Proof Construction and Collaborative Revision in Undergraduate Mathematics

BY
EMILY CILLI-TURNER
B.S., Colorado State University, May 2004
M.A., University of Colorado Boulder, May 2006

THESIS
Submitted as partial fulfillment of the requirements for the degree of Doctor of Arts in Mathematics in the Graduate College of the University of Illinois at Chicago, 2013

Chicago, Illinois

Thesis Committee:
Dr. Mara Martinez, Chair & Advisor
Dr. John Baldwin
Dr. Alison Castro Superfine
Dr. Bonnie Saunders
Dr. Greg Larnell, Curriculum & Instruction
To my parents, Wendy Turner and Dean Cilli, for always believing in me even when I did not believe in myself.
ACKNOWLEDGMENTS

I would first like to thank the wonderful people that have advised and supported me during this endeavor.

Dr. Mara Martinez, thank you for being my advisor. Your encouragement, guidance, patience and support was most appreciated and without which this project would not have been possible. To my committee – Dr. John Baldwin, Dr. Alison Castro Superfine, Dr. Bonnie Saunders and Dr. Greg Larnell – thank you for your helpful comments throughout this process.

I also have amazing colleagues who I have leaned on and learned from during this process.

Andrew Brasile, thank you for helping me with some of the qualitative analysis and for helping to keep me motivated. You always cheer me up and you are the best person to take a road trip with. Gail Tang, thank you for your always helpful advice throughout this process and your assistance with the \LaTeX code. Holly Krieger, you have always believed in my abilities and you have been a great friend to me at UIC.

Finally, I would like to thank some of my wonderful friends that allowed me to take a break once in awhile.

Grant Penfield Haugen, you are one of the funniest and wittiest people I know and we have had amazing adventures together. Alec Boyle, thank you for believing in me. I’m so glad that I met you and I love you so much.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Major Goals of this Study</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Research Questions</td>
<td>3</td>
</tr>
<tr>
<td>1.3 Methodology</td>
<td>4</td>
</tr>
<tr>
<td>1.4 Organization of the Dissertation</td>
<td>5</td>
</tr>
</tbody>
</table>

| 2 FRAMEWORK | 7 |
| 2.1 Proof Frameworks | 8 |
| 2.1.1 Meaning of Proof | 8 |
| 2.1.2 Functions of Proof | 13 |
| 2.1.3 Proof Schemes | 20 |
| 2.2 Summary of Proof Frameworks | 26 |
| 2.3 Collaborative Learning | 27 |
| 2.4 Conclusion | 29 |

| 3 LITERATURE REVIEW | 31 |
| 3.1 Importance of Proof | 31 |
| 3.2 Students’ Understanding of the Functions of Proof | 33 |
| 3.3 Proof Validation | 35 |
| 3.4 Proof Construction | 39 |
| 3.5 Collaborative Learning & Revision | 42 |
| 3.6 Summary | 44 |

| 4 METHODOLOGY & DATA COLLECTION | 46 |
| 4.1 Context | 47 |
| 4.2 Courses & Participants | 48 |
| 4.2.1 Treatment Group | 48 |
| 4.2.1.1 The Emerging Scholars Program (ESP) | 49 |
| 4.2.1.2 Treatment Group Participants | 52 |
| 4.2.2 Comparison Group | 54 |
| 4.2.2.1 Comparison Group Participants | 55 |
| 4.2.3 Similarities and Differences Between Group Participants | 56 |
TABLE OF CONTENTS (Continued)

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3 Data Sources</td>
<td>59</td>
</tr>
<tr>
<td>4.3.1 Assessments</td>
<td>60</td>
</tr>
<tr>
<td>4.3.2 Interviews</td>
<td>63</td>
</tr>
<tr>
<td>4.3.3 Classroom Observations</td>
<td>65</td>
</tr>
<tr>
<td>4.3.4 Proof Portfolios</td>
<td>67</td>
</tr>
<tr>
<td>4.4 Description of Teaching Experiment</td>
<td>68</td>
</tr>
<tr>
<td>4.4.1 Classroom Activities</td>
<td>70</td>
</tr>
<tr>
<td>4.4.2 Role of the teacher</td>
<td>72</td>
</tr>
<tr>
<td>4.4.3 Comparison to Modified Moore Method</td>
<td>73</td>
</tr>
<tr>
<td>4.5 Description of Comparison Course</td>
<td>74</td>
</tr>
<tr>
<td>4.6 Summary</td>
<td>75</td>
</tr>
<tr>
<td>5 DATA ANALYSIS &amp; RESULTS</td>
<td>77</td>
</tr>
<tr>
<td>5.1 Students' Beliefs About Proof Functions</td>
<td>78</td>
</tr>
<tr>
<td>5.1.1 Data Analysis</td>
<td>78</td>
</tr>
<tr>
<td>5.1.1.1 Qualitative Data Analysis</td>
<td>79</td>
</tr>
<tr>
<td>5.1.1.2 Quantitative Data Analysis</td>
<td>83</td>
</tr>
<tr>
<td>5.1.2 Qualitative Results</td>
<td>86</td>
</tr>
<tr>
<td>5.1.2.1 Themes</td>
<td>86</td>
</tr>
<tr>
<td>5.1.2.2 Treatment Group Qualitative Results</td>
<td>94</td>
</tr>
<tr>
<td>5.1.2.3 Comparison Group Qualitative Results</td>
<td>97</td>
</tr>
<tr>
<td>5.1.2.4 Between Groups Qualitative Results</td>
<td>99</td>
</tr>
<tr>
<td>5.1.3 Quantitative Results</td>
<td>101</td>
</tr>
<tr>
<td>5.1.3.1 Treatment Group Quantitative Results</td>
<td>102</td>
</tr>
<tr>
<td>5.1.3.2 Comparison Group Quantitative Results</td>
<td>107</td>
</tr>
<tr>
<td>5.1.3.3 Between Groups Quantitative Results</td>
<td>111</td>
</tr>
<tr>
<td>5.1.4 Synthesis of Results on Students' Proof Function Beliefs</td>
<td>117</td>
</tr>
<tr>
<td>5.2 Students' Proof Validation Skills</td>
<td>119</td>
</tr>
<tr>
<td>5.2.1 Quantitative Data Analysis</td>
<td>120</td>
</tr>
<tr>
<td>5.2.2 Quantitative Results</td>
<td>126</td>
</tr>
<tr>
<td>5.2.2.1 Treatment Group Quantitative Results</td>
<td>126</td>
</tr>
<tr>
<td>5.2.2.2 Comparison Group Quantitative Results</td>
<td>132</td>
</tr>
<tr>
<td>5.2.2.3 Quantitative Differences Between Groups</td>
<td>136</td>
</tr>
<tr>
<td>5.2.3 Interview Data Analysis</td>
<td>143</td>
</tr>
<tr>
<td>5.2.4 Treatment Group Interview Results</td>
<td>144</td>
</tr>
<tr>
<td>5.2.4.1 Stephanie’s Proof Schemes</td>
<td>144</td>
</tr>
<tr>
<td>5.2.4.2 Robert’s Proof Schemes</td>
<td>149</td>
</tr>
<tr>
<td>5.2.5 Comparison Group Interview Results</td>
<td>152</td>
</tr>
<tr>
<td>5.2.5.1 James’ Proof Schemes</td>
<td>152</td>
</tr>
<tr>
<td>5.2.5.2 Francine’s Proof Schemes</td>
<td>156</td>
</tr>
</tbody>
</table>
# TABLE OF CONTENTS (Continued)

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2.6</td>
<td>Between Groups Interview Results</td>
</tr>
<tr>
<td>5.2.7</td>
<td>Synthesis of Results on Students’ Proof Validation Skills</td>
</tr>
<tr>
<td>5.3</td>
<td>Students’ Proof Construction Skills</td>
</tr>
<tr>
<td>5.3.1</td>
<td>Data Analysis</td>
</tr>
<tr>
<td>5.3.2</td>
<td>Results on Students’ Proof Construction Skills</td>
</tr>
<tr>
<td>5.3.2.1</td>
<td>Treatment Group Results</td>
</tr>
<tr>
<td>5.3.2.2</td>
<td>Comparison Group Results</td>
</tr>
<tr>
<td>5.3.2.3</td>
<td>Between Groups Results</td>
</tr>
<tr>
<td>5.3.3</td>
<td>Synthesis of Results on Students’ Proof Construction Skills</td>
</tr>
<tr>
<td>5.4</td>
<td>Conclusion</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>6 DISCUSSION</th>
<th>184</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1 Summary of Results</td>
<td>185</td>
</tr>
<tr>
<td>6.2 Limitations</td>
<td>189</td>
</tr>
<tr>
<td>6.2.1 Proof Portfolio Limitations</td>
<td>192</td>
</tr>
<tr>
<td>6.3 Teaching Implications</td>
<td>195</td>
</tr>
<tr>
<td>6.4 Further Research</td>
<td>199</td>
</tr>
<tr>
<td>6.5 Conclusion</td>
<td>201</td>
</tr>
</tbody>
</table>

| CITED LITERATURE | 202 |
| APPENDICES | 210 |
| Appendix A | 211 |
| Appendix B | 215 |
| Appendix C | 216 |
| Appendix D | 218 |
| Appendix E | 219 |
| Appendix F | 220 |
| Appendix G | 229 |

| VITA | 234 |
LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>SUMMARY OF CHARACTERISTICS OF PARTICIPANTS IN THE TREATMENT GROUP</td>
</tr>
<tr>
<td>II</td>
<td>SUMMARY OF CHARACTERISTICS OF PARTICIPANTS IN THE COMPARISON GROUP</td>
</tr>
<tr>
<td>III</td>
<td>FREQUENCY OF GENDER AMONG STUDENT PARTICIPANTS IN THE STUDY</td>
</tr>
<tr>
<td>IV</td>
<td>FREQUENCY OF MAJORS AMONG STUDENT PARTICIPANTS IN THE STUDY</td>
</tr>
<tr>
<td>V</td>
<td>MEANS AND STANDARD DEVIATIONS OF GPAS FOR ALL STUDENT PARTICIPANTS IN THE STUDY</td>
</tr>
<tr>
<td>VI</td>
<td>FREQUENCY OF STUDENTS IN THE TREATMENT AND COMPARISON GROUPS WHO HAVE TAKEN PROOF-BASED MATHEMATICS COURSES PRIOR TO ENROLLMENT IN THIS COURSE</td>
</tr>
<tr>
<td>VII</td>
<td>CODES USED FOR QUALITATIVE ANALYSIS OF PROOF FUNCTION DATA</td>
</tr>
<tr>
<td>VIII</td>
<td>STATEMENTS ABOUT PROOF FUNCTIONS GIVEN TO STUDENTS ON THE ASSESSMENT</td>
</tr>
<tr>
<td>IX</td>
<td>FREQUENCY OF COMMENTS ON BOTH ASSESSMENTS UNDER EACH PROOF FUNCTION THEME FOR THE TREATMENT GROUP</td>
</tr>
<tr>
<td>X</td>
<td>CODES APPLIED FOR EACH STUDENT IN THE TREATMENT GROUP ON THE PRE AND POST-ASSESSMENT</td>
</tr>
<tr>
<td>XI</td>
<td>FREQUENCY OF COMMENTS ON BOTH ASSESSMENTS UNDER EACH PROOF FUNCTION THEME FOR THE COMPARISON GROUP</td>
</tr>
<tr>
<td>XII</td>
<td>CODES APPLIED FOR EACH STUDENT IN THE COMPARISON GROUP ON THE PRE AND POST-ASSESSMENT</td>
</tr>
</tbody>
</table>
### LIST OF TABLES (Continued)

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>XIII</td>
<td>FREQUENCY OF COMMENTS FOR EACH PROOF FUNCTION TYPE ON THE EACH ASSESSMENT</td>
</tr>
<tr>
<td>XIV</td>
<td>MEANS AND STANDARD DEVIATIONS OF EACH LIKERT ITEM ON THE PRE-ASSESSMENT FOR ALL STUDENTS</td>
</tr>
<tr>
<td>XV</td>
<td>MEANS AND STANDARD DEVIATIONS OF LIKERT ITEM RESPONSES ON THE PRE AND POST-ASSESSMENTS OF THE TREATMENT GROUP</td>
</tr>
<tr>
<td>XVI</td>
<td>MEAN DIFFERENCES AND CORRELATION COEFFICIENTS FOR PRE AND POST-ASSESSMENT RESPONSES FOR THE TREATMENT GROUP</td>
</tr>
<tr>
<td>XVII</td>
<td>MEAN AND STANDARD DEVIATION OF THE TOTAL FUNCTION SCORE (TFS) FOR THE TREATMENT GROUP ON EACH ASSESSMENT</td>
</tr>
<tr>
<td>XVIII</td>
<td>MEANS AND STANDARD DEVIATIONS OF LIKERT ITEM RESPONSES ON THE PRE AND POST-ASSESSMENTS OF THE COMPARISON GROUP</td>
</tr>
<tr>
<td>XIX</td>
<td>MEAN DIFFERENCES AND CORRELATION COEFFICIENTS FOR PRE AND POST-ASSESSMENT RESPONSES FOR THE COMPARISON GROUP</td>
</tr>
<tr>
<td>XX</td>
<td>MEAN AND STANDARD DEVIATION OF THE TOTAL FUNCTION SCORE (TFS) FOR THE COMPARISON GROUP ON EACH ASSESSMENT</td>
</tr>
<tr>
<td>XXI</td>
<td>AVERAGE NUMBER OF PROOF FUNCTIONS IDENTIFIED PER STUDENT ON EACH ASSESSMENT</td>
</tr>
<tr>
<td>XXII</td>
<td>MEANS AND STANDARD DEVIATIONS OF EACH LIKERT ITEM FOR ALL STUDENTS ON THE PRE-ASSESSMENT, SEPARATED BY GROUP</td>
</tr>
<tr>
<td>XXIII</td>
<td>MEANS AND STANDARD DEVIATIONS OF EACH LIKERT ITEM ON THE POST-ASSESSMENT</td>
</tr>
<tr>
<td>XXIV</td>
<td>MEAN AND STANDARD DEVIATION OF TOTAL PROOF FUNCTION SCORES ON EACH ASSESSMENT SEPARATED BY GROUP</td>
</tr>
<tr>
<td>TABLE</td>
<td>PAGE</td>
</tr>
<tr>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>XXV</td>
<td>FREQUENCY AND PERCENTAGE OF STUDENTS IN THE TREATMENT GROUP WHO CORRECTLY CLASSIFIED EACH PROOF</td>
</tr>
<tr>
<td>XXVI</td>
<td>MEANS AND STANDARD DEVIATIONS OF STUDENTS' SELF-REPORTED UNDERSTANDING OF EACH TYPE OF PROOF FOR THE TREATMENT GROUP</td>
</tr>
<tr>
<td>XXVII</td>
<td>MEANS AND STANDARD DEVIATIONS OF STUDENTS' SELF-REPORTED CONFIDENCE ABOUT CORRECTLY CLASSIFYING EACH TYPE OF PROOF FOR THE TREATMENT GROUP</td>
</tr>
<tr>
<td>XXVIII</td>
<td>FREQUENCY AND PERCENTAGE OF STUDENTS IN THE COMPARISON GROUP WHO CORRECTLY CLASSIFIED EACH PROOF</td>
</tr>
<tr>
<td>XXIX</td>
<td>MEANS AND STANDARD DEVIATIONS OF STUDENTS' SELF-REPORTED UNDERSTANDING OF EACH TYPE OF PROOF FOR THE COMPARISON GROUP</td>
</tr>
<tr>
<td>XXX</td>
<td>MEANS AND STANDARD DEVIATIONS OF STUDENTS' SELF-REPORTED CONFIDENCE ABOUT CORRECTLY CLASSIFYING EACH TYPE OF PROOF FOR THE COMPARISON GROUP</td>
</tr>
<tr>
<td>XXXI</td>
<td>MEANS AND STANDARD DEVIATIONS FOR PRE AND POST-ASSESSMENT RESPONSES FROM BOTH GROUPS</td>
</tr>
<tr>
<td>XXXII</td>
<td>FREQUENCY AND PERCENTAGE OF STUDENTS IN BOTH GROUPS WHO CORRECTLY CLASSIFIED EACH PROOF TYPE</td>
</tr>
<tr>
<td>XXXIII</td>
<td>MEANS AND STANDARD DEVIATIONS OF STUDENTS' SELF-REPORTED UNDERSTANDING OF EACH TYPE OF PROOF</td>
</tr>
<tr>
<td>XXXIV</td>
<td>MEANS AND STANDARD DEVIATIONS OF STUDENTS' SELF-REPORTED CONFIDENCE ABOUT CORRECTLY EVALUATING EACH TYPE OF PROOF</td>
</tr>
<tr>
<td>XXXV</td>
<td>PROOF SCHEMES EXHIBITED BY STUDENTS IN EACH GROUP DURING EACH INTERVIEW</td>
</tr>
<tr>
<td>XXXVI</td>
<td>TYPES OF TREATMENT GROUP STUDENTS' PROOFS SEPARATED BY ASSESSMENT</td>
</tr>
</tbody>
</table>
# LIST OF TABLES (Continued)

