interfered with their ability to give a correct multiple-choice response. (Note: The decision not to explicitly state "by contrapositive" was made with the belief that doing so would cause participants to automatically determine the argument to be valid simply because it included a "fancy"

mathematical term they recognized but perhaps did not completely understand. Adding a second argument that also involved the words “by contrapositive” but was not valid perhaps could have helped to distinguish whether participants understood the construction of the arguments or focused merely on the mathematical terminology.)

The typographical error mentioned earlier on item #19 of the Post-Test could have also been avoided if the test had been piloted first. Surprisingly, many of the participants pointed out the typographical error in their explanations for the open-ended portion of the question; that is, they realized that that particular error was typographical and not the intended error in the validity of the argument. Finally, the error in Item #10 (involving the statement “If triangle ABC is an isosceles triangle with congruent sides \overline{AB} and \overline{AC} , then the base angles ($\angle ABC$ and $\angle ACB$) are congruent.”) of the Post-Test was unclear. The argument used the same letter to represent two possibly different lengths of a triangle. The intent of the item was for participants to see that the double use of the variable meant the argument only proved the given statement for one case, that of an equilateral triangle. However, the construction of the rest of the argument did not rely on the incorrect double use of the variable. Many participants therefore stated that the argument represented a valid proof. Six participants identified the incorrect use of the variable, and one specifically stated that the double use of the variable did not actually affect the validity of rest of the argument. Piloting the test first could have led to a clearer presentation of this item.

The third factor having a noticeable negative impact on the study was the confusion created by not being clear from the beginning with the words “grade” and “score.” This was partially due to an incomplete discussion regarding rubrics at the

beginning of the study. Student-participants had differing conceptions about rubrics. Moreover, a lack of experience about how to assign grades caused the participants to make assumptions about the connection between our work with scoring rubrics and its affect on a student's grade.

Further Work

Clearly, the first change in any future study would be to keep the focus on rubrics. The items on the Pre-Test that were geared towards the focus on mathematical examples and counterexamples would be eliminated. Also, a better introduction to rubrics would facilitate classroom discussions regarding their use in scoring proof attempts. In particular, the teacher-researcher would need to present several different types of rubrics, giving examples of each, with the intention of directing student-participants towards the specific style of rubric to be used later with mathematical arguments. The researcher should basically plan to have one full discussion regarding rubrics before ever launching into the idea of using them with mathematical arguments. Additionally, with a clear focus from the beginning, the researcher could start the study earlier in the course, providing a more flexible schedule and allowing the Post-Test to be administered before the last push at the end of the semester.

In addition to correcting and improving the Pre-Test and Post-Test based on the previous discussion of their flaws, other changes could provide more reliable instruments for gauging participant progress. First, the items that were added to the Post-Test were good and should be incorporated into the Pre-Test. Doing so would provide a better basis for comparing results before and after the implementation of the instructional technique. Second, the structure of the tests could be more consistent. In particular, each argument

presented should be followed by both a multiple choice item and an open-ended item.

Third, any changes to the tests should first be piloted with a different set of student-participants, not involved with the study regarding rubrics, in order to check the clarity of the questions.

Another change in the study that I recommend is to be more careful with the use of the words “grading” and “scoring.” I intended for us, the researcher and the participants, to develop a method for scoring mathematical arguments – to delineate the various components of a mathematical argument and rate students on a scale regarding their performance with these components. How teachers convert that score into the grade assigned to a student is a different matter. Being clear and up front about the distinction, as well as discussing ways teachers can convert the score into a grade, would hopefully eliminate the concern some participants may have about the use of the scoring rubrics, thereby allowing the participants to focus better on the content of the rubric.

Finally, recent research articles have exposed the variety of beliefs (many incorrect) that students have about proofs and how this affects their judgment of proofs (Harel & Sowder, 1998; Moore, 1994; Selden, A. & Selden, J., 2003). What do students believe is a proof? Although this study relates strongly to this question, the study did not specifically try to ascertain students’ proof conceptions or beliefs. Instead, the study focused simply on students’ facility with mathematical arguments. That is, the focus was on “what can students do?” not “why do they do it?” Although the discussions regarding the development of the rubric certainly were the result of student beliefs about proofs, trying to investigate their proof conceptions as a separate component of the study would have added significant complexity. Perhaps any further work with the instructional

technique used in this study could try to interconnect student beliefs about proofs, as well as their ability to validate mathematical arguments, with the development of a scoring rubric for those arguments.

Conclusion

The outstanding progress shown by Maria, as well as other participants such as Julie, indicates that the attempted instructional technique (using scoring rubrics to improve teacher understanding of proofs) may still have merit. Moreover, in regards to mathematics teacher education, the attempted technique addresses two important topics at once – scoring rubrics and proofs. I am convinced that correcting and improving the clarity of the two tests, addressing the potential confusion between grades and scores, and presenting a more thorough introduction to scoring rubrics at the beginning of another study has the potential of producing more notable results.

Appendices

Appendix A: Pre-Test	70
Appendix B: Post-Test	77
Appendix C: Consent to Participate in Research Form	84
Appendix D: Rubric Drafts	87
Appendix E: Example Rubrics	89
Appendix F: Arguments Presented to Participants During the Study	91
Appendix G: Writing Prompts	96

Appendix A

Pre-Test

Name _____

Pre-Test, October 6, 2003

(This will be replaced with your pseudonym, unless you are not participating in the study, in which case I will not make a copy of this test. Do NOT put your name on any of the other pages.)

Pre-Test

Answer as many of the following questions as completely as you can. Be aware that some of the terms defined below may relate to “artificial” objects and are not actual mathematical definitions. The point is to analyze the given definition and answer the questions based on that definition.

An “argument” is a set of statements that attempt to prove a given statement. Assume that “valid proof” means a correct formal mathematical proof, based on your understanding of proofs.

When a set of choices is given for a problem, put the letter corresponding to your choice in the appropriate box. Unless otherwise stated, choose only the one best response for each problem.

The test items are in no particular order, so you do not have to answer them in order.

Recall the following facts:

Integers are the set of numbers $\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$.

A number n is divisible by 2 if $n = 2a$ for some integer a .

(Ex: 10 is divisible by 2 since $10 = 2 \cdot 5$)

A number m is divisible by 3 if $m = 3b$ for some integer b .

(Ex: 21 is divisible by 3 since $21 = 3 \cdot 7$)

A number is even if it is divisible by 2. A number is odd if it is an integer but is not even.