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>XXXVII A SUMMARY OF PROOF APPROACHES TAKEN BY STUDENTS IN THE TREATMENT GROUP ON EACH ASSESSMENT</td>
<td>173</td>
</tr>
<tr>
<td>XXXVIII FREQUENCY OF TREATMENT GROUP STUDENTS’ PROOF ERRORS ON THE ASSESSMENTS SEPARATED BY ERROR CODE</td>
<td>174</td>
</tr>
<tr>
<td>XXXIX TYPES OF COMPARISON GROUP STUDENTS’ PROOFS SEPARATED BY ASSESSMENT</td>
<td>176</td>
</tr>
<tr>
<td>XL A SUMMARY OF PROOF APPROACHES TAKEN BY STUDENTS IN THE TREATMENT GROUP ON EACH ASSESSMENT</td>
<td>176</td>
</tr>
<tr>
<td>XLI FREQUENCY OF COMPARISON GROUP STUDENTS’ PROOF ERRORS ON THE ASSESSMENTS SEPARATED BY ERROR CODE</td>
<td>177</td>
</tr>
<tr>
<td>XLII AN ANALYSIS OF STUDENTS’ PROOF ERRORS MADE ON THE ASSESSMENTS SEPARATED BY ERROR CATEGORY</td>
<td>180</td>
</tr>
<tr>
<td>XLIII TYPES OF IMPROVEMENTS MADE ON STUDENTS’ PROOFS SEPARATED BY GROUP</td>
<td>181</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A proof of “The sum of the first n positive integers, $S(n)$, is equal to $\frac{n(n + 1)}{2}$” that proves (Hanna, 1990, pg. 10).</td>
</tr>
<tr>
<td>2</td>
<td>A proof of “The sum of the first n positive integers, $S(n)$, is equal to $\frac{n(n + 1)}{2}$” that explains (Hanna, 1990, pg. 10).</td>
</tr>
<tr>
<td>3</td>
<td>A proof of “The sum of the first n positive integers, $S(n)$, is equal to $\frac{n(n + 1)}{2}$” that explains how the formula is obtained and why it works for all natural numbers (J. Baldwin, personal communication, September 30, 2012).</td>
</tr>
<tr>
<td>4</td>
<td>The organization of students’ proof schemes (Harel and Sowder, 1998).</td>
</tr>
<tr>
<td>5</td>
<td>The informal proof on the assessment given to students.</td>
</tr>
<tr>
<td>6</td>
<td>The empirical proof on the assessment given to students.</td>
</tr>
<tr>
<td>7</td>
<td>The invalid deductive proof on the assessment given to students.</td>
</tr>
<tr>
<td>8</td>
<td>The valid deductive proof on the assessment given to students.</td>
</tr>
<tr>
<td>9</td>
<td>The questions regarding understanding and confidence given to students on the assessment.</td>
</tr>
<tr>
<td>10</td>
<td>Frequency distributions of correct percentages on the pre-assessment (left) and post-assessment (right).</td>
</tr>
<tr>
<td>11</td>
<td>The Proof Error Evaluation Tool (PEET) (Andrew, 2009).</td>
</tr>
<tr>
<td>12</td>
<td>The PEET tool (Andrew, 2009) used to analyze errors from a student proof in Selden and Selden (2003).</td>
</tr>
<tr>
<td>13</td>
<td>A student proof from a pre-assessment and the coding applied.</td>
</tr>
<tr>
<td>14</td>
<td>Another student proof from a post-assessment and the coding applied.</td>
</tr>
<tr>
<td>FIGURE</td>
<td>PAGE</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>15</td>
<td>An example of a student proof that was classified as weak valid.</td>
</tr>
<tr>
<td>16</td>
<td>Two student proofs that were classified as incomplete.</td>
</tr>
</tbody>
</table>
## LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>CUPM</td>
<td>Committee on the Undergraduate Program in Mathematics</td>
</tr>
<tr>
<td>ESP</td>
<td>Emerging Scholars Program</td>
</tr>
<tr>
<td>GPA</td>
<td>Grade Point Average</td>
</tr>
<tr>
<td>IBL</td>
<td>Inquiry-Based Learning</td>
</tr>
<tr>
<td>MSCS</td>
<td>Mathematics, Statistics &amp; Computer Science</td>
</tr>
<tr>
<td>PEET</td>
<td>Proof Error Evaluation Tool</td>
</tr>
<tr>
<td>TFS</td>
<td>Total Function Score</td>
</tr>
</tbody>
</table>
SUMMARY

A study of the impacts of a teaching experiment utilizing a technique called collaborative revision on students’ learning around proof was carried out using a mixed-methods approach. The study was conducted in the context of an introduction to proof course at a large midwestern university. A treatment group, consisting of students enrolled in the course where the teaching experiment was enacted, and a comparison group were used. Student participants were given pre and post-assessments to measure their beliefs about proof functions, proof validation skills and proof construction skills. Additionally, several students were interviewed to determine their thought processes when completing the proof validation tasks and to identify the proof schemes they were exhibiting. Observations were also performed to contrast the collaborative revision classroom learning environment with a comparison classroom.

The findings reveal that collaborative revision does impact student learning in several ways. Students in the treatment group displayed a greater appreciation of several of functions of proof than students in the comparison group on the post-assessment, despite both groups reporting comparable beliefs on the pre-assessment. Neither the treatment group or the comparison group were able to significantly increase their accuracy in validating proofs from the pre to the post-assessment, yet the treatment group reported greater confidence and understanding on the proof validation tasks. Furthermore, even though none of the students were able to write an acceptable proof on the pre-assessment, almost half of the treatment group students were able to construct an acceptable proof on the post-assessment compared to only a quarter of the
SUMMARY (Continued)

students in the comparison group. Implications from this study can aid future educators when
designing classroom interactions to aid students in constructing proofs and understanding the
diverse functions of proof.
CHAPTER 1

INTRODUCTION

The Committee on the Undergraduate Program in Mathematics (CUPM), appointed by the Mathematical Association of America, recommend that “in order to recruit and retain majors and minors, mathematical sciences departments should put a high priority on effective and engaging teaching in introductory courses” (Committee on the Undergraduate Program of Mathematics, 2004). This research project puts forth a novel teaching experiment that was implemented in an introduction to proof course and the impact of this intervention on students’ proving skills is measured to determine its effectiveness. The classroom intervention was designed based on teaching and learning principles from the literature and was enacted during one semester of an introduction to proof course. In what follows, an overview of the research that this project draws from and builds upon will be given as well as a description of how the teaching experiment was administered and how learning outcomes were measured. Further, we detail the data analysis techniques used and interpret the results obtained from this study.

1.1 Major Goals of this Study

The need for instruction that succeeds in teaching students about proof is critical because, as previous research projects have shown (e.g Weber, 2010; Dreyfus, 1999), many undergraduate students struggle in proof-based mathematics courses. In particular, many students emerge
from an introduction to proof course without being able to construct proofs (Sowder and Harel, 2003; Moore, 1994), determine if a given proof is valid (Selden and Selden, 2003; Brandt and Rimmasch, 2012) or understand the function of proof in mathematics (Almeida, 1995; Coe and Ruthven, 1994). Yet in more advanced mathematics courses, proof is essential for communication of ideas and for assessment of students (Weber, 2010) and, thus, there is a need to improve instruction on proof to assist students in the learning process. The importance of proof in the discipline of mathematics and the disappointing data from students measuring students’ proof construction and validation skills warrants the development of new techniques to assist students in learning about proof.

Therefore, this study aimed to investigate the impact on students’ proof validation and construction skills from participating in a teaching experiment using a process called collaborative revision in an introduction to proof course. Collaborative revision refers to the process in which students present a proof they have written to their classmates, who are then encouraged to give critiques to aid the student in revising and resubmitting the proof. It was hypothesized that encouraging students to read proofs for correctness and point out inconsistencies in their classmates proofs during the collaborative revision process would help them to determine the validity a given proof as well as become better proof-writers. Thus, this study was designed to measure the learning gains in proof validation skills and proof construction skills made by students in this experimental course.

Collaborative revision presents proof not as merely a pedantic exercise, but also as a powerful communication and explanation tool, thus putting students directly into an authentic math-
ematical experience. Previous studies (Almeida, 1995; Coe and Ruthven, 1994) have shown that undergraduate students after transition to proof courses may understand that proof can function as a verification or explanation tool, but do not appreciate the functions of proof beyond these. Thus, we also examined students’ beliefs about the functions of proof in mathematics to determine if collaborative revision impacted these beliefs.

1.2 Research Questions

In correspondence with the study goals outlined above, we wish to conclude if participation in collaborative revision results in learning gains for students in the areas of proof function beliefs, proof validation skills and proof construction skills. The study will utilize a treatment and comparison group in order to contrast proving skills between the students that participated in collaborative revision and those that did not. Specifically, we wish to determine the beliefs about proof function of students in the treatment and comparison groups before and after their respective courses. Additionally, the results from the treatment and comparison groups will be contrasted to determine what, if any, impacts collaborative revision has on these beliefs. Similarly, information about the proof construction and validation skills of students in each group (i.e. treatment and comparison) is desired and a comparison will be made to determine to what extent collaborative revision impacts these skills. To this end, we explore the following research questions:

1. (a) What are undergraduate students beliefs about the functions of proof in mathematics before and after a transition to proof course?
(b) Are there differences present in undergraduate students’ beliefs about proof function between students who participated in collaborative revision and those who did not? If so, what are they?

2. (a) What are undergraduate students’ proof validation skills before and after a transition to proof course?

(b) Are there differences present in proof validation skills between students who participated in collaborative revision and those who did not? If so, what are they?

3. (a) What are undergraduate students’ proof construction skills before and after a transition to proof course?

(b) Are there differences present in proof construction skills between students who participated in collaborative revision and those who did not? If so, what are they?

1.3 Methodology

This study used a treatment group comprised of students who participated in the collaborative revision process in addition to their enrollment in a lecture-based transition to proof course. To determine impacts of collaborative revision a comparison group was also used, comprised of students enrolled only in the lecture-based course. Students in each group (i.e. treatment and comparison) were administered pre and post-assessments designed to measure their appreciation of different functions of proof, their proof validation skills and their ability to construct a proof. Several students were also interviewed to determine the aspects of a proof that are convincing to them at the beginning of the courses and if these attributes evolve throughout
the semester. The research questions were addressed by a mixed-method analysis of this data collected from the treatment and comparison groups.

1.4 Organization of the Dissertation

We begin by outlining the framework underlying this study in detail in Chapter 2. This chapter includes important considerations that were made when designing the study and performing the data analysis, such as the definition of proof that is used for this research project and the different functions of proof from the literature, of which we measure students’ appreciation. Also, the concept of proof schemes as defined by Harel and Sowder (1998) is presented in preparation for use of this framework in data analysis of student interviews. To end this chapter, a rationale for the collaborative nature of the teaching experiment used in this research project is given.

A review of the literature related to this study and a discussion of how this study builds upon previous works follows in Chapter 3. Previous students regarding the importance of proof as well as students’ understanding of proof functions are summarized. Moreover, studies assessing students’ proof validation skills and proof construction skills are presented.

The methodology and data sources used in this research project are described in Chapter 4. A description of the context of this study and demographic data about participants is presented. Also, an overview of how the teaching experiment was enacted and various activities done in the treatment course is given. Finally, all of the data sources are detailed.

A summary of the data analysis techniques used to extract meaning from the data, in addition to the results of the data analysis, can be found in Chapter 5. For each of the three
research questions in this study, we present the analysis performed on data collected around the question and the relevant results. Results are presented as results from the treatment group, results from the comparison group and then between groups results. Each of the sections in this chapter finishes with a synthesis of the main results.

Finally, Chapter 6 summarizes the results and gives a discussion about what was found. The limitations of this study are discussed and teaching implications from our results are given. This chapter concludes with a presentation of further research that stems from this project.
CHAPTER 2

FRAMEWORK

This chapter gives an overview of the constructs that motivate the design and methodology used in this study. We begin with a consideration of the theoretical issues around proof needed to effectively answer the research questions. First, we address the definition of proof that we are working under on this research project. As there are multiple definitions of proof presented in the literature and the context of the research influences the definition used, it is important to chose one for this study. Secondly, we discuss the various roles that proof can play in mathematics as presented by mathematics education researchers. This is done because a goal of this study is to determine the beliefs that students hold about functions of proof and we use these proof functions as a grounding for the data analysis. Third, the idea of proof schemes developed by Harel and Sowder (1998) is presented as a way to categorize how students gain conviction about mathematical statements. Proof schemes are used as a framework for data analysis to answer the research question about any impact collaborative revision may have on students’ proof validation skills. This chapter concludes with an examination of collaborative learning in mathematics as this is a major consideration in the design of the teaching experiment utilized this study.
2.1 Proof Frameworks

The following three sections present an examination of important frameworks needed for meaningful results from this study. Since this study is mainly about issues around proving skills, it is essential that we consider the differing definitions of proof in various fields of mathematics and choose one to operate under. This becomes valuable when we consider what constitutes a valid proof later in this report. Moreover, frameworks are required for some of the data analysis in this study. The first research question in this study references the functions of proof in mathematics, which need to be defined and rooted in the literature. This is done here and a further discussion of students’ beliefs around proof functions is given in Chapter 3. Additionally, when studying students’ proof validation skills we wish to determine what aspects of a proof are convincing to students. These aspects are categorized and organized into a hierarchy by Harel and Sowder (1998) and proves useful in analyzing and interpreting data in this study.

2.1.1 Meaning of Proof

The answer to the seemingly simple question “what is proof?” turns out to be highly complex and context dependent. This section will review the dominant definitions of proof presented in three types of fields: mathematics, mathematical logic, and mathematics education. The definitions presented in each of these areas differ because each field interacts with proof in a distinct way. Thus, it is important to present the interpretation of the meaning of proof in each field before a definition for use in this study is settled upon.

We first turn to defining the concept of proof in the discipline of pure mathematics. A simple definition, given by Griffiths (2000), outlines the basic structure of a proof in mathematics: “A
mathematical proof is a formal and logical line of reasoning that begins with a set of axioms and moves through logical steps to a conclusion” (pg. 2). Hanna (1990) gives a broader definition of what she terms an *acceptable proof* in mathematics: “The proof must proceed from specific and accepted premises, must present an argument that is not flawed, and must lead to a result which, even if unexpected, seems upon reflection to make sense in the context of other mathematical knowledge” (p. 8). The idea of acceptable proof is useful to us because the context of this research is a transition to proof course, where many students may be constructing proofs for the first time. Thus, arguments made by the students most likely will not have the same level of rigor as those in journals or in graduate courses. To evaluate these student arguments, we need a definition of proof as a sequence of statements proceeding from axioms and definitions, but where certain steps may be omitted for the sake of the emergence of the ‘big picture’ of the argument.

However useful Hanna’s definition may be, there are some precautions to take when applying it. First, what exactly is meant by the phrase “that is not flawed.” It is assumed that Hanna meant that a proof relies on deductive inference using only statements that are assumed to be true (such as axioms and definitions) or have been already (or could be) proven. Second, this definition is ignorant of the audience for the proof. A result that an undergraduate in a transition to proof course understands in relation to other mathematical results is far different that one that a mathematician understands. Therefore, we propose a modification of this definition to make it more explicit and clear up confusion about how to apply it to arguments (with emphasis to denote changes): “A proof must proceed from specific and accepted premises,
must present an argument that *is deductive and uses valid statements*, and must lead to a result which, even if unexpected, seems upon reflection to make sense in the context of other mathematical knowledge *for the specified audience.*”

These definitions implicitly state that a proof is accepted as valid in the mathematical community by consensus and that proof has a social aspect to it. The mathematician Krantz (2007) incorporates this into his definition, which he states as “a proof is a rhetorical device for convincing another mathematician that a given statement (the theorem) is true” (pg. 3). Balacheff agrees and suggests that proving is very much a cultural artifact and defines proof as “an explanation accepted by a given community at a given time” (as quoted in Weber, 2003, p. 1). Additionally, a study by Inglis et al. (2013) show that mathematicians in different fields have discordant standards when evaluating a proof’s validity. Thus, proof acceptance is a social process that mathematician’s participate in and the definition of an acceptable proof can, and has, change over time.