Note: Some of these are technical definitions, but they are basically the same as your previous knowledge about divisibility and even/odd.

1) Consider the following mathematical argument:

$0 = 2 \cdot 0$, and 0 is an integer.

Thus, 0 is divisible by 2.

Hence, 0 is even.

1)

Does this argument correctly demonstrate that 0 is an even number?

- A) Yes, this argument correctly demonstrates that 0 is an even number.
- B) Zero is neither even nor odd, so this argument cannot be correct.
- C) Although 0 is an even number, this argument does not correctly demonstrate that 0 is even because it only shows one example.
- D) This argument contains arithmetic errors and thus does not correctly demonstrate that 0 is even.

Name _____

Pre-Test, October 6, 2003

2) A *hextant* number is defined as a number that is divisible by both 2 and 3.

i) Based on this definition, give an example of a *hextant* number, or state that no such example exists.

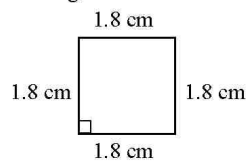
ii) Based on this definition, give an example of a number that is divisible by 2 but is NOT a *hextant* number, or state that no such example exists.

3) Which one of the following represents the correct logical structure of a valid proof for the statement “All squares are rectangles”? That is, which of the following is a correct outline for the proof that “All squares are rectangles”?

3)

A) Let R be any rectangle. Show that the rectangle R is not a square.

B) Consider the square drawn below with side lengths of 1.8 centimeters. Show that this square is also a rectangle.



C) Let S be any square. Show that the square S satisfies the definition of a rectangle.

D) Let R be any rectangle. Show that the rectangle R satisfies the definition of a square.

4) Briefly explain the reason(s) for your choice in #3 above.

Name _____

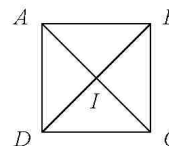
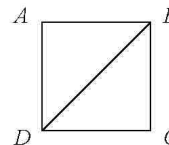
Pre-Test, October 6, 2003

- 5) Below is an argument for the statement:
“If the diagonals of a rectangle are perpendicular, then it must be a square.”

Let ABCD be a square. Then the diagonal BD forms a 45-degree angle with the sides since it is the hypotenuse of an isosceles right triangle. Similarly the diagonal AC forms a 45-degree angle with the sides.

Let I be the point of intersection of the diagonals. Then triangle ABI is a right triangle since two of its angles are 45 degrees.

Thus $AC \perp BD$, that is, the diagonals of the square are perpendicular to each other, as desired.



Which of the following best describes this argument?

- A) The argument represents a valid mathematical proof for the statement.
B) The argument does not represent a valid mathematical proof for the statement because it does not prove the statement for all cases.
C) There can be no valid proof because the statement is false.
D) The argument may be a valid proof, but does not prove the statement given.
- 6) Briefly explain the reason(s) for your choice in #5 above.
- 7) Consider the statement “If x and y are odd integers, then $x \cdot y$ is an odd integer.”
- i) Does the following example prove the statement is false? Explain briefly.
If $x = 4$ and $y = 5$, then $x \cdot y = 20$.
- ii) Does the following example prove the statement is true? Explain briefly.
If $x = 3$ and $y = 7$, then $x \cdot y = 21$.

5)

Name _____

Pre-Test, October 6, 2003

- 8) Below is an argument for the statement “The sum of two even integers is even.”

8)

Let n be an even integer.

Then $n = 2k$ for some integer k .

Hence, $n + n = 2k + 2k = 2(k + k) = 2(2k) = 2j$ for some integer j .

Thus, $n + n$ is even.

Therefore, the sum of two even integers is even.

Which of the following best describes this argument?

- A) The argument represents a valid mathematical proof for the statement.
- B) The argument does not represent a valid mathematical proof for the statement because it does not prove the statement for all cases.
- C) There can be no valid proof because the statement is false.
- D) The argument does not represent a valid mathematical proof for the statement because it contains algebraic and/or arithmetic errors.
- 9) Briefly explain the reason(s) for your choice in #8 above.

- 10) Consider the statement “If $x + y$ is odd and $y + z$ is odd, then $x + z$ is odd.”
Give a counterexample for this statement.

Name _____

Pre-Test, October 6, 2003

- 11) Consider the statement "If $x - 2 < 0$ and $x^2 - 4 = 0$, then $|x| \geq 2$." Is this statement true or false?
If it is false, provide a counterexample. If it is true, explain briefly.

- 12) Let n be a positive integer. Below is an argument for the statement:
"If $n(n+2)$ is divisible by 2, then n is divisible by 2."

12)

Let n be divisible by 2.

Then $n = 2x$ for some integer x .

Hence, $n(n+2) = 2x(2x+2) = 4x^2 + 4x = 2(2x^2 + 2x) = 2y$, where $y = 2x^2 + 2x$.

That is, $n(n+2) = 2y$ for some integer y .

Thus, $n(n+2)$ is divisible by 2.

Therefore, if $n(n+2)$ is divisible by 2, then n is divisible by 2.

Which of the following best describes this argument?

- A) The argument represents a valid mathematical proof for the statement.
B) The argument does not represent a valid mathematical proof for the statement because it does not prove the statement for all cases.
C) The argument does not represent a valid mathematical proof for the statement because it contains algebraic and/or arithmetic errors.
D) The argument may be a valid proof, but does not prove the statement given.
- 13) Briefly explain the reason(s) for your choice in #12 above.

Name _____

Pre-Test, October 6, 2003

14) Note: A *rational* number is any number that can be written as a ratio of integers, that is, as $\frac{p}{q}$ (with $q \neq 0$), where p and q are integers.

14)

An *irrational* number is any real number that is not rational.

Below is an argument for the statement:

“There exists an irrational number x such that $x^{\sqrt{2}}$ is rational.”

Consider $(\sqrt{3})^{\sqrt{2}}$. Either $(\sqrt{3})^{\sqrt{2}}$ is rational or $(\sqrt{3})^{\sqrt{2}}$ is irrational.

If $(\sqrt{3})^{\sqrt{2}}$ is rational, let $x = \sqrt{3}$. Then x is irrational but $x^{\sqrt{2}} = (\sqrt{3})^{\sqrt{2}}$ is rational, and the desired x is found.

If $(\sqrt{3})^{\sqrt{2}}$ is irrational, let $x = (\sqrt{3})^{\sqrt{2}}$. Then x is irrational but $x^{\sqrt{2}} = \left((\sqrt{3})^{\sqrt{2}}\right)^{\sqrt{2}} = (\sqrt{3})^2 = 3$ is rational, and the desired x is found.

Thus, either $x = \sqrt{3}$ or $x = (\sqrt{3})^{\sqrt{2}}$ is an irrational number x such that $x^{\sqrt{2}}$ is rational.

Which of the following best describes this argument?

- A) The argument represents a valid mathematical proof for the statement.
- B) The argument does not represent a valid mathematical proof for the statement because it does not prove the statement for all cases.
- C) The argument does not represent a valid mathematical proof for the statement because it contains algebraic and/or arithmetic errors.
- D) The argument may be a valid proof, but does not prove the statement given.

15) Briefly explain the reason(s) for your choice in #14 above.

16) We can think of the SAS (side-angle-side) theorem in geometry as saying that Triangle E is congruent to Triangle F if the following two conditions are true:

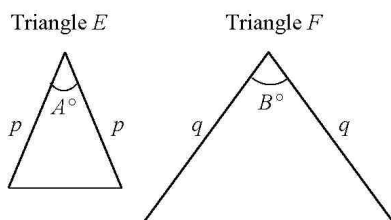
- I. Two sides of Triangle E are congruent to two (corresponding) sides of Triangle F .
- II. The included angle (angle between the two given sides) of Triangle E is congruent to the (corresponding) included angle of Triangle F .

Which one of the following diagrams provides an example to specifically show that condition (I.) alone is not enough to prove Triangle E is congruent to Triangle F ?

16)

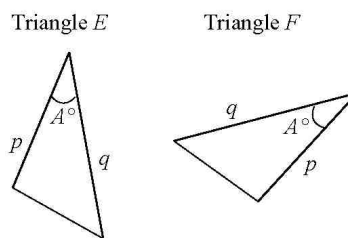
(Note that p and q represent different side lengths, whereas A and B represent different angle measures.)

A)



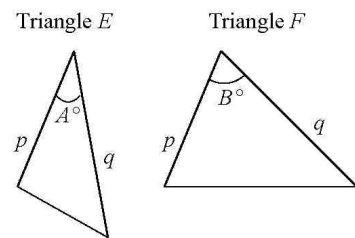
Triangle E is not congruent to Triangle F

B)



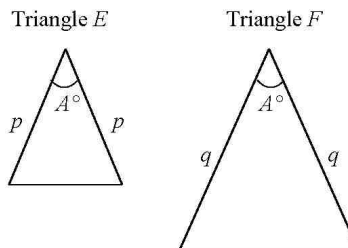
Triangle E is congruent to Triangle F

C)



Triangle E is not congruent to Triangle F

D)



Triangle E is not congruent to Triangle F

17) Briefly describe what you were looking for in #16 above. That is, how did you decide which choice correctly answered the question?

Appendix B

Post-Test

Name _____

Post-Test, December 8, 2003

(This will be replaced with your pseudonym, unless you are not participating in the study, in which case I will not make a copy of this test. Do NOT put your name on any of the other pages.)

Post-Test

Answer as many of the following questions as completely as you can.

An “argument” is a set of statements that attempt to prove a given statement. Assume that “valid proof” means a correct formal mathematical proof, based on your understanding of proofs.

When a set of choices is given for a problem, put the letter corresponding to your choice in the appropriate box. Unless otherwise stated, choose only the one best response for each problem.

The test items are in no particular order, so you do not have to answer them in order.

Recall the following facts:

Integers are the set of numbers $\{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$.

A number n is divisible by 2 if $n = 2a$ for some integer a .

(Ex: 10 is divisible by 2 since $10 = 2 \cdot 5$)

A number m is divisible by 3 if $m = 3b$ for some integer b .

(Ex: 21 is divisible by 3 since $21 = 3 \cdot 7$)

A number is even if it is divisible by 2. A number is odd if it is an integer but is not even.

Note: Some of these are technical definitions, but they are basically the same as your previous knowledge about divisibility and even/odd.

1) Consider the following mathematical argument:

$0 = 2 \cdot 0$, and 0 is an integer.

Thus, 0 is divisible by 2.

Hence, 0 is even.

1)

Does this argument correctly demonstrate that 0 is an even number?

- A) Yes, this argument correctly demonstrates that 0 is an even number.
- B) Zero is neither even nor odd, so this argument cannot be correct.
- C) Although 0 is an even number, this argument does not correctly demonstrate that 0 is even because it only shows one example.
- D) This argument contains arithmetic errors and thus does not correctly demonstrate that 0 is even.

Name _____

Post-Test, December 8, 2003

- 2) Below is an argument for the statement "If $Ax + By = C$ determines y as a function of x , then $B \neq 0$."

If $Ax + By = C$, then $By = C - Ax$.

If $B \neq 0$, then $y = \frac{C}{B} - \frac{A}{B}x$.

Thus, y is a function of x for $B \neq 0$.

2)

Which of the following best describes this argument?

- A) The argument represents a valid mathematical proof for the statement.
- B) The argument does not represent a valid mathematical proof for the statement because it does not prove the statement for all cases.
- C) The argument does not represent a valid mathematical proof for the statement because it contains algebraic and/or arithmetic errors.
- D) The argument may be a valid proof, but does not prove the statement given.
- 3) Briefly explain the reason(s) for your choice in #2 above.

- 4) Below is an argument for the statement "The sum of two even integers is even."

Let n be an even integer.

Then $n = 2k$ for some integer k .

Hence, $n + n = 2k + 2k = 2(k + k) = 2(2k) = 2j$ for some integer j .

Thus, $n + n$ is even.

Therefore, the sum of two even integers is even.

4)

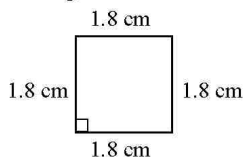
Which of the following best describes this argument?

- A) The argument represents a valid mathematical proof for the statement.
- B) The argument does not represent a valid mathematical proof for the statement because it does not prove the statement for all cases.
- C) There can be no valid proof because the statement is false.
- D) The argument does not represent a valid mathematical proof for the statement because it contains algebraic and/or arithmetic errors.
- 5) Briefly explain the reason(s) for your choice in #4 above.

- 6) Which one of the following represents the correct logical structure of a valid proof for the statement “All squares are rectangles”? That is, which of the following is a correct outline for the proof that “All squares are rectangles”?