We now consider the concept of proof in mathematical logic. There are various misconceptions propagated in the mathematics education literature about the logicians view of proof (e.g. Weber, 2003). For instance, Segal (2000) writes: [T]here is another aspect in which mathematical and logical proofs differ, that of completeness. With logical proofs, the information required, the axioms, laws of inference, are all explicitly set down. Such is not true of mathematical proofs. Also, Hanna (1990) defines formal proof as a “theoretical concept in formal logic (or metalogic), which may be thought of as the ideal which actual mathematical practice only approximates” (pg. 6).
Although formalism in mathematical proof can be a beneficial tool for exploring mathematics, as was Hilbert’s exposition of Euclid’s postulates for creation of new types of geometries (Renz, 1981), it is not the case that proofs published in mathematical logic journals are fully formalized. Gowers (2008) writes that the concept of a formalized proof in logic is “a considerable idealization of what actually appears in a normal mathematical paper under the heading ‘Proof’” (pg. 74). He goes on to say:

“[A] purely formal proof would be very long and almost impossible to read. And yet, the fact that arguments can in principle be formalized provides a very valuable underpinning for the edifice of mathematics. If a mathematician produces an argument that is strangely unconvincing, then the best way to see whether it is correct is to ask him or her to explain it more formally and in greater detail. This will usually either expose a mistake or make it clearer why the argument works” (pg. 74).

Thus, the misconception lies in stating that logicians prefer that proofs are formalized; they merely prefer that proofs could be formalized. Corry (2008) elaborates further:

“[A]ctual proofs devised and published by mathematicians since that time are seldom presented as fully formalized texts. They typically present a clearly articulated argument in a language that is precise enough to convince the reader that it could – in principle, and perhaps with straightforward (if sustained) effort – be turned into one.” (pg. 140)

Taking the above quotes into account, we believe that a good definition of proof in the context of mathematical logic would be the one offered by Tymoczko (1979): that proofs should be convincing, surveyable and formalizable. The characteristics of proofs to be convincing ties back to the social aspect of proof outlined above; an arbitrary reader (with the appropriate knowledge) must be convinced by the argument made in the proof. The definition that Ty-
moczko gives about they surveyability of a proof is that a proof must be able to be “looked over, reviewed, verified by a rational agent” (pg. 59). In this definition, the first two aspects of a proof, the ability to convince and the ability to be reviewed, are similar to the definition in the context of pure mathematics. It is the final component of this definition that couches it in mathematical logic; that a proof must have the ability to be formalizable. Again, this does not imply that the proof must be written formally, just that it would be possible to “find an appropriate formal language and theory in which the informal proof can be embedded and ‘filled out’ into a rigorous formal proof” (pg. 60).

We now systematically review the definition of proof in a mathematics education context. Balacheff (2002) notes that definitions of proof varies widely among mathematics education researchers. Hanna (1990) argues that proof should be less formal in a pedagogical context and emphasize understanding over rigor. For example, she cites Leron’s (1983) proposed restructuring of presented proofs in order to make explicit what is being assumed and where special constructs in the proof come from. Leron gives as an example the traditional proof of the statement: There are infinitely many primes of the form $4k + 3$. This proof is done by contradiction so it is assumed that $p_1, p_2, \ldots, p_n$ represent all the primes. Then, the integer $M = 4p_1 \cdot p_2 \cdots p_n - 1$ is formed without any explanation as to why this integer would be useful in the proof. A proof for teaching may contain a sidebar that justifies the creation of this number and foreshadows its utility in the rest of the proof. Also, bringing the social aspect of proof acceptance into the classroom could also be effective when teaching proof. Alibert and Thomas (1991) write about the a teaching experiment using scientific debate when testing the
truth of conjectures in the classroom, where proofs are considered valid by consensus of the class.

DeVilliers (2012) cautions that defining proofs as having purpose only for verification can obscure the other functions of proof that are perhaps more fruitful in a pedagogical context. In fact, he proposes a simpler definition: “Instead of defining proof in terms of its verification function (or any other function for that matter), it is suggested that proof should rather be defined simply as a deductive or logical argument that shows how a particular result can be derived from other proven or assumed results; nothing more, nothing less” (pg. 7). Thus, we aim to settle on a definition of proof for this study that is broad enough to include the many roles that proof plays in mathematics\(^1\), yet simple enough to apply to the types of student arguments that are typically seen in an introduction to proof course. As discussed above, there are many notions of what proof could be and should be in different contexts of mathematics. In the context of this study, we adopt our modified version of Hanna’s notion of acceptable proof, as this is the type of proof that the undergraduate students will encounter in their courses and their textbooks and will be required in their own proof constructions.

### 2.1.2 Functions of Proof

As we will now argue, proof can fulfill many different roles in mathematics beyond what is considered its primary function, that of verification. The role of proof as useful for verification of mathematical statements has been well-documented (e.g. Harel and Sowder, 1998; Almeida,

\(^1\)An overview of these functions of proof is given in the next section, 2.1.2.
1995; deVilliers, 2012; Hanna, 1990; Mason et al., 1982). For instance, Harel (2007) identifies two distinct processes in proving: the process of ascertaining, in which one convinces him or herself about the truth of a statement, and the process of persuading, in which one convinces others of the truth of a statement. This reflects a common adage, taken from Mason et al. (1982), that one should develop an argument by first convincing yourself, then by convincing a friend and finally by convincing an enemy. DeVilliers (2012) notes that “traditionally, the verification (justification or conviction) of the validity of conjectures has been seen as virtually the only function or purpose of proof” (pg. 1) and “even the majority of research conducted in the area of proof has been done from this perspective” (pg. 1). However, besides verification, mathematics education researchers have identified many other functions of proof in the practice of mathematics. This section will discuss eight additional functions of proof from the literature that will be used as a framework for analysis of student assessment data regarding proof function in this study: proof for explanation, discovery, intellectual challenge, axiomatizing, communication, justification for a definition, illustrating techniques and providing autonomy.

Hanna (1990) and deVilliers (1990) demonstrate that some proofs can be explanatory and provide insight as to why a certain statement is true. Although Steiner adds that the explanation function of a proof is dependent on the part of the proof that we are trying to explain: “An explanatory proof makes reference to a characterizing property of an entity or structure mentioned in the theorem, such that from the proof it is evident that the result depends on the property” (as quoted in Mancosu, 2008, pg. 143).
**A proof that proves.**
Proof by induction:
For $n = 1$ the theorem is true.
Assume it is true for an arbitrary $k$.
Then consider:
$$S(k + 1) = S(k) + (k + 1) = \frac{k(k + 1)}{2} + (k + 1) = \frac{(k + 1)(k + 2)}{2}$$
Thus, the statement is true for $k + 1$ if it is true for $k$.
By induction, the statement is true for all $n$.

Figure 1. A proof of “The sum of the first $n$ positive integers, $S(n)$, is equal to $\frac{n(n + 1)}{2}$” that proves (Hanna, 1990, pg. 10).

For example, Hanna (1990) gives the induction proof shown in Figure 1 of the statement “The sum of the first $n$ positive integers, $S(n)$, is equal to $\frac{n(n + 1)}{2}$” as a proof that proves.

This proof does give insight into why this formula works for all natural numbers $n$, but gives no clarification about where the formula $\frac{n(n + 1)}{2}$ is obtained from. On the other hand, Hanna gives the proof in Figure 2 as one that proves the statement and also gives an explanation of how

**A proof that explains.**
Let $S(n) = 1 + 2 + \cdots + n$
Then, $S(n) = n + (n - 1) + \cdots + 1$

\[\begin{align*}
\text{So,} & \quad 2S(n) = (n + 1) + (n + 1) + \cdots + (n + 1) = n(n + 1) \\
\text{Thus,} & \quad S(n) = \frac{n(n + 1)}{2}
\end{align*}\]

Figure 2. A proof of “The sum of the first $n$ positive integers, $S(n)$, is equal to $\frac{n(n + 1)}{2}$” that explains (Hanna, 1990, pg. 10).
this formula was obtained, which she terms a *proof that explains*. However, the proof in Figure 2
does nothing to show that this should work for all natural numbers. Thus, the explanatory
nature of the proof is determined by the structure that is focussed on in the statement of the theorem.

A third proof is now given that is claimed to be explanatory of both the reason that this particular formula is chosen and that the formula is valid for all natural numbers (J. Baldwin,
personal communication, September 30, 2012). This proof, shown in Figure 3, uses properties about series to produce a result that shows how the formula is obtained (in that, it is actually solved for as part of the proof) as well as gives an explanation\(^1\) of why this result should remain valid for all natural numbers.

These explanatory proofs are often seen by researchers as more valuable for educational purposes. Hanna (1990) writes that “as mathematics educators it is our mission to make students understand mathematics. It is my contention that in support of this mission we should give a more prominent place in the mathematics curriculum to proofs that explain” (pg. 12) and deVilliers (1990) adds that “the more fundamental function of explanation should be exploited to present proof as a meaningful activity to pupils” (pg. 23).

DeVilliers (1990, 2002) also provides evidence for four more functions of proof in mathematics. The first is the use of proof as a means of discovery since proof can sometimes lead to

\(^1\)Although an explanation of why this formula holds for all \(n\) is given in this proof, the proof relies implicitly on the generalized distribution and associativity properties of the real numbers, which must be proven by induction. Thus, this proof is explanatory, however it is not the most rigorous proof.
Another proof that explains.
Consider the telescoping series \( \sum_{k=1}^{n} [(k+1)^2 - k^2] \).
Notice first that \([ (k+1)^2 - k^2] = 2k + 1 \) and so the summand can be rewritten. Notice also that
\[
\sum_{k=1}^{n} [(k+1)^2 - k^2] = (2^2 - 1^2) + (3^2 - 2^2) + \ldots + ((n^2 - (n-1)^2) + ((n+1)^2 - n^2) = (n+1)^2 - 1.
\]
Thus, we have \((n+1)^2 - 1 = \sum_{k=1}^{n} [2k + 1] = 2 \sum_{k=1}^{n} k + \sum_{k=1}^{n} 1 = 2 \sum_{k=1}^{n} k + n \).
This can be simplified to \(n^2 + 2n = 2 \sum_{k=1}^{n} k + n \) and solving for the sum we obtain \( \sum_{k=1}^{n} k = \frac{n(n+1)}{2} \).

Figure 3. A proof of “The sum of the first \( n \) positive integers, \( S(n) \), is equal to \( \frac{n(n+1)}{2} \)” that explains how the formula is obtained and why it works for all natural numbers (J. Baldwin, personal communication, September 30, 2012).

new results in a field. An example of this is when proofs are analyzed and a general case conjectured. Indeed, Lakatos (1976) shows that the method of proof analysis, where, upon emergence of a counterexample, proofs and definitions are reanalyzed, is exactly the way that mathematics progressed throughout history and how new fields of mathematics opened up. Aristotle also appreciated the role that proof plays in creating new knowledge and he elucidated a type of proof, that he called demonstrative logic, that was not proof for persuasion, but for creation of new knowledge. He defined proof as “starting with premises known to be true and a conclusion not known to be true, the knower demonstrates the conclusion by deducing it from the premises—thereby acquiring knowledge of the conclusion” (Corcoran, 2009, pg.1).
Also, proof can function as an intellectual challenge to the prover. The completion of a proof can be very satisfying for mathematicians and analogous to solving a puzzle. This view is that of proof for proof’s sake and many mathematicians espouse a delight in the activity of proving. Third, proof can expose logical relationships between statements and can serve as a tool for axiomatizing results in a mathematical system. If another statement is used in a proof, we can consider that an axiom or lemma needed for that result and, comparatively, if a further statement is the consequence of that proof we consider it a corollary. This provides a way to build up the hierarchy of statements in mathematics. DeVilliers (1990) gives the example: “The primary function of a proof for the intermediate value theorem for continuous functions is purely that of systemization, as a simple picture combined with an informal argument is sufficient for the purposes of both verification and illumination” (pg. 21).

Moreover, proofs serve an important communicative role, since they are the main way mathematical knowledge is transmitted. Rav (1999) proposes that proof are the main bearers of mathematical knowledge, not only between mathematicians, but also between mathematics teachers and their students. Hanna and Barbeau (2011) write that “educators have long recognized the explanatory value of many proofs, but they have had in mind primarily the light such explanatory proofs can shed on the mathematical subject matter with which they deal... proofs can also be bearers of mathematical knowledge in the classroom in another sense, the sense proposed by Rav: that proofs have the potential to convey to students ‘methods, tools, strategies and concepts for solving problems’” (pg. 85).
Furthermore, Weber (2002) claims that proof can serve as the justification for a definition in mathematics. For instance, the author cites the use of Peano arithmetic axioms to prove simple statements such as $2 + 2 = 4$. Clearly, this statement would not be proved for verification purposes, but instead to determine if the axioms of Peano arithmetic are reasonable ones. In the context of teaching and learning, Weber (2002) also comments that proof can illustrate techniques in advanced mathematics. The author gives the example of instructors asking students to prove that $f(x) = x^2$ is a continuous function. Most students have an intuitive understanding of why this statement should be true, however the construction of a proof this statement can demonstrate certain techniques for proving statements about continuous functions in general. A proof of this statement can also clarify issues about how to make the concept of continuity precise. Yackel and Cobb (1996) comment that another role of proof can be to provide autonomy to students by allowing them to verify statements for themselves and create their own mathematical knowledge.

While it is important to enumerate all these functions of proof, we strongly agree with the statement by Weber (2002) that “some proofs can fall into multiple categories; others might fall into none.” He gives the example of Euclid’s proof of the infinitude of primes. The proof of this traditionally presented to introduction to proof students can serve to illustrate the technique of proof by contradiction, convince the student that this statement is indeed true and explain why the statement is true simultaneously. De Villiers (1990) also believes that the “although the...functions of proof above can be distinguished from one another, they are often intricately linked in specific cases” (pg. 23).
It is the nine functions of proof presented here (i.e. verification, explanation, discovery, intellectual challenge, axiomatizing, communication, justification of a definition, that we wish determine student understanding and appreciation of in this study. Therefore, we use these functions as a framework for the data analysis in this study. In Chapter 3, studies about what is currently known about students’ understanding of some of these proof functions are reviewed. This study builds on studies of a similar nature that investigate several of these proof functions, but we expand our scope to include all nine of functions of proof presented above.

2.1.3 Proof Schemes

Students of mathematics often have a different way of approaching proof than experts in the domain (Weber, 2001; Dreyfus, 1999). This may be because students gain conviction about a statement from other aspects of a proof than mathematicians do. For instance, a student may verify several cases of a statement and consider it a proof. Others may accept the truth of a statement because it appears in the textbook or it is written with symbols and ‘looks mathematical’. This idea of the varying ways that students gain conviction is termed students’ proof schemes by Harel and Sowder (1998). The authors refer to the characteristics of a proof that convince a person as that person’s proof scheme and they define it as “what constitutes ascertaining and persuading for that person” (pg. 244).

Regarding these differing modes of gaining conviction, Harel and Sowder (1998) identified three broad categories, each with several subcategories, of proof schemes that students may
employ: external conviction, empirical, and analytical\(^1\). An *external conviction* proof scheme is characterized by gaining conviction from superficial aspects of a proof, while an *empirical* proof scheme is one where “conjectures are validated, impugned, or subverted by appeals to physical facts or sensory experiences” (pg. 252). *Analytical* proof schemes are where statements are validated by logical deductions, however not necessarily in the way that constitutes mathematical proof. This section will elaborate on each of these categories, which will be used as a framework for data analysis of interviews asking students to elaborate on their thinking on proof validation tasks. The organization of the proof schemes presented here is shown in Figure 4.

An *external conviction* proof scheme is characterized by obtaining conviction about the proof of a statement from an outside source. This category is further divided into three subcategories: *ritual*, *authoritarian* and *symbolic* proof schemes. A *ritual* proof scheme is when students accept a proof based on appearance alone. For example, a student may be convinced that a proof is valid because it looks ‘mathematical’ or contains symbolic reasoning. The *authoritarian* proof scheme is when students accept a proof based solely on its presentation by an authority, such as a teacher or a textbook. In other words, the student is convinced by the fact that the source of the proof is reliable and not from any logical aspects about the proof itself.