6)

- A) Let R be any rectangle. Show that the rectangle R is not a square.
 B) Consider the square drawn below with side lengths of 1.8 centimeters. Show that this square is also a rectangle.



- C) Let S be any square. Show that the square S satisfies the definition of a rectangle.
 D) Let R be any rectangle. Show that the rectangle R satisfies the definition of a square.
- 7) Briefly explain the reason(s) for your choice in #6 above.

- 8) Below is an argument for the statement: “For any integer n , if n^2 is odd, then n is odd.”

Suppose n is not odd. Then n is even and $n = 2k$ for some integer k .

Hence, $n^2 = (2k)^2 = 4k^2 = 2(2k^2) = 2j$,

where j is an integer since $2k^2$ is an integer and $j = 2k^2$.

Since $n^2 = 2j$, then n^2 is even. That is, n^2 is not odd.

Therefore, if n^2 is odd, then n is odd.

8)

Which of the following best describes this argument?

- A) The argument represents a valid mathematical proof for the statement.
 B) The argument does not represent a valid mathematical proof for the statement because it does not prove the statement for all cases.
 C) There can be no valid proof because the statement is false.
 D) The argument may be a valid proof, but does not prove the statement given.
- 9) Briefly explain the reason(s) for your choice in #8 above.

10) Below is an argument for the statement:

“If triangle ABC is an isosceles triangle with congruent sides \overline{AB} and \overline{AC} , then the base angles ($\angle ABC$ and $\angle ACB$) are congruent.”

Given triangle ABC is an isosceles triangle, let p be the length of the congruent sides.

Let M be the midpoint of the segment \overline{BC} .

If the length of side \overline{BC} is p , then the length of segment \overline{BM} is $\frac{1}{2}p$ and the length of segment \overline{CM} is $\frac{1}{2}p$.

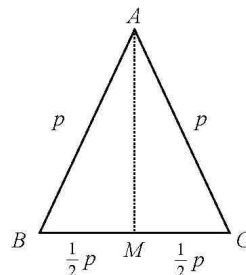
Therefore, segment \overline{BM} is congruent to segment \overline{CM} .

By the reflexive property, the segment \overline{AM} that joins point A to point M is congruent to itself.

So, \overline{AM} is congruent to \overline{AM} , \overline{AB} is congruent to \overline{AC} , and \overline{BM} is congruent to \overline{CM} .

Thus, triangle ABM is congruent to triangle ACM by SSS (side-side-side).

Since corresponding parts of congruent triangles are congruent, then $\angle ABC$ and $\angle ACB$ are congruent.



Which of the following best describes this argument?

10)

- A) The argument represents a valid mathematical proof for the statement.
- B) The argument does not represent a valid mathematical proof for the statement because it does not prove the statement for all cases.
- C) The argument does not represent a valid mathematical proof for the statement because it contains algebraic and/or arithmetic errors.
- D) There can be no valid proof because the statement is false.

11) Briefly explain the reason(s) for your choice in #10 above.

Name _____

Post-Test, December 8, 2003

12) Note: A *rational* number is any number that can be written as a ratio of integers, that is, as $\frac{p}{q}$ (with $q \neq 0$), where p and q are integers.

12)

An *irrational* number is any real number that is not rational.

Below is an argument for the statement:

“There exists an irrational number x such that $x^{\sqrt{2}}$ is rational.”

Consider $(\sqrt{3})^{\sqrt{2}}$. Either $(\sqrt{3})^{\sqrt{2}}$ is rational or $(\sqrt{3})^{\sqrt{2}}$ is irrational.

If $(\sqrt{3})^{\sqrt{2}}$ is rational, let $x = \sqrt{3}$. Then x is irrational but $x^{\sqrt{2}} = (\sqrt{3})^{\sqrt{2}}$ is rational, and the desired x is found.

If $(\sqrt{3})^{\sqrt{2}}$ is irrational, let $x = (\sqrt{3})^{\sqrt{2}}$. Then x is irrational but $x^{\sqrt{2}} = \left((\sqrt{3})^{\sqrt{2}}\right)^{\sqrt{2}} = (\sqrt{3})^2 = 3$ is rational, and the desired x is found.

Thus, either $x = \sqrt{3}$ or $x = (\sqrt{3})^{\sqrt{2}}$ is an irrational number x such that $x^{\sqrt{2}}$ is rational.

Which of the following best describes this argument?

- A) The argument represents a valid mathematical proof for the statement.
- B) The argument does not represent a valid mathematical proof for the statement because it does not prove the statement for all cases.
- C) The argument does not represent a valid mathematical proof for the statement because it contains algebraic and/or arithmetic errors.
- D) The argument may be a valid proof, but does not prove the statement given.

13) Briefly explain the reason(s) for your choice in #12 above.

Name _____

Post-Test, December 8, 2003

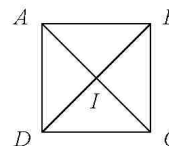
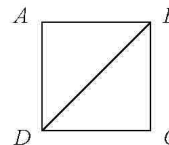
14) Below is an argument for the statement:

“If the diagonals of a rectangle are perpendicular, then it must be a square.”

Let ABCD be a square. Then the diagonal BD forms a 45-degree angle with the sides since it is the hypotenuse of an isosceles right triangle. Similarly the diagonal AC forms a 45-degree angle with the sides.

Let I be the point of intersection of the diagonals. Then triangle ABI is a right triangle since two of its angles are 45 degrees.

Thus $AC \perp BD$, that is, the diagonals of the square are perpendicular to each other, as desired.



Which of the following best describes this argument?

- A) The argument represents a valid mathematical proof for the statement.
- B) The argument does not represent a valid mathematical proof for the statement because it does not prove the statement for all cases.
- C) There can be no valid proof because the statement is false.
- D) The argument may be a valid proof, but does not prove the statement given.

14)

15) Briefly explain the reason(s) for your choice in #14 above.

16) Consider the statement “If $x - 2 < 0$ and $x^2 - 4 = 0$, then $|x| \geq 2$.” Is this statement true or false? If it is false, provide a counterexample. If it is true, explain briefly.

Name _____

Post-Test, December 8, 2003

17) Let n be a positive integer. Below is an argument for the statement:

“If $n(n+2)$ is divisible by 2, then n is divisible by 2.”

17)

Let n be divisible by 2.

Then $n = 2x$ for some integer x .

Hence, $n(n+2) = 2x(2x+2) = 4x^2 + 4x = 2(2x^2 + 2x) = 2y$, where $y = 2x^2 + 2x$.