\(^1\)These proof scheme classifications are revisited in Harel (2007). The main changes to the hierarchy of proof schemes in this paper are in the names of the categories, thus we feel it is appropriate to use the categories from Harel and Sowder (1998).
where arguments originate from definitions and axioms. Additionally, the transformational proof scheme is subdivided into internalized, where the student uses a proof heuristic; interiorized, where the student has reflected upon an internalized proof scheme, making it explicit; and restrictive, where the student has assumed certain restrictions on the statement. Depending on the restriction made, the restrictive category is subdivided into contextual, where the conjecture is proved only in a specific context; generic, where conjectures are interpreted generally, but proved in a specific context; and constructive, where students are convinced only by construction of a mathematical object (and not by contradiction). The axiomatic proof scheme is further divided into intuitive; axiomatic, where only axioms that correspond to the student's own intuition are used; structural, where the student understands that axioms describe a mathematical structure or system and the theorems are proven within this system; and axiomatizing, where the student is able to vary the axioms used and explore the outcomes.

The hierarchies of these proof schemes are shown in Figure 1. According to Harel and Sowder (1998), the goal of instruction on proof should be to move a student from an empirical or external conviction proof scheme into an axiomatic proof scheme and the highest level of understanding corresponds to the axiomatizing proof scheme, where the student understands the "big picture" and how statements fit together.

Figure 4. The organization of students’ proof schemes (Harel and Sowder, 1998).
The *symbolic* proof scheme is when symbols are used in a proof without retaining their meaning. For this proof scheme, Harel and Sowder (1998) give the example of a student constructing a linear algebra proof and letting $A$ represent a matrix, but later dividing by $A$ in an equation. In this example, the student realizes that some symbolic logic must be used, but does not attend to the fact that properties of matrices must be adhered to.

The second category of proof schemes, *empirical* proof schemes, is one where statements are verified by considering specific cases or representations of mathematical objects. This category has two subcategories: *inductive* and *perceptual*. *Inductive* proof schemes are characterized by evaluating several specific cases of a statement. This refers to when students are convinced by noticing a pattern when checking several cases and then assume that the pattern must continue. The *perceptual* proof scheme is when a specific picture or mental image is used, but the results of transformations of those images is not attended to. An example of this proof scheme would be if a student was proving a statement about all quadrilaterals, yet failed to consider the consequences of having a convex quadrilateral.

*Analytical* proof schemes are characterized by recognizing that statements must be proven using some type of logical deduction. This proof scheme has two subcategories, *transformational* and *axiomatic*, each with several more subcategories. Harel and Sowder (1998) write that the *transformational* proof scheme is “characterized by (a) consideration of the generality aspects of the conjecture, (b) application of mental operations that are goal oriented and anticipatory, and (c) transformations of images as part of a deduction process” (pg. 261). This could occur as an *internalized*, *interiorized* or *restrictive* proof scheme. An internalized proof scheme is one
where a student uses or observes a proof heuristic. The authors give an example of a linear algebra student using row reduction on a matrix in a proof because the student has seen this tool used frequently in other proofs. Another form of the transformational proof scheme is an interiorized proof, which is an internalized proof scheme that a student has reflected upon to make it explicit. An example of this is mathematical induction; often students, after learning induction, become aware of the benefits and limitations of this method and look for cues in the problem statement as to when it would beneficial to use in a proof.

The last type of transformational proof scheme is called a restrictive proof schemes and occurs when a student has assumed certain restrictions on the statement. The type of restriction that a student imposes in a restrictive proof scheme further subdivides this category. A contextual restriction would relate to only focusing on a statement within a certain context. For example, proving the general statement “$n + 1$ vectors in an $n$-dimensional vector space are linearly dependent” in the specific context of the vector space $\mathbb{R}^n$. A contextual restriction could also be of a spatial nature, where “students are able to think of the problem situation only in terms of their imaginative space” (pg. 268), in which case it is considered a spatial proof scheme. Also, a generic restriction could be made by a student in a proof, which would be when the student realizes that a statement must be proven in general terms, but still proves it in a specific context. Additionally, a student may need to actually construct certain mathematical objects to gain conviction of their existence, rather than using a particular existence theorem. This would then be considered a constructive proof scheme.
The final category of proof schemes are *axiomatic* proof schemes, of which there are three types: *intuitive-axiomatic*, *structural* and *axiomatizing*. If a student has an axiomatic proof scheme, he or she appreciates that mathematical arguments must originate from definitions and axioms. However, if a student will only attend to axioms that correspond to the student’s own intuition, this is defined as an *intuitive-axiomatic* proof scheme. A *structural* proof scheme is when a student understands that axioms describe a mathematical structure or system and the theorems are proven within this system. The highest level of understanding according to Harel and Sowder (1998) occurs when a student is able to vary the axioms used and explore the outcomes, which they call the *axiomatizing* proof scheme.

These are all classifications of what convinces a student, but it is important to note that the given person’s proof schemes is not static; a student may translate from one proof scheme to another over time or for different types of mathematical statements. Additionally, the organization of proof schemes shown in Figure 4 should not be interpreted as a strict hierarchy. Although the transformational proof scheme is at a higher level than external conviction or empirical proof schemes, several of the other proof schemes can occur in relation to each other. For example, the interiorized proof scheme can happen only after a student has created and reflected upon an internalized proof scheme. Sowder and Harel (2003) write that:

“...
are sometimes incorrect and that a generalization supported by even a large number of examples cannot be trusted unequivocally.”

Thus, while external conviction and inductive proof schemes may be employed when exploring an argument, a proof that is considered a final product must appeal to one’s axiomatic proof scheme.

In this study, information was desired about the proof schemes that students begin and end the semester with and if there is a pronounced difference in the proof schemes held by students in the treatment and comparison courses. We use this idea of proof schemes as a framework for data analysis of assessments and student interviews to gain insight about any impacts collaborative revision may have on students’ proof validation skills.

2.2 Summary of Proof Frameworks

The previous sections lay the groundwork for the study design and analysis of data collected to measure students’ proving skills. First, we rationalized why we adopt in this study a modified version of Hanna’s definition of acceptable proof. An appropriate definition is essential when later we approach the idea of proof validity. Second, a summary was given proof functions from the literature reviewed above that we will address in this study are: verification, explanation, discovery, communication, axiomatization, intellectual challenge, justification for a definition, illustration of techniques and providing autonomy. These functions of proof provide a framework for analysis of data on students’ appreciation of proof function.
2.3 Collaborative Learning

The teaching experiment designed and implemented in this study draws upon the sociocultural perspective. This perspective, pioneered by Vygotsky (Goos et al., 2002), focuses on the role that participation and social interactions play in psychological development. A number of studies (Goos, 2004; Weber et al., 2008; Sfard, 2001) have brought this perspective into classroom learning and examined how students’ social interactions can impact students’ learning about various topics. In the classroom, the sociocultural perspective usually manifests as collaborative learning and Scott and Palinscar (2009) writes:

“Informed by a sociocultural perspective, learning is thought to occur through interaction, negotiation, and collaboration. While these features are characteristic of ‘cooperative learning,’ what sets instruction that is informed by sociocultural theory apart is that there is also attention to the discourse, norms, and practices associated with particular discourse and practice communities” (pg. 854).

This study also uses collaborative learning and we concentrate specifically on how collaboration between students can impact the learning of proof at the level of an undergraduate introduction to proof course. We borrow the definition of collaborative learning presented by Rasmussen and Stephen (2008) as “the collective activity of a mathematics class refers to the normative ways of reasoning that develop as learners solve problems, explain their thinking, represent their ideas, etc.” (pg. 195).

A natural extension of sociocultural theory into the classroom is in small group interactions. There are several studies that explore small group interactions. For instance, Goos (2004) writes that “from an educational perspective, there is learning potential in peer groups where students have incomplete but relatively equal expertise, each partner possessing some knowledge and skill
but requiring others contribution in order to make progress” (pg. 263). Weber et al. (2008) augment this argument when they claim

“group discussion can also facilitate learning by inviting students to be explicit both about the ways in which they make new claims from previously established facts and about the standards they are using in deciding whether an argument is acceptable. Challenges from classmates can encourage students to debate whether a particular method of argumentation is appropriate and provide students with the opportunity either to justify their methods when their reasoning is sound or revise or abandon their methods when their reasoning is flawed” (pg. 248).

It has been hypothesized that the learning attributed to small group interactions is due to the communication and explanations involved in this activity. Noss and Hoyles (2006) write: “important progress in our understanding of how the community of practice paradigm works has come from studies of the way the scientific community produces knowledge. These studies have shown that, even more than individual or team efforts, the open process of publishing, critiquing, exchanging, and debating ideas significantly contributes to the growth of science” (pg. 390). Sfard (2001) goes even further and says “that communication should be viewed not as a mere aid to thinking, but as almost tantamount to the thinking itself” (pg. 13).

Collaborative learning may be difficult to enact, however, as students may be unfamiliar with the new norms of the classroom, which now requires participation and communication, due to overexposure to the traditional classroom where discussion and interaction may be discouraged. Weissglass (1993) writes: “One obstacle when college students begin to work in groups is their lack of experience communicating about mathematics. It is important therefore that early group activities develop communication skills rather than stress solutions or proof” (pg. 664). A study by Smith et al. (2009) found that the instructor of a collaborative inquiry-
based number theory course had to intervene in differing ways throughout the course. At the beginning of the course, the students were unsure what was expected of them and the instructor had to step back from the front of the classroom to get them to participate. During the middle of the course, however, students were more comfortable speaking up so the instructor had to direct the conversation more and help students struggling with the harder content. At the end of the course, the instructor spoke only when he determined students were stuck and then only to suggest ways to proceed. Thus, collaborative learning often requires familiarizing students with new standards of what is expected of them in the classroom.

Drawing upon this sociocultural perspective, the design of this study uses small group interactions to promote mathematical communication amongst students and to help students draw on each other’s expertise to create meaning and to aid them with proving skills. Cuoco et al. (1996) states, and we agree, that “part of students experience should be in a classroom culture in which they work in collaboration with each other and in which they feel free to ask questions of each other and to comment on each others work” (pg. 379).

2.4 Conclusion

This chapter presented a framework for conducting an exploration into the extent that collaborative revision may impact students' proving skills. The meaning of proof for this study, the roles that proof can fulfill in mathematics and the idea of proof schemes were well-defined, which was necessary as a precursor to our explanation of the data analysis and results. Additionally, an overview of collaborative learning theories was given as a rationale for the use of
collective activity in the study design. The next chapter, Chapter 3, will review studies that are related to our research questions.
CHAPTER 3

LITERATURE REVIEW

This study examines the impacts of collaborative revision on students’ proof attitudes and abilities, therefore we begin this section with a discussion of why proof is important for undergraduate students to learn. This chapter gives an outline of the literature emphasizing the integral nature of proof in mathematics. Furthermore, this chapter reviews studies related to students’ understanding of the proof from the literature discussed in the last chapter. Research regarding students’ difficulties with proof validation will be discussed to motivate the study as well as research on the cause of and solutions to these struggles. Proof construction and overview what is already known about student learning in these areas. Finally, research on collaborative learning and proof revision will be summarized to provide a rationale for this design of the teaching experiment performed in this research project.

3.1 Importance of Proof

Rav (1999) proposes, and Hanna and Barbeau (2011) agree, that proofs are of the utmost importance in the field of mathematics, since they, instead of theorems, are the main vehicles in which mathematical knowledge is contained and transferred. Rav (1999) uses the examples of mathematicians’ proof attempts of Goldbach’s conjecture leading to the new field of sieve theory and proof attempts of the Continuum Hypothesis leading to developments in topology and
contributed to the work of Gödel. Krantz (2007) gives a powerful argument for the essentialness of proof to mathematics.

“If one were to remove ‘proof’ from mathematics then all that would remain is a descriptive language. We could examine right triangles, and congruences, and parallel lines and attempt to learn something. We could look at pictures of fractals and make descriptive remarks. We could generate computer printouts and offer witty observations. We could let the computer crank out reams of numerical data and attempt to evaluate those data. We could post beautiful computer graphics and endeavor to assess them. But we would not be doing mathematics. Mathematics is (i) coming up with new ideas and (ii) validating those ideas by way of proof. The timelessness and intrinsic value of the subject come from the methodology, and that methodology is proof” (pg.33).

Thus, it seems that proof is the crux of the pure mathematical discipline.

The ability to construct and evaluate proofs is also important for preparing our mathematics majors and future teachers. The Committee on the Undergraduate Program in Mathematics (CUPM) recommends in their curriculum guide incorporating the ability to “develop mathematical thinking and communication skills” (Committee on the Undergraduate Program of Mathematics, 2004) as one of its six recommendations for all mathematics courses in the undergraduate curriculum. This recommendation included teaching student to “learn to apply precise, logical reasoning to problem solving”, “develop persistence and skill in exploration, conjecture, and generalization”, and “read and communicate mathematics with understanding and clarity”; all of which are skills related to the learning of proof. Moreover, the widely adopted Common Core State Standards identify the ability to “construct viable arguments and critique the reasoning of others” as one of the standards for mathematical practice. This means that pre-service teachers must acquire the ability to aid students in producing and testing conjec-
tures, using and recognizing counterexamples and making plausible arguments and evaluating the reasoning of others (Common Core State Standards Initiative, 2010).

3.2 Students’ Understanding of the Functions of Proof

A presentation of nine functions of proof written about in the mathematics education literature (de Villiers, 1990; Hanna, 1990; Weber, 2002; Yackel and Cobb, 1996) was given in Section 2.1.2. Although many mathematicians and mathematics educators would agree that verification is an important function of proof, we reviewed eight additional functions of proof: explanation, discovery, communication, intellectual challenge, axiomatizing, justification for a definition, illustrating techniques, and providing autonomy. A review of the literature around which of these additional proof functions students do and do not understand is presented in this section.

De Villiers (1990) argues that informing students about these additional functions of proof when teaching can be advantageous and Hanna (1990) believes instructors can help students understand certain concepts by substituting proofs that are explanatory for proofs that are not, whenever possible. Given the teaching implications, it is desirable to determine what functions of proof students do or do not appreciate. Healy and Hoyles (2000) found in a study of 14-15 year old students in the United Kingdom that 50% (N=2459) of the students referred to proof as having a verification function, 35% identified an explanation function and only 1% perceived a discovery or axiomatizing function of proof. Additionally, they found that 28% of the students surveyed “had little or no idea of the meaning of proof and what it was for” (pg. 418).
Another study by Knuth (2002a) interviewed in-service secondary school teachers about their conceptions of proof in mathematics. Results indicate that verification was the primary function of proof identified, with all of the sixteen teacher participants referring to proof for establishing truth of a statement. Twelve of the teachers spoke about the communication function of proof and eight referred to the role of proof in axiomatizing results. Only three of the teachers provided explanation as a function of proof, which indicates that in-service teachers do understand multiple roles of proof, but may not be teaching these functions to their students. Knuth (2002a) writes:

“These roles suggest that teachers have a diverse and, pedagogically speaking, potentially powerful understanding of the function of proof in mathematics. Perhaps if teachers were to pay explicit attention to these roles during their instruction, they would provide classroom experiences with proof that would enable their students to go beyond the limited conceptions of proof that students have traditionally developed. For example, one might question whether high school geometry students are able to view the proofs that they construct in class as interrelated – that is, whether these students are cognizant of the particular axiomatic system (typically Euclidean geometry) that provides the structure for their work. Teachers holding a view of proof as a means of systematization might be more likely to provide opportunities for students to reflect on their work through this particular lens.” (pg. 399-400).

Teachers also referred to other functions of proof beyond those in the literature; namely, proof as a way of developing logical thinking skills and, to a lesser degree, proof as a way of promoting understanding Knuth (2002b). The role of proof in increasing understanding is echoed in Pinto and Tall (1999) when they conclude that exploring the way definitions are used in proofs can help one to develop intuition around a concept.

Almeida (1995) found in a study of second-year undergraduates that students did appreciate proof as explanatory as well as for verification, but that they did not identify proof as a means
of communicating mathematics. Similarly, a study with advanced undergraduate mathematics students by Coe and Ruthven (1994) found that “the most commonly expressed function of proof was to ensure the certainty of what is proved” (pg. 50). However, Coe did find that several students recognized the role of proof in organizing mathematical statements into a hierarchy. Furthermore, in this study several students related a different function of proof, that of understanding better the concepts in mathematics, which is consistent with Knuth (2002b).

These studies agree with the hypotheses of Moore (1994) and Weber (2002) that undergraduate students appreciate the role of proof to explain and convince, but most likely do not understand the other functions of proof outlined in the literature. Weber (2002) argues that this may apply to mathematicians as well when he writes “I suspect that most students and perhaps some mathematicians believe that convincing and explaining are the only purposes of proof” (pg. 15). There is very little research studying undergraduate student perceptions of functions of proof and essentially no research studying any of the functions of proof beyond verification, explanation, communication, axiomatizing and discovery. Thus, a primary goal of this study is to determine which of these roles of proof are understood by undergraduate students in an introduction to proof course, where they will likely be exposed to proof-writing for the first time in their studies, and to determine if students beliefs about proof evolve during this course.

3.3 Proof Validation

The importance of developing proof construction and validation skills in a transition to proof course cannot be overlooked. Students majoring in pure mathematics are expected to
learn not only mathematical content, but also be able to construct proofs involving that content (Selden and Selden, 2003; Weber, 2010). Hanna and Barbeau (2011) make an argument that proofs, not theorems, are the main transmitters of mathematical knowledge and thus it is imperative that anyone with a degree in mathematics be able to construct and evaluate proofs. Additionally, Selden and Selden (2003) add that it is important for pre-service teachers to be able to evaluate the validity of arguments, since the increased emphasis on proof in school mathematics will make this a necessary skill when evaluating their students’ arguments.

However, it is clear from the research that students at all levels have trouble determining if a given argument is valid. Selden and Selden (2003) found that undergraduate students in an introduction to proof course had trouble differentiating between a valid and an invalid proof. In a proof verification task, undergraduate students in a transition to proof course initially judged proofs correctly less than half the time. However, when prompted for reflection about the proofs by the interviewer, the students were able to correctly evaluate proofs over 80% correctly. The authors conjecture that this may be due to the fact that proof verification skills are not explicitly taught to students and the vast majority of proofs seen in this type of course are valid, either from the instructor or the textbook. Furthermore, the texts for transition to proof courses include very few proof validation tasks and those that do exist are specifically designed to only have a single error to be detected.

Alcock and Weber (2005) found similar results to Selden and Selden (2003) with students in an undergraduate real analysis course; that students were unable to assess proofs correctly unless prompting was involved on the part of the interviewer. Another study by Brandt and
Rimmasch (2012) showed a group of undergraduate students in a transition to proof course video presentations of proofs and asked them first to evaluate the validity of proofs individually and then participate in a small group discussion with an interviewer facilitator. After the group discussion students and were asked to reassess their judgement. This study found that the group discussion had a positive impact on the students ability to identify correct proofs and confidence about their assessment. Weber (2010) found that although students who had completed a transition to proof course did not find empirical arguments (arguments where only a few examples are given) convincing, students “exhibited difficulty in detecting flaws in deductive arguments and were prone to accepting flawed deductive arguments as convincing valid proofs” (pg 330).

The trouble that students have with proof validation is attributed to a number of misunderstandings. For instance, students may hold varying conceptions of what constitutes a convincing argument and, thus, may be convinced by an argument that a mathematician would not consider a rigorous proof (Harel and Sowder, 1998). Recio and Godino (2001) add that students may not be familiar yet with the type of formal proof required in mathematics and so they draw on ideas about proof from the context of daily life or from the empirical sciences, both of which require a lesser degree of formalism. Another possibility for poor proof validation skills is that students attend to different aspects of proof than mathematicians would. Selden and Selden (2003) found that students often focused on surface features of a proof, such as what symbols and formulas were used in the proof. Inglis and Alcock (2012) found similar results and noticed that when compared to mathematicians readings of proofs, students “spend pro-
portionately more time focusing on surface features of arguments, suggesting that they attend
less to logical structure” (pg. 358). A study by Weber (2010) shows an additional incongruity
in students’ proof validation skills is that undergraduate students may find a written proof to
be convincing, but not consider it a valid proof. As an example, several students in Weber’s
study did not find a proof containing a graph to be valid, yet they were convinced by the proof.
Healy and Hoyles (2000) found similar results with 14-15 year olds, where students often chose a
more algebraic or symbolic proof as the one they thought would receive the best grade, whereas
they chose a more explanatory proof written in plain words as the proof they found most con-
vincing. Another study by Segal (2000) found that first and second year university students
tended to see the ability to supply conviction and the ability to prove as distinct attributes of
an argument.

Several studies give suggestions as to how instructors can increase students’ proof validation
skills. Alcock and Weber (2005) write that the results of their study, as well as results from
proof validation studies of Selden and Selden (2003) and Weber (2010), “suggests that the
ability to validate proofs may be in many students zone of proximal development and that
students abilities in this regard might improve substantially with relatively little instruction”
(pg. 131). That is, students can benefit by making their own evaluations of the validity of
proofs and then discussing these evaluations with an instructor or with other students. This is
echoed in Pfeiffer (2009) when she argues that “practice of proof validation can not only improve
students’ validation skills but can also lead them to a better understanding of mathematical
content and to improved appreciation of deductive reasoning” (pg. 84). Thus, a goal of this
study is to expose students to proofs written by students at the same level, which may have varying degrees of validity, as a way to teach validation skills.

3.4 Proof Construction

Even though proof construction is an integral part of mathematics, it is well documented (e.g. Harel and Sowder, 1998; Weber, 2002; Moore, 1994) that students often do not leave an introduction to proof course with strong proof construction skills. Pinto and Tall (1999) write that “the transition from elementary mathematics to formal proof is a huge chasm for many students whose underlying concept image is unable to sustain the formalism” (pg. 7). Moore (1994) conducted a study to identify specifically what difficulties students were having with proof construction. This study was of 16 students in an introduction to proof course and found seven main obstacles students encountered when writing proofs:

D1. The students did not know the definitions, that is, they were unable to state the definitions.
D2. The students had little intuitive understanding of the concepts.
D3. The students concept images were inadequate for doing the proofs.
D4. The students were unable, or unwilling, to generate and use their own examples.
D5. The students did not know how to use definitions to obtain the overall structure of proofs.
D6. The students were unable to understand and use mathematical language and notation.
D7. The students did not know how to begin proofs (pg. 251-252).

Moore (1994) also writes that “the students’ perceptions of mathematics and proof influenced their proof-writing performance and were sometimes a hindrance to their success” (pg. 252).

The problems students have when constructing proofs are addressed more generally by Weber (2001), who hypothesizes that students’ difficulty writing proofs can sometimes be at-
tributed to the lack of a specific type of understanding, which he terms *strategic knowledge*. That is, even if a student understands what constitutes a proof and is aware of the appropriate theorems and definitions that can be used, he or she may still not be able to construct a proof because they are unable to narrow down the number of inferences that can be drawn from the ones that are relevant. Weber notes that this type of knowledge is possessed by domain experts, such as graduate students and professors, but usually not by undergraduates. Raman (2003) expands upon Weber’s determination of the types of knowledge that separate expert from novice proof writers. She refers to three types of ideas used in proof production: a *heuristic idea*, which is one rooted in informal understandings that may give insight but not necessarily pave the way toward a formal proof; a *procedural idea*, which is a formal manipulation which may lead to conviction, but does not offer understanding; and a *key idea*, which is a heuristic idea that can be translated into a rigorous proof. Raman claims that mathematicians tend to think in mainly key ideas, but do not tend to teach these to their students and that emphasis of key ideas when teaching students could be an important step in helping students understand proof.

Fischbein (1982) proposes that an explanation for students’ poor proving skills could be that formal proof is at odds with students’ intuition. He writes: “A formal proof offers an absolute guarantee to a mathematical statement. Even a single check is superfluous. This way of thinking, knowing and proving, basically contradicts the practical adaptive way of knowing which is permanently in search of additional confirmation” (pg. 17). He claims that rigorous proof is not natural and thus students struggle when faced with writing such a proof. Another explanation is that students may not even understand what constitutes a mathematical proof.
Recalling the discussion of proof schemes from Section 2.1.3, Harel and Sowder (1998) intimate that the goal of instruction on proof should be to move students towards axiomatic proof schemes and ultimately towards an axiomatizing proof scheme, where the student understands the “big picture” and how statements fit together.

Moreover, students may see the proof construction as pedantic or as something required only by an instructor for a class. Thus, involving students more in the real work of a mathematician has been a way suggested by the literature to facilitate student learning of proof. This includes the use of conjecture in the classroom and the retooling of proofs based on comments from colleagues. Alibert and Thomas (1991) argue that while formulating conjectures are essential to the work of a mathematician, students are often taught about proofs in a manner that obscures the need for conjecture. The authors describe a teaching experiment where students are presented with a mathematical situation and then asked to make conjectures about the situation, which are then decided as true or false by consensus of the class. Martinez and Li (2010) agree that conjecturing is an important part of the proving process and depriving students of the opportunity to create conjectures and determine their truth one to prove hinders their ability to construct a proof. Their work was with 9th and 10th grade classrooms where students were given a problem situation and asked to make conjectures. During the conjecturing process, students identified competing conjectures, dismissed some conjectures due to new evidence and identified a conjecture to prove. The study found that participation in the conjecturing process used knowledge that was then used in proving the conjecture, which suggests that these processes are highly related.
3.5 Collaborative Learning & Revision

As outlined in Section 2.3, this study utilizes sociocultural theory to rationalize the inclusion of small group interactions into the design of the teaching experiment. This is done to exploit the power of communication between students in assisting with learning and understanding. Additionally, the CUPM recommends that students in all courses in the mathematical sciences be able to “read and communicate mathematics with understanding and clarity.” (Committee on the Undergraduate Program of Mathematics, 2004, pg. 11). The detailed recommendation specifically identifies classroom discussion and group work as beneficial for all students.

“Our classroom discussion and informal oral presentations are important (but often overlooked) ways to help students improve thinking and problem-solving skills. For example, students can write homework solutions or partial solutions on the board each day at the beginning of class. Lively class discussions can arise from the students board work, and they can set the tone for the rest of the class period...Group work can further encourage students to verbalize mathematics both in and out of the classroom.” (pg. 16).

Several studies (Selden and Selden, 2009; Smith et al., 2009) have explored collaborative learning and small group interaction and were able to find enhancements to student learning. Goos et al. (2002) found that collaboration amongst secondary students promoted metacognition. This study analyzed student conversations around mathematical tasks and the results indicate that when students make their thinking explicit to other students it can contribute to “productive metacognitive decisions” (pg. 219) which aid students in solving challenging problems. Bonsangue (1994) found that small group collaboration resulted in significantly improved test scores and course grades in an undergraduate elementary statistics course compared to a control group. He conjectures that “the connection that the students in the cooperative
group made amongst themselves and their colleagues and the instructor may have increased their commitment to the course and their mathematical self-confidence” (pg. 115).

Inquiry-Based learning (IBL) has also been suggested as a way to encourage collaborative learning and sense making among students (Smith et al., 2009; Dhaher, 2007). IBL is an alternative way to set up the classroom, where the role of the teacher is merely that of mediator and group discussion uses up the majority of class time. Smith et al. (2009) used IBL in an undergraduate number theory course and found that, when compared to a control classroom, IBL students tended to emphasize sense-making when presented with proof tasks, whereas the control group students tended to be concerned with finding shortcuts to complete the tasks. This is similar to Selden and Selden’s (2009) suggestion that an effective way to teach proofs is to provide students with statements to prove with little or no explanation and allow students to provide the appropriate definitions, examples and proofs.

Allowing students a chance to revise and resubmit proofs may also have benefits. A study by Strickland and Rand (2012) allowed students to submit multiple revisions of proofs in response to teacher feedback and measured the impacts on student learning. The teacher comments given were minimal, often just circling a confusing or incorrect passage of the proof, and students were allowed as many revisions as needed. Although the data set was small, on average, students in the revision group did better on the final exam. However, we are not aware of other studies that consider the impacts of revision on undergraduate students’ learning about proofs. Since revision and reassessment of proofs in a natural part of the mathematician’s practice,
the teaching experiment in this study turns the classroom environment into a more authentic picture of mathematics.

We close this section with a quote from Selden and Selden (2003): “[W]e suggest that additional practice in validating such actual student ‘proofs’ as those in this study, together with small-group discussion could be beneficial, especially for preservice secondary teachers who one day may need to judge the correctness of their own students’ proofs or novel solutions to problems” (pg. 28–29). This is precisely what this study aims to do, with students evaluating each other’s proofs for validity and discussing and critiquing the proofs in small groups.

3.6 Summary

Proof is an integral part of the mathematical discipline and therefore we should expect students of mathematics to learn about and embrace proof throughout their undergraduate experience. However, the studies reviewed above show that this may not be the case and that many mathematics students struggle with proving skills. While students do appreciate proof as useful for gaining conviction and explanation, the studies outlined in this chapter show that students understand little about proof function beyond this. Students also have trouble identifying if a given argument is a valid one. However, there is evidence that discussion, either in a group or with an interviewer or instructor, may help students when evaluating arguments. Proof construction is also a difficulty for students and may due to misconceptions about what constitutes a proof or may be at odds with students’ natural intuitions about arguments.

There are several ways that have been suggested in the literature about how to remedy the student difficulties presented here. This study uses collaborative learning as a way to overcome
struggles students have with proof construction and validation. Proof validation studies (Selden and Selden, 2003; Alcock and Weber, 2005) show that reflection about written proofs can aid in correctly evaluating the validity of a given proof and Strickland and Rand (2012) showed that revision can aid in student learning. Thus, collaborative revision is a way to explore the benefits of combining these proven techniques and this study examines the impact on students validation skills when using a collaborative revision teaching intervention in an introduction to proof classroom.
CHAPTER 4

METHODOLOGY & DATA COLLECTION

This research investigates undergraduate student thinking regarding proof when subjected to a teaching experiment that was designed and implemented in a transition to proof course for undergraduate students. This teaching intervention employed collaborative revision in the classroom, which refers to the process in which students explain a proof they have written to their classmates and the other students are encouraged to make comments and critiques to ensure that the proof is valid. In order to properly document this environment and measure the impacts on student learning, qualitative and quantitative analyses were performed on the collected data. Students participating in this teaching intervention were contrasted with students in a comparison course to determine the effects of collaborative revision on proof construction and validation skills.

The research questions were answered using a mixed method analysis, with qualitative and quantitative components, of data collected from these treatment and comparison groups. Data was collected from pre and post-assessments and classroom observations in the comparison classrooms were completed to discern the learning environment for the students and document similarities and differences between the treatment and comparison classrooms. Also, student proof portfolios were collected from students in the experimental course and included the first to final draft of several proofs and the other students’ comments. This was to answer the third research question about the effects of collaborative revision on students’ proof construction
skills as well as to document the collaborative revision process. Several students from both types of course were also interviewed about their responses on the proof validation and proof construction tasks to determine the proof schemes (Harel and Sowder, 1998) employed by students and how each student’s proof schemes evolved during the semester. The interviews also gave information about the proof schemes held by students in the treatment course versus students in the comparison course.

4.1 Context

This study was conducted at a large midwestern university in the Mathematics, Statistics & Computer Science (MSCS) department. Student participants were enrolled in a course entitled “Special Topics in Mathematics” where a teaching intervention was enacted. We subsequently refer to this group as the treatment or experimental group. Additionally, a comparison group was formed from students enrolled in a correlated course entitled “Introduction to Advanced Mathematics”. The treatment course is a supplementary workshop for the comparison course. These courses serve as bridge courses to help students transition from calculus, where the problems are mainly procedural, to higher level proof-based mathematics courses. A detailed description of the two courses, including educational goals and content, will be given below.

It is important to note that this was a quasi-experiment, with the comparison course and treatment course not formed by random assignation. The students in the treatment course were self-selected by their enrollment. Also, it was possible for students to be enrolled in both the treatment and comparison course, which was the case for many of the study participants (80% of the treatment group). However, there have been some precautions taken to ensure
that the comparison group would still give valid data for the study. Following the guidelines outlined in Shadish et al. (2002) for strengthening a quasi-experimental study, this study uses an internal comparison group, taken from the same class as the students in the treatment group. Additionally, a pre-test is used to “examine selection biases and attrition as sources of observed effects” (Shadish et al., 2002, p. 158). A further discussion of the participants in the two groups is given in the next two sections.

4.2 Courses & Participants

Data was collected from a treatment group comprised of students who were enrolled in the treatment course and consented to participate in this study. In order to answer the research questions in this study, a comparison group was also desired, which included students enrolled in a comparison course that consented to participate in the research. We describe the details about each of the courses in the next sections as well as demographic data about the participants in each group.

4.2.1 Treatment Group

The course where the teaching experiment was enacted was a content-correlated, one-credit supplemental course offered through the Emerging Scholars Program (ESP) at the university. This course was designed to be highly collaborative and allow students the opportunity to work on challenging problems. The goal of the course, according to the MSCS department, is to supplement the introduction to proof course, so the content is the same as described above, but students are given opportunities to delve deeper into the material. Traditionally, this course is taught by a graduate student teaching assistant at the university and the design of the course
varies widely depending on the instructor. For the Fall 2012 semester, I was the instructor of this course and thus designed it to incorporate the collaborative revision process.