That is, $n(n+2) = 2y$ for some integer y .

Thus, $n(n+2)$ is divisible by 2.

Therefore, if $n(n+2)$ is divisible by 2, then n is divisible by 2.

Which of the following best describes this argument?

- A) The argument represents a valid mathematical proof for the statement.
- B) The argument does not represent a valid mathematical proof for the statement because it does not prove the statement for all cases.
- C) The argument does not represent a valid mathematical proof for the statement because it contains algebraic and/or arithmetic errors.
- D) The argument may be a valid proof, but does not prove the statement given.

18) Briefly explain the reason(s) for your choice in #17 above.

19) A student makes the following conjecture:

“If x is divisible by 5 and y is divisible by 3, then $x + y$ is divisible by 8.”

Below is the student’s argument for this conjecture:

If x is divisible by 5, then $x = 5k$ for some integer k .

If y is divisible by 3, then $y = 3k$ for some integer k .

Thus, $x + y = 5k + 3y = (5 + 3)k = 8k$.

That is, $x + y$ is divisible by 8.

Is this a valid proof or not? Explain.

Appendix C

Consent to Participate in Research Form

CONSENT TO PARTICIPATE IN RESEARCH

- **INTRODUCTION**

You are invited to participate in a research study conducted by myself, Timothy J. Middleton, a Graduate Student in the Department of Mathematics and Statistics at the University of New Mexico. The results of the study will be used in my master's thesis. The study is entitled "Development of Scoring Rubrics as a Method for Improving Pre-Service Teachers' Ability to Validate Mathematical Proofs."

You were selected as a possible participant in the study because you are enrolled in Math 338, Section 001, which is a mathematics course designed for pre-service teachers.

- **PURPOSE OF THE STUDY**

Several research studies conducted over the past three decades have demonstrated the various problems encountered by students and teachers regarding formal mathematical proofs. However, few researchers have tested new instructional strategies in this area. The basic aim of the research is to determine if a specific instructional strategy, that of developing scoring rubrics within a cooperative classroom setting, improves pre-service teachers' facility with proofs. In particular, the study will ascertain if the design, utilization, and refinement of rubrics for grading student proofs, as well as the associated classroom discussions, improve pre-service teachers' ability to determine the correctness of proofs, to specify the logical structure required for a proof, and to point out the specific errors in an invalid proof.

- **PROCEDURES**

The principal investigator for this study is myself, Timothy J. Middleton. I intend to employ action research methods in order to demonstrate a link between the development of scoring rubrics in class and improvements in students' ability to validate proofs. Here, action research refers to the practice in education of studying an instructional strategy by actually implementing it in the classroom and documenting the results. Hence, I will collect data from a variety of sources that arise as normal instructional activities in the class: responses to a pre-test, reflective writing assignments, student-generated proofs, instructor's class observation notes, responses to a post-test, and the various versions of scoring rubrics developed cooperatively by the class.

The pre-test and post-test will consist of 15 to 30 questions regarding the accuracy of proofs, the logical structure required for a proof, and the specific errors in an invalid proof. The pre-test will be given the 7th week during class and should take 30 to 45 minutes to complete. The post-test will be given approximately the 15th week during class and should take 45 to 60 minutes to complete.

Approximately the 10th week of class, the class will cooperatively develop and utilize an initial scoring rubric for sample proofs that I provide. Further development and refinement of scoring rubrics will occur approximately the 12th and 14th weeks in the course, using proofs that the students write (related to the current area of mathematics studied in the course) and possibly other proofs that I provide. I will collect the student-generated proofs and any written work associated with developing and implementing the scoring rubrics. Also, at least three reflective writing assignments will be given regarding proofs and scoring rubrics. These are one or two paragraph responses to an instructor-selected prompt, given in class or as homework, and will be assigned approximately the 10th, 12th, and 14th weeks of class.

Please note that ALL students in this course will be required to complete the pre-test, reflective writing assignments, proofs, rubric development and implementation assignments, and post-test as normal instructional activities of the class. These assignments will be collected and graded as part of the course expectations without regard to your participation in the study. However, your work will ONLY be used in the research study if you allow me to do so by signing and returning this consent form. In other words, participating in the study neither increases nor decreases the amount of work you must do in this course, nor does it improve your grade in any way. It simply gives me permission to use your work in my research study. Procedures for protecting the privacy and confidentiality of all students are described below.

- **POTENTIAL RISKS AND DISCOMFORTS**

Participation in this study will involve only the normal minimal risks associated with taking an upper-level undergraduate mathematics class. You should feel comfortable with the procedures described below for protecting your identity before signing this form.

- **POTENTIAL BENEFITS TO PARTICIPANTS AND/OR SOCIETY**

Results of this study will add to the knowledge base of instructional procedures in mathematics education. Specifically, the research study will give mathematics educators insight for improving instruction related to formal mathematical proofs, an area of considerable concern in recent years. The study may also have a direct benefit for you. Students will experience the process for developing and implementing a scoring rubric, will experience how cooperative groups can be utilized in the classroom, and will practice analyzing and scoring proofs. Through these important activities, you will hopefully become better prepared to teach secondary mathematics. Of course, you may experience this benefit regardless of your participation in the study.

All data collection methods described above will occur as part of the normal instructional activities of the course and therefore could constitute a portion of all students' grades. For example, all students in the class will complete the previously described reflective writing assignments and receive a grade. However, data for the study will ONLY be drawn from the work of those students who voluntarily choose to participate and have signed the "Consent to Participate in Research" form. Although everyone in the class is invited to participate in the study, declining to do so will not affect your grade in any way. Grades will be assigned to all student work using the standard procedures of the course and without regard to your participation in the study. Students choosing to participate will receive no payment, no extra points towards their grade, no special allowances for deletion or addition of assignments towards their grade, and no other compensation for participating except the personal satisfaction of helping to advance knowledge in the discipline of mathematics education and to advance your own growth as a mathematics teacher.

- **CONFIDENTIALITY**

The following procedures will be employed to protect the identity of all students in the Math 338 course.

No personal or identifying information will be collected from participants except their names. Any and all references in the study to participating students will be made (or replaced) using their corresponding pseudonyms as described below. Any and all references to non-participants will be made (or replaced) with phrases such as "a classmate" or "another student." These procedures apply to student work and my own classroom observation notes.