The course met once a week for one hour and fifty minutes. For a comprehensive overview of the classroom activities during the semester, refer to Section 4.4. Since this workshop already incorporated collaborative learning techniques, the instruction given was within the educational goals of the program. The next section gives an overview of the ESP program and its history and implementation at this university. Then we report details about the demographics and characteristics of the participants in this treatment group.

4.2.1.1 The Emerging Scholars Program (ESP)

The Emerging Scholars Program (ESP) was conceived first in the late 1970’s by mathematics education researcher Uri Treisman while at University of California at Berkeley to counteract the low passing rates in calculus of minority students (Brugueras et al., 2005). An analysis of students passing first-semester calculus, found that 60% of African-American students at Berkeley were receiving grades of D or F (Treisman, 1992). However, interviews with students found that these attrition problems were not due to any of the hypotheses put forth in the study, which were that African-American students may have lower motivation, inadequate preparation for the course, unsupportive communities and families, or lower income (Oppland, 2010). Thus, a comparison of the study habits between groups of African-American students and Asian students was conducted to find an explanation for the low passing rates of the African-Americans versus the relatively high passing rates of the Asian students. It was found that:
“both the African American and Chinese students practiced the mathematical study habits recommended by their institution’s study skill courses while engaging in their calculus courses (e.g. regularly attending class, taking notes, studying individually for several hours each week, and completing and submitting homework assignments on time). However, the Chinese students’ study habits differed from the Black students in that they studied individually for approximately eight hours more per week, discussed solutions to problems with their peer, categorized problems by level of difficulty (which aided in minimizing the level of frustration they experienced), worked through old mathematics exams, completed tests administered by family members, and were aware of how they were performing in the course in relation to their peers” (Oppland, 2010, pg. 15).

Thus, the African-American students were practicing good study skills, but they were not interacting with other students outside of class or discuss the mathematics with other students. Unlike the African-American students, the Chinese students were strongly incorporating mathematics into their lifestyle and spending time interacting with each other to enhance their understanding of the material (Brugueras et al., 2005). Based on this research, Berkeley implemented a program, encouraging social interaction in the classroom by having students work on challenging problems in small groups in hopes that these students would also collaborative outside of the classroom, that became the Emerging Scholars Program.

In 1989, the university in this study implemented their own version of the Emerging Scholars Program as a response to similar low performance of minorities in pre-calculus and calculus courses (Brugueras et al., 2005). The university gives the following description of the program: “The Emerging Scholars Program (ESP) offers an opportunity to work on challenging mathematics problems with classmates through innovative techniques of cooperative learning. Students meet for a regularly scheduled workshops weekly to work collaboratively in small
groups on math problems that emphasize the key ideas of the corresponding math course”1.

ESP workshops are currently offered as supplemental courses correlated with Pre-calculus, Calculus I - III (i.e. differential calculus, integral calculus, and multi-variable calculus), and the introduction to proof course. The workshops meet twice a week for one hour and fifty minutes for Precalculus, Calculus I and Calculus II and once a week for one hour and fifty minutes for Calculus III and introduction to proof.

The instructors of each workshop are graduate teaching assistants at the university, who are hand-selected by the ESP coordinator. Since, “instructors highly dedicated are preferred to conduct the workshops, consequently the ESP coordinator tries to recruit teaching assistants recommended by the assistant director of graduate studies, who is in constant contact with graduate students. Also, recommendations from other teaching assistants that have participated in the program are obtained” (Brugueras et al., 2005, pg. 4). Monthly meetings are held for the instructors by the ESP coordinator so problems can be discussed and ideas can be shared. I have been an instructor in ESP for six semesters and, during this time, I have taught workshops for Pre-calculus, Calculus I (i.e. differential calculus), Calculus III (i.e. multi-variable calculus) and introduction to proof.

Although originally designed for underrepresented students to participate, ESP courses are open-enrollment and any student can enroll with the ESP coordinator’s permission. However, minority students are aggressively recruited for the program. During the first year that ESP

1The Emerging Scholars Program website can be found at http://homepages.math.uic.edu/~uicamp/esp.html
courses were offered, “letters inviting 200 recipients of the Presidents Award Program scholar-
ships and to approximately 50 students who obtained A or B in Pre-Calculus Algebra or
Trigonometry” (Brugueras et al., 2005, pg. 4) were sent out to recruit students into the course.
However, today recruiting is done by making announcements in courses during the first week of
classes, sending emails or letters to all students enrolled in the correlated course, or by directing
advisors to encourage students to take the course.

4.2.1.2 Treatment Group Participants

All students in my ESP workshop during the Fall 2012 semester were recruited to participate
in the treatment group of this study. There were two sections of this ESP workshop offered
during this semester and I was the instructor of both. There were a total of 15 participants\(^1\)
in this group, with 12 students enrolled in both the treatment and comparison courses and 3
enrolled only in the treatment course. There were 9 male and 6 female participants in this
group. A third of the participants were majoring or double majoring in mathematics, with
the rest split between computer science or mathematical computer science, civil engineering,
biology, biochemistry, physics, economics and one who was undeclared. Two students from
the treatment course were selected for pre and post interviews regarding their responses on an
assessment tool (see Section 4.3.1) used in this study. Of the students interviewed, one was a
male mathematics major and the other was a female Computer Science major. The average
grade point average (GPA) of students in the treatment course prior to the Fall 2012 semester

\(^1\)A total of 19 students were enrolled in this ESP workshop in Fall 2012. Of the 19 students, three
did not complete the course and one did not consent to participate in this study.
TABLE I

SUMMARY OF CHARACTERISTICS OF PARTICIPANTS IN THE TREATMENT GROUP

was 3.32 and the average mathematics GPA\(^1\) of these students was 3.33. The demographics of this group are summarized in Table I.

The students in the treatment group were given a survey, shown in Appendix E to gain information on prior ESP courses that they have taken and if they prefer to work on mathematics alone or in groups. Of the 15 participants from the treatment group, two students had taken an ESP course previously. When asked how they like to work on mathematics, only one person said she preferred to work on mathematics in groups. Two other students mentioned that they liked to work on problems first by themselves and then they like to check their work and go

\(^1\)This was calculated by computing the GPA of only the mathematics classes taken prior to the Fall 2012 semester using the standard GPA formula of total points divided by total credits.
over problems they are stuck on in groups. This suggests that students in the treatment group are not predisposed to working in groups on mathematics problems.

4.2.2 Comparison Group

Introduction to Advanced Mathematics serves as a first course in proof for undergraduate students at this university. The description from the course catalog reads that it functions as an “introduction to methods of proofs used in different fields in mathematics.” Content of the course varies widely by instructor, but normally includes the introduction of several proof methods, such as direct proof, induction, contradiction, set theory and counting proofs, as well as some special topics chosen by the instructor such as countability of sets, combinatorics or logic. The usual student population of this course is second year undergraduates who have completed the Calculus sequence (i.e. single and multi-variable calculus). However this course is sometimes taken concurrently with Calculus III (i.e. multi-variable calculus) as the only prerequisite is Calculus II (i.e. integral calculus). Thus, the proofs are primarily basic set theory and number theory proofs and do not include any concepts from real analysis or linear algebras. A textbook, An Introduction to Mathematical Reasoning (Eccles, 1998) was used for this course, but often instructors would make up their own homework assignments instead of choosing problems from the book.

Students were sampled from two sections (out of four offered) for the comparison group in this study. These sections were taught in the traditional lecture style, where the instructor introduces the material and demonstrates the writing of several proofs in class and then students are expected to write their own proofs for homework and on assessments. This course met three
times a week for 50 minutes. This was determined from communication with the professors teaching these sections, syllabus analysis, and classroom observations. Classroom observations were completed twice during the course of the semester to determine the learning environment for the students and more information can be found in Section 4.3.3. Two of the four sections offered were excluded from this study. The first section was not included because the professor communicated to me that he would be teaching the course using a Modified Moore Method (see Section 4.4.3) approach. This was undesirable to use as a comparison because information was needed about students’ proof construction and validation skills as compared to a lecture-based course. Another section was excluded because it was added to the schedule late, which did not allow enough time to communicate with the professor about participation in the study. The student participants from the comparison group are described below.

4.2.2.1 Comparison Group Participants

Students from two sections of the comparison course offered during the Fall 2012 semester were recruited to use as the comparison group. There were a total of 12 participants in this group and none of these students were also enrolled in the treatment course. This group had 5 male and 7 female participants. Only two of these participants were majoring in mathematics, with the majority majoring in computer science or mathematical computer science. The remaining participants were majoring in mathematics education, industrial design and one student did not disclose this information. Two students from the comparison course were also selected for interviews. The students interviewed in this group were one male Mathematics Education major and one female Mathematics major. The average grade point average (GPA) of students
### TABLE II

**SUMMARY OF CHARACTERISTICS OF PARTICIPANTS IN THE COMPARISON GROUP**

in the treatment course prior to the Fall 2012 semester was 3.22 and the average mathematics GPA\(^1\) of these students was 3.36 as shown in Table II.

#### 4.2.3 Similarities and Differences Between Group Participants

There were 27 total student participants\(^2\) in this study, with 12 students enrolled in both the treatment and comparison courses, 12 in the comparison course only and 3 in the treatment course only. As summarized in Table III there were roughly the same number of participants

---

\(^1\)This was calculated by computing the GPA of only the mathematics classes taken prior to the Fall 2012 semester using the standard formula of total points divided by total credits.

\(^2\)There were 43 original participants, however some students did not participate in the post-assessment because they dropped out of the course or were absent on the day it was administered. This sample includes only participants who participated in both assessments.
from each gender (14 males and 13 females). However, the treatment group had more male than female participants, while the comparison group had more females than males.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Treatment</th>
<th>Comparison</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>9</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>Female</td>
<td>6</td>
<td>7</td>
<td>13</td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>12</td>
<td>27</td>
</tr>
</tbody>
</table>

TABLE III

FREQUENCY OF GENDER AMONG STUDENT PARTICIPANTS IN THE STUDY

The frequency of majors among the participants is shown in Table IV separated by students in the treatment group and students in the comparison group. There were a wide range of majors represented in this study. The majors that were represented by one person in the study (classified in the “Other” category) included: economics, physics, biochemistry, industrial design, biology, undeclared and one participant who did not volunteer this information. This course traditionally attracts students majoring in mathematical fields, however our sample was comprised of only about 41% of participants who reported a major or double major of mathematics or mathematics education. There were 5 mathematics majors in the treatment course, while there were only 2 in the comparison course. However, there were 3 mathematics education majors enrolled in the comparison course, which is required for all secondary mathematics education majors.
The mean grade point average (GPA) and mathematics GPA was calculated for all students in this study and is reported in Table V. Note that while the average mathematics GPA for students in the treatment group is slightly higher than for the comparison group, an independent samples \( t \)-test found that this difference was not significant. The average total GPA was roughly the same for both the treatment and comparison group and the difference was also found not statistically significant using an independent samples \( t \)-test.

The characteristics of these two groups suggest, that even though students self-selected to be in the treatment group, the comparison group serves a function similar to that of a control group. It could be hypothesized that the students in the treatment group signed up to receive two additional hours of instruction per week and thus were more motivated, however the GPA data do not show this to be the case. Also, it does not seem that mathematics majors are more or less likely to enroll in the treatment course. Another consideration for comparing these two groups would be the level of mathematical maturity of the students. We can measure this by
<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mathematics GPA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>3.33</td>
<td>0.66</td>
</tr>
<tr>
<td>Comparison</td>
<td>3.22</td>
<td>0.68</td>
</tr>
<tr>
<td>Total GPA</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>3.32</td>
<td>0.45</td>
</tr>
<tr>
<td>Comparison</td>
<td>3.36</td>
<td>0.36</td>
</tr>
</tbody>
</table>

*a There were no significant differences \((p > 0.60)\) between means from the treatment and comparison groups at the \(\alpha = 0.05\) level.

TABLE V

MEANS AND STANDARD DEVIATIONS OF GPAS FOR ALL STUDENT PARTICIPANTS IN THE STUDY

the number of proof-based mathematics courses the students have taken prior to enrolling in this transition to proof course. This course, being a transition to proof course, could be the first time that many students have had to construct proofs. However, if a large number of students in the treatment or comparison course had taken a previous proof-based course, this could affect the mathematical maturity and, thus, results from that group. As shown in Table VI, there are about the same number of students in treatment and comparison group who have taken a previous proof-based mathematics course. The most common proof-based course for students to have taken prior to enrollment in this course was Linear Algebra. However, a few students had taken a finite mathematics course, analysis or abstract algebra. The table shows that the mathematical maturity is comparable for both groups.

4.3 Data Sources

Multiple data sources, which included an assessment, student interviews, classroom observations and note-taking, collection of student work in proof portfolios and field notes about the
classroom experience, were used in this study. Collection of this data allowed for a holistic view of the collaborative revision process as well as a measurement of the impacts of collaborative revision on student proof construction and validation skills.

4.3.1 Assessments

Since information was desired about students’ views on the nature and function of mathematical proof as well as the abilities of students to identify valid proofs, assessments were administered to both the experimental and comparison courses on these topics. The same assessment was given as a pre and post-test in order to obtain data about student learning throughout the semester and to compare student learning between the experimental and comparison courses. The assessment, which can be found in Appendix A, was given the second week of classes and again during the last week of classes of a 16-week semester. Three parts were contained in the assessment: questions concerning the role of proof in mathematics, four “proofs” of a statement that students evaluated as either valid or invalid, and a mathematical statement that required students to construct a proof.

<table>
<thead>
<tr>
<th>Previous Courses</th>
<th>Group</th>
<th>1</th>
<th>2</th>
<th>3+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of Students</td>
<td>Treatment</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>4</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

**TABLE VI**

FREQUENCY OF STUDENTS IN THE TREATMENT AND COMPARISON GROUPS WHO HAVE TAKEN PROOF-BASED MATHEMATICS COURSES PRIOR TO ENROLLMENT IN THIS COURSE.
The first part of the assessment focused on students’ beliefs about the functions of proof in mathematics. The first question, modeled on a study by Healy and Hoyles (2000), is open response and asks students for general thoughts on the purpose of proof in mathematics. This question is an opportunity to determine whether students see proof as integral to the practice of mathematics, merely as exercises to do in mathematics classes or as serving some other purpose. Additional questions in the first part of the assessment concerned the roles of proof outlined in the literature. Statements are given about the exploratory, verification, intellectual challenge, communication, discovery, axiomatization, justification of definition and creating autonomy roles of proof. Students were asked to rate their agreement with each statement on a five-point Likert scale, asking students if they strongly agree, agree, disagree, strongly disagree or are neutral on the statements. The purpose of these questions was to determine if students consider other functions of proof besides verification and if students believe that some functions of proof are more important than others.

The second part of the assessment asked students to examine four possible ‘proofs’ of a given statement and determine whether each was a valid or invalid proof. The proofs presented to the students are adapted for this study from the proofs given to high school students in Healy and Hoyles (2000) to be appropriate for undergraduates. This included adapting the language to be more formal including details and wording similar to what the students would see in their textbook. The proof validation task did not require outside concepts or theorems that students might not remember from a previous course and the ‘proofs’ were short enough that the students should have been able to digest them in a reasonable amount of time. The
students were first asked to determine if each proof is valid or invalid. The terms valid and invalid were intentionally not defined in order to see if students gain more of an understanding of what a valid proof entails throughout the course of the semester. In addition to the open response question, there were two more questions on a three-point Likert scale asking students whether or not they felt they understood the proof and how certain they were about their classification. The same proofs to be validated were given on both the pre-test and post-test to gain comparative data on whether students changed their minds about the validity of the proofs after exposure to the course.

Each of the proofs given in the validation component of the assessment was modified from a similar assessment given by Healy and Hoyles (2000). Thus, we adapt the same framework from Balacheff (1988) and Coe and Ruthven (1994) for inclusion of each proof type as they did in this study. This framework relies on two types of proof approaches identified by Balacheff: naïve empiricism and the crucial experiment as well as Coe and Ruthven’s (1994) codification of proofs as empirical, weak deductive or strong deductive. According to Healy and Hoyles (2000), an attempt was made to include “(a) an argument or arguments characterized as specific, empirical, or requiring an action or concrete demonstration with little or no explanation; (b) an argument that relied on common properties or a generic case; (c) an argument that suggested underlying reasons and explanations written in a narrative everyday style; and (d) a deductive proof, written in a formal style, presenting a logical argument with explicit links made between premises and conclusions” (pg. 401-402). An analysis of the assessment given in Appendix A
will reveal that the first proof is an example of (c), the second proof is an example of (a), the third proof is an example of (b) and the fourth proof is an example of (d).

The third part of the assessment required students to construct a proof of the mathematical statement: *For any positive integer* \( n \), *if* \( n^2 \) *is divisible by 3, then* \( n \) *is divisible by 3*. The aim of this question was to gain insight about students’ prior knowledge coming into the courses, as well as to determine the impact of the course on students’ proof writing skills and to allow a comparison of proof writing skills between the comparison and experimental section, to answer the research question in this study regarding students’ proof construction skills and any impact collaborative revision may have on those skills. This statement, adapted from Selden and Selden (2003), was chosen as the one for students to prove for several reasons. First, this statement requires a short proof and does not contain any ideas that students may not be familiar with or need to recall from prior courses. Thus, it is at a sufficient level for students in an introduction to proof course. Additionally, this statement can be proved in a variety of ways and the way that students approached this proof can give insight into their level of proof construction skills. Finally, the converse of this statement is much easier to prove so a student’s proof can give information about their logical thinking skills if they attempt a proof of the converse. Furthermore, this proof does not appear in the students’ textbook so they are unlikely to have already written or seen a proof of the statement.

### 4.3.2 Interviews

In addition to the assessments, student interviews were conducted with two students in the treatment course and two students in the comparison course. Student interviews were conducted
to gather data from students about how their beliefs about proof functions have transformed during the semester. Specifically, information was desired about the proof schemes (Harel and Sowder, 1998) that students held and if those proof schemes evolved during the semester. Since the assessment asked students merely to choose valid or invalid when faced with a given argument, no information was obtained about thought processes of students around these concepts.

This study sought to determine not only if proof validation skills were improved during the semester for students in the treatment course, but if students completing the treatment course were exhibiting higher level proof schemes than students in the comparison course. This information aided in answering the second research question about specifically how the collaborative revision course could impact students’ proof validation skills. Thus, the interviews were completed to gain insight on student thought processes and to conclude the type of proof schemes that students are using in proof validation.

The sample of students interviewed was chosen purposefully, which is recommended by Fraenkel and Wallen (2009), based on their performance on the pre-assessment and on gender. From the students consenting to an interview, I looked at each student’s pre-assessment and chose one student at each of the high, middle and low performing levels in each type of class. Assessing the level of the students was performed by looking at the responses in the pre-assessment and determining which students who consented to interviews had the most correct responses on the proof validation and proof construction tasks, which students had the least correct responses and which had some correct responses. This was done to be able to compare students at each level and determine the impacts of the comparison course and the treatment
course on each type of student. Additionally, a sample was sought with an equal number of men and women to minimize gender bias in the study.

Students were interviewed individually twice during the course of the semester; first shortly after the pre-assessment was administered and again shortly after the post-assessment. Interviews were semi-structured and an interview guideline can be found in Appendix C. The questions asked to students, according to Zazkis and Hazzan (1998), are unexpected “why” questions and reflection questions, requiring students to explain thought processes during the proof validation tasks on the assessment. Unexpected “why” questions are employed to clarify student responses when they assert something to be true, while “reflection questions, rather than solving a mathematical problem, interviewees are asked to reflect on a solution presented by an imaginary third person” (pg. 434). In this context, the students were reflecting on their own solutions in one of two ways. Either students were asked to look at their own proof construction or proof validation tasks and reflect on why they answered that way or they were shown their answers on both the pre and post-assessments and asked to compare them and reflect on why their responses had or had not changed.

4.3.3 Classroom Observations

Classroom visits were conducted in the comparison course to document the similarities and differences between the learning environments in each type of course. The study was designed such that the experimental course spent minimal time on lecture, with small group discussion dominating the time spent in class. In contrast, a hypothesis of this study is that the comparison course used the majority of class time for lecture, with minimal student interaction taking place
during class meetings. The goal of the classroom observations was two-fold; first, to make sure
that the assumed differences between the courses were indeed present in the students’ learning
environment and, second, to determine how collaborative revision could enhance a lecture-based
classroom.

Observation of the comparison course took place twice during the semester and the entire
50-minute class meeting time was observed. Notes were taken using the observation protocol,
which can be found in Appendix B. During these observations, I recorded the estimated per-
centage of class time spent on lecture, small group work, whole-class discussion, and other tasks,
such as student assessment. Since the experimental course consisted mainly of small group work
and very little time spent on lecture, these percentages allow for a direct comparison between
the courses about usage of class time. The role of the instructor during these class sessions was
also observed and described. I classified the role of the teacher in one of several ways: trans-
mitter of knowledge, facilitator of discussion, passive observer, or other, after descriptions in
Blanton et al. (2009). The instructor as transmitter of knowledge would be describe a lecture-
based course, where the teacher spends the whole time in front of the class teaching material
or answering questions, with no student to student discussion. The facilitator of discussion
role refers to when the instructor acts as mediator of a whole-class discussion, posing questions
and keeping discussion on track as needed, but not lecturing. Passive observer refers to when
the teacher organizes group or whole-class discussion, but then just observes the conversations
taking place without adding input. Note that it is possible for a teacher to take on each of
these roles during a single class session or to take on a role that is distinct from these.
As instructor of the experimental course, I experienced this course first hand and took field notes about student interactions and the usage of class time during each class meeting. These notes included information about which students were in each group, which statements were given to which students to prove and any notable interactions that took place during class. After each class, I would note what was completed during that session and make estimations of usage of class time on lecture, small group work and whole-class discussion.

4.3.4 Proof Portfolios

During the course of the semester, each student in the experimental course was instructed to submit proofs for a portfolio, which included first and subsequent drafts of five mathematical statements. The course syllabus is included in Appendix D for reference. Students were given statements to prove, drawn from various introduction to proof textbooks and other sources, based around whatever topic we were addressing that week. After given a statement, a student was to attempt the proof at home and bring to class a written copy of their proof. They then took turns in small groups presenting the proof to their classmates on the board or on paper and eliciting comments from the group. One group member was responsible for taking notes and turning in the notes to me at the end of class. The student note taker was instructed to document the comments made to a presenter by the other students. Once a student’s group determined that the proof was valid and well-written, the presenting student submitted to me all the drafts they had and I put these together with the notes taken during each presentation into the student’s portfolio. Since the revision process took a number of weeks and different
students were absent week to week, not all students submitted five complete proofs during the course of the semester.

Collection of this data allowed for a comprehensive analysis of the revision process and allow me to see how the other students comments affected the subsequent drafts of a proof. This was in order to answer the third research question regarding the impacts of this treatment course on proof construction skills of the students. In an attempt to standardize the collaborative revision process and obtain data that is useful, I implemented some guidelines for the students to use when writing and revising their proofs. I provided each student with a folder to use as their proof portfolio and I brought the portfolios to class each session and then collected them at the end of class to make sure that the students would not lose them. Additionally, I provided students with paper packets in which to write all of their proofs, providing more paper as needed. This meant that the students used the provided paper exclusively when doing their proofs and then transferred the paper directly into their portfolio, making it easy to keep all their work together. Furthermore, students were instructed not to erase any work, but instead to cross out any incorrect work so that I could still see it for my data analysis.

4.4 Description of Teaching Experiment

In order to answer the second and third research questions about impacts of collaborative revision on students’ proving skills, in the Fall 2012 semester a teaching experiment was conducted. According to Steffe and Thompson (2000), “a primary purpose for using teaching experiment methodology is for researchers to experience, firsthand, students’ mathematical learning and reasoning”. This teaching experiment was designed to maximize the amount of
time spent working in groups having students explain and critique proofs in order to enhance students' proof validation skills and students' proof construction skills.

This teaching intervention employed collaborative revision in the classroom, which refers to the process in which students explain a proof they have written to their classmates and the other students are encouraged to make comments and point out inconsistencies in order to ensure that the proof is valid and that it is written coherently. Based on feedback from classmates, the student then revises the proof and presents it again, repeating the process until the proof is valid and includes all the relevant details. The experimental course was structured so that a student would present a proof they have constructed outside of class to a small group and the group would then provide reactions, observations, and criticisms of the proof. The student presenter would then revise the proof according to the group discussion and present it again, repeating the process until the group reaches a consensus about the validity of the proof.

The concept of collaborative revision is borrowed from writing classes typically taught in an English department, where they tend to call this process “workshopping.” A description of this process is given in a Creative Nonfiction course syllabus from Duke University: “A writer shares copies of a piece she is working on with the members of her group, reads a section of the text aloud, and then quietly takes notes as her readers offer their impressions and advice. After all the readers have responded, the writer may ask them questions and the group can have a less structured conversation about the piece. Then the group moves on to the next writer and his piece and the same process is repeated until everyone in the group has received feedback on their work” (Harris, 2011). This research project implements this technique, but instead of
essay pieces students shared copies of their proofs with their classmates and then the class will have a structured discussion about the proof. Even though similar types of experiments have been done in mathematics education\(^1\), there is no literature on this exact process and thus the term *collaborative revision* has been coined for this study.

### 4.4.1 Classroom Activities

The following is a description of the day to day activities that took place in the teaching experiment classroom from field notes taken during the sixteen week semester. For the first two weeks of the semester, class time was spent not on the collaborative revision process, but on familiarizing the students with the collaborative norms of the classroom. In a traditional classroom, students are often used to a lecture environment, where they may ask questions occasionally, but are otherwise not encourage to talk. Thus, it took training for the students to properly engage in the collaborative revision process. These two weeks were spent on the study of mathematical logic and tasks were given to students concerning logical operators, negations of logical statements, truth tables, and logical equivalence, quantifiers and negations of quantifiers. Part of the first class included minimal instruction on the basic structure of direct proofs to make sure students were all at the same level. During these two weeks, students were often asked to work in small groups and discuss prompts or work on problems. The rationale for this approach was that it takes time to get students used to the adjusted classroom norms that are

\(^1\)For example, see 4.4.3 for information about the Modified Moore Method.
required when using collaborative revision, such as the role of the teacher as a facilitator and addressing questions directly to each other when working in groups.

Starting in the third week of the semester, the collaborative revision process was explained to students and at the end of class during this week, the students were given their first statement to prove at home. The statements to prove were simple number theory statements regarding divisibility or properties of even and odd numbers. See Sample Proof Set 1 in Appendix G for the exact statements that students were given to prove during this week.

Then, during the fourth week, the first round of collaborative revision took place. Students were put into groups of 3–4 and asked to go through each person’s proof, one at a time. The student whose turn it was would write their proof onto a packet of paper handed out by the instructor, meanwhile explaining each step to the other students. The other students were encouraged to ask questions to the presenter and make comments on parts they like, parts they don’t understand or parts they had logical concerns about. At the end of class, each student handed in their paper to the instructor for inclusion into their proof portfolio.

This structure continued for each of the remaining weeks of the semester. Students would be put into groups with revolving members and either re-present a revision of a proof from a prior week or present a new proof. Each week a new statement to be proven was handed out to each student, which they were to bring in a proof of for the following week. In Appendix G several sample proof sets are shown. Note that there are nine proofs in each set because there were up to nine students enrolled in each section of the treatment course. Notice also that each set of statements to be proven are focused around a single mathematical topic. This was done
on purpose so that students would presumably have had some experience with the topic while attempting a proof of the statement given to them and, thus, be better able to evaluate their classmates’ proofs. All drafts of each proof were collected and put into a proof portfolio to document the revision process.

4.4.2 Role of the teacher

During this process, the instructor had several roles. The first was to do the preparation work of gathering appropriate proofs to give to the students and to assign groups. As the instructor, I have identified many proofs that would be challenging yet solvable and did not appear in the students regular text, to avoid giving students a mathematical statement that they have already proven. Several sample proof sets can be found in Appendix G. Formation of groups also becomes an important task for the instructor in a collaborative revision classroom. Groups of students were assigned for each new round of student presentations and the same groups were kept during each iteration of proof revisions. Group size was set at a maximum of 4, which is consistent with the rationale of Davidson (1990): “groups with four members seem to work best. They are large enough to generate ideas for the discussion and solution of challenging problems, yet not be decimated by the absence of one member. They are small enough to permit active participation, to allow clustering around a chalkboard panel, and not to require a leader or elaborate organizational structure” (pg. 56).

Another function of the instructor was to regulate and facilitate discussions. When students were off-task it was my job to redirect the students to the task at hand. Additionally, as instructor I often needed to aid groups that were stuck on a particular task. This sometimes
took the form of the presenting a certain example or counterexample to the students to help them with their arguments or suggesting a direction in which to proceed. Finally, the most important role of the instructor was to create an environment of collaboration in the classroom. This included asking students to direct their comments and questions towards other students instead of the instructor and stepping away from the front of the classroom to denote a facilitator role. This was often also the most difficult role, as students were unaccustomed to this type of classroom setup and were used to the instructor being in front of the class and transmitting all of the information. Students also tended to be somewhat uncritical of their colleagues' proofs, which could be a function of inexperience at validating proofs or a function of not wanting to hurt a classmate's feelings. Thus, I spent a lot of time at the beginning of the course asking my own questions about the students' proofs and asking some of the other students if they understood certain parts and if they could explain that argument in their own words. This often resulted in many more questions being asked about the proof and during the later weeks of the semester I did not need to do this as much as the students had adopted a more questioning attitude.

4.4.3 Comparison to Modified Moore Method

At this point, it is important to note the differences between the collaborative revision process and the Modified Moore method. Although the exact process is new for the purposes of this study, the idea of students generating and presenting proofs dates back to at least the 1920s with the method of R.L. Moore (Mahavier, 1999). This approach involved giving students a list of axioms and definitions and then a list of results to prove throughout the
semester. The students were responsible for attempting to prove the theorems on the list at home, without the use of any outside resources. When a student thought they had a complete proof, they would present it to the class, who would then ask questions and critique the proof. The Modified Moore method gives the students a bit more direction, which may include, among other things, assignments of short exercises on the definitions before the students attempt the theorems (Chalice, 1995), the use of outside resources by students or supplemental lecture by the instructor (Dhaher, 2007). Collaborative revision does share some common elements with the Moore method, such as that students are responsible for discovery outside of the classroom and it requires student presentations. However, collaborative revision is designed to be done in small groups and not as a whole-class exercise. Additionally, collaborative revision in the transition to proof course does require some lecture on the basic concepts of proof construction, which the students will receive in the comparison course. Revision also plays a crucial part in the design of this study and students will be required to present the same proof multiple times, taking into consideration their classmates feedback.

4.5 Description of Comparison Course

What follows is a description of the learning environment provided to students in the comparison course. The descriptions are taken from field notes taken using the observation protocol shown in Appendix B. Course websites and syllabi were also examined to determine the topics being covered in each section and the course grading structure.

The comparison group included students enrolled in two sections of the introduction to proof course at the university. As described in Section 4.3.3, I visited each of these sections
twice during the semester to observe the learning environment. During my first observation, each of the section instructors lectured on the topic chosen for that day and presented definitions and axioms and then gave proofs of several theorems. The environment was such that students were encouraged to ask questions if they didn’t understand a concept or a part of a proof. In my observation, several students did ask such questions and the instructor expertly answered their concern. There was no group work or whole-class discussion that took place in either section during my first observation. The instructor lectured for roughly 90% of the class time with the other 10% used up by student questions and housekeeping things (i.e. passing back homework, talking about test dates, etc.).

During the second observation, one of the sections remained unchanged. Again the professor went through the material to be covered and presented more definitions and proofs. It was still about 90% lecture and 10% questions and miscellaneous things to be discussed. The other section was run slightly differently. Since it was the last week of the semester, the other section was preparing for the final and having a review session. Students were asking about statements to be proved off of a review sheet provided by the professor and the professor was going through the proof. This resulted in the instructor at the board filling about 80% of the time and the students asking questions from the review sheet and about the proofs the instructor was writing taking up about 20% of the time.

4.6 Summary

As outlined above, this research project employs a teaching intervention using a process called collaborative revision and is designed to determine if there are any impacts on students’
understanding of proof function, proof validation skills and proof construction skills. The study design used participants from two groups to draw inferences; a treatment group and a comparison group. Data was collected from four sources to answer the research questions and analyze the collaborative revision process itself. In Chapter 5, we discuss how each of these data sources was analyzed and we give results related to our research questions.
CHAPTER 5

DATA ANALYSIS & RESULTS

As previously stated, this study was designed to answer the following research questions:

1. (a) What are undergraduate students beliefs about the functions of proof in mathematics before and after a transition to proof course?
   
   (b) Are there differences present in undergraduate students’ beliefs about proof function between students who participated in collaborative revision and those who did not? If so, what are they?