Once I know which students will be participating in the study and have received their signed "Consent to Participate in Research" informed consent form, I will generate a hand-written list associating each participating student with a pseudonym. The pseudonyms will be random generic student names, none of which are equivalent to the names of any students in class. The original of this pseudonym list will be kept by my advisor, Dr. Umland, in Humanities 456. A copy of the list will be kept either with me while working privately on the study or in my office in Humanities 456. Besides myself, only the members of my thesis committee (Dr. Kristin Umland, Dr. James Ellison, and Dr. Richard Kitchen) will possibly be given access to this list or to any data collected from participants. Both the original list and the copy will be destroyed by March 31, 2004 unless I obtain an extension from the Institutional Review Board prior to this time due to extenuating circumstances.

Only photocopies of student work will be used for the study and only from those students choosing to participate. I will use removable (Post-It®) tape to cover over names PRIOR to photocopying. Thus, the sole connection between the data collected for my study and your name will be the temporary list (1 original and 1 copy) generated at the beginning of the study correlating participants with a pseudonym. Your original work will be returned according to the normal instructional practices of the course: promptly after being reviewed and/or graded, and no later than the end of the semester. I will keep the photocopies indefinitely.

- **PARTICIPATION AND WITHDRAWAL**

Participation in this research study is completely voluntary. You will receive no compensation for your participation. You can choose whether to participate in this study or not, and refusal to participate will involve no penalty or loss of benefits to which you might otherwise be entitled. If you volunteer to participate, you may discontinue participation at any time without penalty or loss of benefits to which you might otherwise be entitled. To withdraw from the study at any time for any reason, simply provide me with a signed note stating your desire to withdraw. (In consideration of the deadlines for my thesis, if you decide to withdraw, please inform me prior to January 23, 2004.)

I maintain the right to terminate your participation in the study without your consent if you drop the course before final grades are submitted. However, maintaining this right does not prevent me from deciding instead to use any data already collected from you for my study, *unless* you withdraw from the study as described above. In other words, it is your responsibility to withdraw from the study if you drop the course. Otherwise, I may or may not include your work in my study.

- **IDENTIFICATION OF INVESTIGATORS AND REVIEW BOARD**

If you have any questions or concerns about the research, please feel free to contact:

Principal Investigator:	Timothy J. Middleton (505) 453-2621	Humanities, Room 456 Department of Mathematics & Statistics University of New Mexico Albuquerque, NM 87131
Responsible Faculty Member:	Dr. Kristin Umland 277-2514	Humanities, Room 456 Department of Mathematics & Statistics University of New Mexico Albuquerque, NM 87131
Institutional Research Board:	Professor José Rivera 277-2257	Scholes Hall, Room 255 University of New Mexico Albuquerque, NM 87131

SIGNATURE OF RESEARCH PARTICIPANT

I understand the procedures described above. My questions have been answered to my satisfaction, and I agree to participate in this study. I have been provided a copy of this form.

Name of Participant (please print)

Signature of Participant

Date

SIGNATURE OF INVESTIGATOR

In my judgment the participant is voluntarily and knowingly providing informed consent and possesses the legal capacity to provide informed consent to participate in this research study.

Signature of Investigator

Date

Appendix D

Rubric Drafts

Below are the two major drafts of the scoring rubric as it was developed collaboratively in class. Rubric Draft 1 was written by hand at the board and is simply retyped here.

Rubric Draft 2 appears basically in the same form that was printed for participants.

Rubric Draft 1

1) Overall logical structure	0	3	6
2) Valid steps			
a) Algebra correct	0	2	4
b) Flow/connections	0	1	2
3) Proves what was to be proven			
a) Demonstrates grasp of concepts	0	2	4
b) Attempts to prove general case	0	2	4

Rubric Draft 2

<u>Scoring Rubric for Proofs</u>				
1. <u>Process</u>				<u>Comments</u>
A. Argument clearly addresses the general case (if appropriate)	0	2	4	
B. Overall logical structure of the argument is appropriate for the assertion	0	3	6	
C. Flow of the argument and connections among steps are clear and appropriate	0	1	2	
2. <u>Mechanics</u>				
A. Individual steps and statements in the argument are correct	0	2	4	
B. Argument demonstrates mastery of all pertinent concepts	0	2	4	

Appendix E

Example Rubrics

Grading Rubric for Daily Assignments

The grading scheme below will be used to assess the quality and accuracy of some daily assignments. The purpose of this system is to center student attention on accuracy and thoroughness, rather than merely on point values. Although the following rubric is intended to serve as a guide for students, the descriptions are NOT entirely comprehensive nor binding in every case.

A

- The problems or proofs are complete and all answers are correct.
- Any writing or algebra errors are *truly* minor and have little bearing on the problem's accuracy.
- "Word problems" have "word answers" with appropriate units of measure. Any graphs are drawn neatly on a set of clearly labeled axes with at least two exact points specified directly on the graph.
- All the steps displayed in the solutions are appropriate. The steps demonstrate a clear understanding of each problem and the reasoning used to solve it, while exemplifying a "mastery" of the necessary math skills. Also, the assignment clearly indicates that effort and time were expended to complete the problems.
- The solution uses the correct methods, as specified in the directions, and the correct mathematical notation for each problem.

B

- The assignment is complete and clearly displays the reasoning used to solve each problem.
- The methods used are mostly correct and generally arrive at the correct answers.
- NO more than two noticeable flaws exist throughout the entire assignment or problem set, such as failure to provide a "word problem" with a "word answer," using a decimal approximation inappropriately, showing incorrect algebraic techniques, not including an essential step in a proof, not labeling axes or points on a graph, or being illegible.
- The assignment clearly indicates that effort and time were expended to complete the problems.
- The assignment definitely exemplifies a "mastery" of the necessary math skills.

C

- The student attempts to solve the problems in the correct manner, yet often fails to produce correct results due to major algebraic errors or other incorrect techniques.
- OR The assignment demonstrates a grasp of the mathematics involved and the reasoning needed to solve the problems, yet arrives at the answers without showing supporting work or through methods other than those specified in the directions.
- OR More than two noticeable flaws (listed above) exist throughout the assignment or problem set.
- OR The assignment is obviously less than complete.

D

- The student demonstrates some familiarity with the problems, yet the assignment falls very short of completion due to a lack of comprehension about some of the mathematics involved.
- OR The assignment contains combinations of the errors discussed above for a 'C.'
- OR The assignment is only halfway complete.

<D

- The student clearly demonstrates an attempt, yet indicates very little knowledge about the problems or their solution.
- OR The assignment is very incomplete.

Name _____ Honors PreCalculus Project

ACCURACY/CRITERIA

(Inclusion of necessary items, following of specifications, depth of research)

Functions 0 1 2
 Domain 0 1 2
 Range 0 1 2
 Zeros 0 1 2
 11 Basic Functions 0 1 2
 Composition of Functions 0 1 2
 Size/Shape 0 1 2
 Name Card 0 1 2
 Visible Title 0 1 2
 Border/Bulletin Board Appearance 0 1 2

_____ Other ± _____

TOTAL Accuracy/Criteria Points _____

TOTAL POINTS ABOVE _____

x2.5

Preliminary Sketch _____

Bonus (if any) _____

NEATNESS/PLANNING

Layout/Lettering 0 1 2
 Coherence 0 1 2
 Effort 0 1 2
 Artwork/Materials 0 1 2

_____ Other ± _____

TOTAL Neatness/Planning Points _____

CREATIVITY

Effectiveness 0 1 2
 Use of Colors/Materials 0 1 2
 Originality 0 1 2

_____ Other ± _____

TOTAL Creativity Points _____

FINAL GRADE _____

Appendix F

Arguments Presented to Participants During the Study

Argument L

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

Proof: If $n^2 = 9$, then $n = 3$.

↑ ↑
divisible also divisible
by 3 by 3

If $n^2 = 36$, then $n = 6$.

If $n^2 = 81$, then $n = 9$.

Thus, by induction, if n^2 is a multiple of 3, so is n .

Argument M

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

Proof: Let n^2 be a multiple of 3; that is, $n^2 = 3k$ for some integer k .

Since $n^2 = 3k$, then $n \cdot n = 3k$.

So, $n = \frac{3k}{n}$ or $n = 3\left(\frac{k}{n}\right)$.

Thus, n is a multiple of 3.

Hence, n^2 is a multiple of 3 implies n is a multiple of 3.

Argument E

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

Proof: Let n be a positive integer such that $n = 3k$ for some integer k .

$$\text{Then } n^2 = (3k)^2 = 9k^2 = 3(3k^2).$$

$$\text{That is, } n^2 = 3q \text{ for some integer } q = 3k^2.$$

Hence, if n^2 is a multiple of 3, then n is a multiple of 3.

Argument J

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

Proof: Assume that n^2 is an odd positive integer that is divisible by 3.

$$\text{That is, } n^2 = (3n+1)^2 = 9n^2 + 6n + 1 = 3n(n+2) + 1.$$

Therefore, n^2 is divisible by 3.

Assume that n^2 is even and a multiple of 3.

$$\text{That is, } n^2 = (3n)^2 = 9n^2 = 3n(3n).$$

Therefore, n^2 is a multiple of 3.

If we factor $n^2 = 9n^2$, we get $3n(3n)$, which means n is a multiple of 3.

Argument G

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

Proof: Let n be a positive integer such that n^2 is a multiple of 3.

Then $n = 3m$, where $m \in \mathbb{Z}^+$ (set of positive integers).

So, $n^2 = (3m)^2 = 9m^2 = 3(3m^2)$.

This breaks down into $3m$ times $3m$, which shows m is a multiple of 3.

Argument K

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

Proof: Let n be an integer such that $n^2 = 3x$, where x is any integer.

Then 3 divides n^2 , or $3 \mid n^2$.

Since $n^2 = 3x$, then $n \mid 3x$. Thus, $3 \mid n$.

Therefore, if n^2 is a multiple of 3, then n is a multiple of 3.

Argument Y

Statement: If a number is not divisible by 2, then it is not divisible by 6.

Proof: Let n be a number that is divisible by 6.

Then $n = 6k$ for some integer k .

Hence, $n = 2(3k) = 2q$ for some integer q .

Therefore, n is divisible by 2.

Thus, by contrapositive, the statement is correct.

Argument P

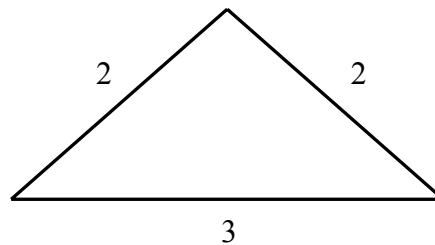
Statement: In a triangle with side lengths a , b , and c , if $a^2 + b^2 = c^2$, then the triangle is a right triangle.

Proof: Consider the triangle shown with side lengths $a = 2$, $b = 2$, and $c = 3$.

Then $a^2 + b^2 \neq c^2$ since $2^2 + 2^2 \neq 3^2$.

Therefore, the triangle is not a right triangle.

Thus, by contrapositive, the statement is correct.



Argument Q

Statement: In a triangle with side lengths a , b , and c , if $a^2 + b^2 = c^2$, then the triangle is a right triangle.

Proof: Consider an equilateral triangle ABC with side lengths of s (with $s \neq 0$).

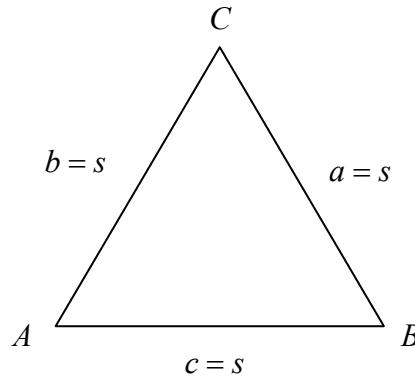
That is, $a = s$, $b = s$, and $c = s$.

Since $m\angle A = m\angle B = m\angle C = 60^\circ$, triangle ABC is not a right triangle.

Also, $a^2 + b^2 \neq c^2$ since $s^2 + s^2 \neq s^2$ (with $s \neq 0$).

That is, triangle ABC is not a right triangle implies $a^2 + b^2 \neq c^2$.

Thus, by contrapositive, the statement is correct.



Appendix G

Writing Prompts

The following prompts were used to solicit written responses from study participants. Participants completed responses for Prompts 1, 2, and 4 in class the same day the prompts were presented. Participants completed a response for Prompt 3 as homework.

Prompt 1

To ascertain what you already know about rubrics, provide a one or two sentence response to each of the following items.

- 1) What is a rubric?
- 2) Describe your experience(s) with rubrics. (What courses have you been in that used rubrics for grading your work? Was the rubric approach helpful? Have you ever graded student work using a rubric?)
- 3) Why or how is a rubric helpful?
- 4) What do you believe is involved in developing rubrics for scoring student work?

Prompt 2

- 1) How does the new score for Argument L (after discussion) compare/relate to the original score you gave the “proof” last week?
- 2) Do you think the new score or the previous score is more appropriate? (fair, accurate) Explain briefly.
- 3) Has developing the rubric reminded you of or taught you anything about proofs? Explain.

Prompt 3

Do you prefer scoring the proofs individually or in a group? Why? What is an advantage of each?

Prompt 4

- 1) What was most beneficial for you from the work with scoring rubrics for proofs?
- 2) Do you feel this work improved your facility with proofs? How so, or why not?
- 3) Would you be very likely to use scoring rubrics in your own classroom?
(Maybe not for proofs; probably not for every assignment) Why or why not?

References

- Almeida, D. (1996). Variations in proof standards: Implications for mathematics education. *International Journal of Mathematical Education in Science and Technology*, 27, 659–665.
- Bell, A. (1976). A study of pupils' proof-explanations in mathematical situations. *Educational Studies in Mathematics*, 7, 23–40.
- Chazan, D. (1993). High school geometry students' justification for their views of empirical evidence and mathematical proof. *Educational Studies in Mathematics*, 24, 359–387.
- Colorado Department of Education. (2003, September 26). *3-Point Rubric for Medium Constructed-Response Items*. Retrieved October 27, 2003, from the Unit of Student Assessment's CSAP Scoring information website: http://www.cde.state.co.us/cdeassess/documents/csap/rubrics/Math_3pt_Rubric.html
- Creswell, J. W. (2003). *Research Design: Qualitative, Quantitative, and Mixed Method Approaches*. Thousand Oaks, CA: Sage Publications.
- de Villiers, M. (1995). An alternative introduction to proof in dynamic geometry. *Micromath*, 11(1), 14–19.
- Dean, E. (1996). Teaching the proof process: A model for discovery learning. *College Teaching*, 44, 52–55.
- Epp, S. (1994). The role of proof in problem solving. In A. Schoenfeld (Ed.), *Mathematical thinking and problem solving* (pp. 257–269). Hillsdale, NJ: Erlbaum.

- Even, R. (1993). Subject-matter knowledge and pedagogical content knowledge: Prospective secondary teachers and the function concept. *Journal for Research in Mathematics Education*, 24, 94–116.
- Exemplars®. (2003). *Classic Math Rubric*. Retrieved October 27, 2003, from <http://www.exemplars.com/resources/rubrics/classic.html>
- Fischbein, E. (1982). Intuition and proof. *For the Learning of Mathematics*, 3(2), 9–18, 24.
- Goff, J. (2002). Using composition techniques to improve classroom instruction and students' understanding of proof. *Primus*, 12, 239–246.
- Hanna, G. (1989). Proofs that prove and proofs that explain. In G. Vergnaud, J. Rogalski, & M. Artigue (Eds.), *Proceedings of the Thirteenth International Conference for the Psychology of Mathematics Education* (Vol. 2, pp. 45–51). Paris: CNRS—Paris V.
- Hanna, G. (1995). Challenges to the importance of proof. *For the Learning of Mathematics*, 15(3), 42–49.
- Hanna, G. (2000). Proof, explanation and exploration: An overview. *Educational Studies in Mathematics*, 44, 5–23.
- Harel, G., & Sowder, L. (1998). Students' proof schemes: Results from exploratory studies. In A. Schoenfeld, J. Kaput, & E. Dubinsky (Eds.), *Issues in mathematics education, Vol. 7. Research in collegiate mathematics education, III* (pp. 234–283). Providence, RI: American Mathematical Society.
- Healy, L., & Hoyles, C. (2000). A study of proof conceptions in algebra. *Journal for Research in Mathematics Education*, 31, 396–428.

- Herbst, P. (2002). Engaging students in proving: A double bind on the teacher. *Journal for Research in Mathematics Education*, 33, 176–203.
- Hersh, R. (1993). Proving is convincing and explaining. *Educational Studies in Mathematics*, 24, 389–399.
- Hewitt, D. (1996). Towards proof. *Mathematics Teaching*, 155, 27–31.
- Hoyles, C. (1997). The curricular shaping of students' approaches to proof. *For the Learning of Mathematics*, 17(1), 7–16.
- Jones, K. (2000). The student experience of mathematical proof at university level. *International Journal of Mathematical Education in Science and Technology*, 31, 53–60.
- Kleiner, I., & Movshovitz-Hadar, N. (1990). Aspects of the pluralistic nature of proofs. *Interchange*, 21(1), 28–25.
- Knuth, E. (2002a). Secondary school mathematics teachers' conception of proof. *Journal for Research in Mathematics Education*, 33, 379–405.
- Knuth, E. (2002b). Teachers' conceptions of proof in the context of secondary school mathematics. *Journal of Mathematics Teacher Education*, 5, 61–88.
- MacKernan, J. (1996). What's the point of proof? *Mathematics Teaching*, 155, 14–20.
- Moore, R. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, 27, 249–266.
- Movshovitz-Hadar, N. (1988). School mathematics theorems—an endless source of surprise. *For the Learning of Mathematics*, 8(3), 34–40.
- National Council of Teachers of Mathematics. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

- O'Brien, R. (1998) *An Overview of the Methodological Approach of Action Research*.
Retrieved June 13, 2006, from <http://www.web.ca/~robrien/papers/arfinal.html>
- Selden, A., & Selden, J. (2003). Validations of proofs considered as texts: Can undergraduates tell whether an argument proves a theorem? *Journal for Research in Mathematics Education*, 34, 4–36.
- Selden, J., & Selden, A. (1995). Unpacking the logic of mathematical statements. *Educational Studies in Mathematics*, 29, 123–151.
- Senk, S. (1985). How well do students write geometry proofs? *Mathematics Teacher*, 78, 448–456.
- Slomson, A. (1996). Teaching and the role of proof. *Mathematics Teaching*, 155, 23–26.
- Sowder, L., & Harel, G. (1998). Types of students' justifications. *Mathematics Teacher*, 91, 670–675.
- Thompson, D. (1996). Learning and teaching indirect proof. *Mathematics Teacher*, 89, 474–481.
- Touval, A. (1997). Investigating a definite integral—from graphing calculator to rigorous proof. *Mathematics Teacher*, 90, 230–232.
- Wheeler, D. (1990). Aspects of mathematical proof. *Interchange*, 21(1), 1–5.