2. (a) What are undergraduate students’ proof validation skills before and after a transition to proof course?
   
   (b) Are there differences present in proof validation skills between students who participated in collaborative revision and those who did not? If so, what are they?

3. (a) What are undergraduate students’ proof construction skills before and after a transition to proof course?
   
   (b) Are there differences present in proof construction skills between students who participated in collaborative revision and those who did not? If so, what are they?

This chapter gives a detailed report of the data analysis techniques used and the results in this study. Each section addresses one of these research questions and outlines the methods of
analyzing the data collected to answer that question and gives the results from the analysis. A variety of qualitative and quantitative methods were used to interpret the data and to get a comprehensive understanding of the impacts that collaborative revision has on student learning about proof.

5.1 Students’ Beliefs About Proof Functions

To answer the first research question regarding students’ initial beliefs about the functions of proof and if these beliefs evolve due to participation in collaborative revision, data from the pre and post-assessments was analyzed. Qualitative analysis was performed on students’ comments to an open response question about the role of proof in mathematics. The open response question asked “What do you believe is the purpose of proof in mathematics?” and provided space for students to write their comments. Additionally, the assessment contained eight Likert scale items which were quantitatively analyzed to gain a comprehensive understanding of the students’ beliefs about proof function in the treatment course and determine if they were any different from the comparison group. The functions of proof identified in mathematics education research (i.e. verification, explanation, discovery, intellectual challenge, axiomatizing, justification for a definition, illustrating techniques, communication and providing autonomy) were used as a framework for data analysis. All student answers to the open response question and the coding applied can be found in Appendix F.

5.1.1 Data Analysis

Since students were asked not only to write comments about their beliefs about proof, but also were asked about their level of agreement on a Likert scale with a variety of statements
about the role of proof, both qualitative and quantitative techniques were used to analyze data. First, a discussion of the qualitative data analysis performed on the constructed response question from the assessment is given. The coding scheme applied to students’ responses is detailed and a discussion of how reliability was obtained in the coding process is outlined. Second, we outline the quantitative data analysis performed from the attitudinal questions about proof function and report on the statistical tests used to gain results.

5.1.1.1 Qualitative Data Analysis

Students were asked on both the pre and post-assessment the open response question “What do you believe is the purpose of proof in mathematics?” A grounded theory approach (Char- maz, 2006) was used in the data analysis, using the following functions of proof outlined in the literature as a framework for coding the data: verification, explanation, discovery, intellectual challenge, communication, axiomatization, illustration of techniques, justification of a definition and providing autonomy. Analysis of the data provided evidence for the creation of four additional codes: understanding, thinking skills, building mathematics and logical skills. A summary of each code that was developed during the data analysis can be found in Table VII along with a short description of that code and a sample student comment.

The four additional codes that were created during the data analysis process speak to proof as more of a pedagogical tool than the functions of proof from the literature. The understanding function of proof refers to using proof as a tool to understand certain concepts in mathematics and identifying connections between branches of mathematics. The function of thinking skills encapsulates the many comments that discussed the different thought processes needed when
constructing a proof. Several comments referred to mathematics as a discipline where concepts are built on one another and how proof exposes this relationship and these were categorized under *building mathematics*. Other comments spoke to logic as a basis for mathematical proof or reasoning skills as a necessary component of proof construction, and these were classified under *logical skills*. Students’ identification of proof for understanding and for developing logical skills is consistent with the findings of Knuth (2002a) when a similar question was asked to secondary teachers.

Each student response was coded for each instance of a proof function that appeared. Codes were then separated and tallied first by group (treatment or comparison) and then by assessment (pre or post). Refer to Appendix F for a table showing all the comments coded for a proof function separated by group and assessment. The number of codes applied to each student’s response on each assessment were tallied for the treatment group and comparison group students to determine if students, on average, recognize that proof has multiple functions in mathematics. Often responses contained mention of several different functions of proof, thus a student may have more than one code applied to his or her response. Additionally, if a student spoke about a single function of proof in several different ways a single code could be applied to a student more than once.

When applying codes, certain keywords were looked for in order to apply the correct code. For instance, the verification code (V) was often applied to a statement if it had the words “verify”, “validity”, “true” or “false”. Also, if a response had the words “explain” or “why”, it was coded as explanation (E). Additionally, several responses had the words ”build”, ”logic”
or "challenge" in them, which led to a code of building mathematics (B), logical skills (L) or intellectual challenge (C), respectively. However, context was very important when coding a response and comments were also looked at comprehensively to see if the words were being used in a way that is consistent with the code descriptions given in Table VII. Once all the responses were coded, all the comments were organized into the table found in Appendix F. Any code that was applied at least once during the coding process became a theme heading. Thus, the code names are the same as the theme headings that we will discuss in the next section on results.

To obtain reliability in the coding process, it was decided that performing a member-check with the students themselves would be too cumbersome. A member-check is when the researcher takes her interpretation of the data to the participants themselves to check for authenticity, thereby improving the accuracy and credibility of the report (Fraenkel and Wallen, 2009). However, it would have been difficult to track down all the students a semester later when the analysis was completed. Therefore, the technique of triangulation was used to ensure credibility and validity (Creswell, 2008) in this study. This was done by asking a qualified colleague to review the themes and comments that I had coded. The colleague was a graduate student in pure mathematics who was interested in mathematics education and had extensive experience teaching in the Emerging Scholars Program. The course in this study where the teaching experiment was implemented he had previously taught and was thus familiar with the level and the content. My colleague was given the students’ comments from both the pre-assessment and post-assessment without names attached and the code list with explanations shown in
<table>
<thead>
<tr>
<th>Code</th>
<th>Proof Function</th>
<th>Description</th>
<th>Sample Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>V</td>
<td>Verification</td>
<td>Proof as necessary for verifying truth/falsity of a statement</td>
<td>“To accept or deny proposed ideas.”</td>
</tr>
<tr>
<td>E</td>
<td>Explanation</td>
<td>Proof as being able to explain why a statement is true/false</td>
<td>“To explain why certain statements/operation are true.”</td>
</tr>
<tr>
<td>D</td>
<td>Discovery</td>
<td>Proof as a tool for discovering new mathematics</td>
<td>“...to use known facts and discover new insight into mathematics.”</td>
</tr>
<tr>
<td>C</td>
<td>Intellectual Challenge</td>
<td>Proof as being intellectually challenging</td>
<td>“They challenge intellectually.”</td>
</tr>
<tr>
<td>Co</td>
<td>Communication</td>
<td>Proof as important for mathematical communication</td>
<td>“...an undeniable argument of some sort which is readily interpreted by the intended audience.”</td>
</tr>
<tr>
<td>A</td>
<td>Axiomatizing</td>
<td>Proof as a tool for axiomatizing mathematics or incorporating a statement into a hierarchy</td>
<td>“To incorporate a logical statement into the logical hierarchy that already exists.”</td>
</tr>
<tr>
<td>J</td>
<td>Justification of a Definition</td>
<td>Proof as a tool for justifying mathematical definitions</td>
<td>No student examples found</td>
</tr>
<tr>
<td>IT</td>
<td>Illustrating Techniques</td>
<td>Proof as a tool for showing different proof techniques</td>
<td>“To learn how to prove things we already know...”</td>
</tr>
<tr>
<td>Au</td>
<td>Providing autonomy</td>
<td>Proof as a tool for students to understand and create new mathematics</td>
<td>No student examples found</td>
</tr>
<tr>
<td>U</td>
<td>Understanding</td>
<td>Proof for understanding or understanding deeper the concepts of mathematics</td>
<td>“...to understand the basics of mathematics and to see the origin of where things come from.”</td>
</tr>
<tr>
<td>TS</td>
<td>Thinking skills</td>
<td>Proofs as requiring a “different” sort of thinking skills</td>
<td>“Also, to be forced to think more critically and use a part of our mind that isn’t always required in math classes.”</td>
</tr>
<tr>
<td>B</td>
<td>Building Mathematics</td>
<td>Proof as a way to build and categorize mathematics</td>
<td>“To derive conclusions or further information from earlier information which has been assumed or already proven.”</td>
</tr>
<tr>
<td>L</td>
<td>Logic Skills</td>
<td>Proof as being related to logic and requiring reasoning skills</td>
<td>“Also, to develop and expand logic skills.”</td>
</tr>
</tbody>
</table>

**TABLE VII**

CODES USED FOR QUALITATIVE ANALYSIS OF PROOF FUNCTION DATA
Table VII. He was instructed to assign codes to each student’s comments and to create new codes if he felt it necessary. Additionally, he was told that a student’s comments could have multiple codes and that the same code could be used twice on the same comment. He did not create any new codes and there were very few comments that we disagreed on. After he analyzed all of the comments, we met to resolve any differences in the coding.

5.1.1.2 Quantitative Data Analysis

As shown in Appendix A, students were also asked on the pre and post-assessments about their agreement on a 5-point Likert scale with 8 statements designed around the functions of proof outlined in the literature. Statements were given to students regarding the importance of proof and the verification, explanation, discovery, intellectual challenge, communication, axiomatizing and providing autonomy functions of proof. The exact statements are shown in Table VIII. Students were asked to give a rating of 1 (strongly disagree), 2 (disagree), 3 (neutral), 4 (agree) or 5 (strongly agree) to each statement. All statements were framed positively, with a higher rating on a statement indicating a greater appreciation of that proof function, except for the second statement. Thus, a strong understanding of the functions of proof given in the literature would correspond to a rating of 5 on statement 1 and 3–8 and a rating of 1 on statement 2.

Responses were analyzed in a variety of ways to determine differences in appreciation of proof functions between students in the treatment and comparison courses, as well as differences between the beginning and end of the semester. First, descriptive statistics were computed on the Likert scale responses for the pre and post-assessment and separated by group (i.e.
1. I believe that proofs are very important in the field of mathematics.
2. I believe that proofs are only done to verify whether a given statement is true or false.
3. A proof can explain why a given statement is true.
4. I enjoy writing proofs because they are intellectually challenging.
5. I believe that proofs are an important communication tool in mathematics.
6. I believe that proving can be used to discover new mathematics.
7. Proofs are important in organizing mathematical knowledge.
8. I want to learn proof construction so I can create understand the language of mathematics and create my own proofs.

**TABLE VIII**

**STATEMENTS ABOUT PROOF FUNCTIONS GIVEN TO STUDENTS ON THE ASSESSMENT**

treatment and comparison). Then, independent samples $t$-tests were used to determine if there were any significant differences in the means. Norman (2010) writes that “both theory and data converge on the conclusion that parametric methods examining differences between means, for sample sizes greater than 5, do not require the assumption of normality, and will yield nearly correct answers even for manifestly nonnormal and asymmetric distributions like exponentials,” thus this is an acceptable test to use on this data. These tests serve to answer the first research question about students’ initial beliefs about proof function, how these beliefs change during the semester and differences between the treatment and comparison groups.

In addition to $t$-tests, Pearson’s $r$ correlation coefficients were computed between ratings on each question on the pre-assessment and the post-assessment. This parametric method is appropriate for use in this case because, as Norman (2010) argues: “If the Likert ratings
are ordinal which in turn means that the distributions are highly skewed or have some other undesirable property, then it is a statistical issue about whether or not we can go ahead and calculate correlations. However, there have been a number of studies that...have all shown, using theoretical distributions, that the Pearson correlation is robust with respect to skewness and nonnormality” (pg. 629). Thus, despite the possibility that the distributions are not normal curves it is still appropriate to calculate Pearson’s $r$ correlation coefficient. Calculation of $r$ was to determine if there was a high correlation between a student’s rating of a particular proof function on the pre-assessment and the post-assessment, which would indicate that the student’s beliefs had not changed during the semester. This test was done to determine if students in either group have mutable beliefs about proof or if their beliefs remained static throughout the course.

A total measure of a particular student’s beliefs about roles of proof was also desired. We call this measure the Total Function Score (TFS) and it was computed by adding up the Likert scale responses on each of the eight statements. Since statement 2 was written negatively, the responses on this question were first transformed using a feature of the statistics program SPSS. So, if a student responded on statement 2 with a 1, it was changed to a 5, a response of 2 was changed to a 4, a response of 4 was changed to a 2, a response of 5 was changed to a 1, and a response of 3 was left unchanged. Any statements left blank were treated as a rating of zero as
it was seen as a lack of interest in that particular proof function\textsuperscript{1}. Thus, if all statements were responded to, a student’s TFS could range from 5 to 40. One-sample t-tests were performed on these means to determine if the TFS for each group on the pre-assessment and the post-assessment were significantly different from a score of 24. This score was chosen because it is equivalent to a rating of 3 (neutral) on each statement on the assessment.

5.1.2 Qualitative Results

We first present a detailed description of each theme including representative quotes under each theme heading. The results are then discussed, organized into results from the treatment group followed by results from the comparison group. A discussion of the differences between the two groups completes this section.

5.1.2.1 Themes

We now discuss each of the eleven themes that emerged from the data in detail and give representative quotes from students’ responses on the pre and post-assessments. If a code was applied to at least one student response, then that code became one of the theme headings that we outline below.

Verification

As hypothesized, many students (17 on the pre-assessment, 14 on the post-assessment) identified the verification function of proof; that is, proof is used to determine the truth or falsity of a statement.

\textsuperscript{1}Since ample time was given to the students for the assessments and since the Likert scale statements were the first part of the assessment, there were only four blank responses on either assessment to a statement.
statement. This proof function was the most frequent comment in the pre-assessment and post-assessment responses, although did not appear as frequently in this study as in other studies of the same nature (e.g. Almeida, 1995; Coe and Ruthven, 1994). The quotes falling under this theme were very similar to each other; for example:

“Proofs demonstrate the truth of mathematical statements...”

“To accept or deny proposed ideas.”

“...to explain why a mathematical conclusion is true based on already known truths.”

**Discovery**

Several (6 on the pre-assessment, 4 on the post-assessment) students, in both the pre- and post-assessment, commented that proof can aid in discovering new mathematics theorems. This result contradicts the hypotheses of Moore (1994) and Weber (2002) that students may only understand the verification function of proof. In fact, there were several insightful quotes under this theme, demonstrating that some students understand the role of proof in creation of new mathematics theorems.

“We build the tools to discover and prove new things.”

“...to use known facts and discover new insight into mathematics.”

“Proofs can push math forward into new fields...”

**Illustrating Techniques**

Although there were a small number of comments (6 on the pre-assessment, 2 on the post-
assessment) in this category, it is important to note that a few students did understand the pedagogical role of proof in teaching students various proving techniques.

“To learn how to prove things we already know”

“Use the methods we know or we don’t know yet to prove the mathematical concepts and other methods.”

Explanation

Several students (3 on the pre-assessment, 5 on the post-assessment) in this study identified using proof as a tool for explanation of why a certain theorem was true.

“...explain reasoning and evidence why certain theorems/processes work.”

“...to understand where concepts in mathematics are derived from and why they work the way that they do.”

Several students commented that proof serves both a verification and an explanation role in mathematics, thus connecting these two themes. This finding is similar to arguments by Hanna (1990) when she writes “a proof that explains and a proof that proves are both legitimate proofs. By this I mean that both types of proof meet the requirements for a mathematical proof, and thus serve in equal measure to establish the validity of a mathematical assertion” (pg. 9).

“A mathematical proof serves to not only justify a statement, but also to explain the (or a possible) method by which the statement is justified.”
“I think it is to have a deeper understanding about things, like why and how it is true or false.”

Axiomatization

There were a few comments (2 on the pre-assessment, 2 on the post-assessment) focusing on the role of proof in axiomatizing mathematics; that is, determining a hierarchy of definitions, axioms, lemmas and theorems and illuminating the implicit relationships among them. In contrast to the building mathematics category, these comments explicitly refer to axioms and definitions as the tools to build mathematics. For example,

“...to incorporate a logical statement into the logical hierarchy that already exists.”

“I believe proof in mathematics establishes a (frequently long) connection between known or assumed knowledge in the form of axioms and knowledge found to be true dependent on these initial facts or foundations.”

Intellectual Challenge

This theme is the only one that refers to proof as not serving some utilitarian purpose, but instead having importance merely due to causing enjoyment. Two students commented about this role of proof, but responses under this theme only occurred on the pre-assessment.

“...proofs are fun!”

“They challenge intellectually...”

Communication

Proof plays an integral role in communication in mathematics. Indeed, as previously discussed,
Rav (1999) and Hanna and Barbeau (2011) agree that proofs are of the utmost importance in mathematics, since much more than theorems, proofs are the main vehicles by which mathematical knowledge is contained and transferred. One student commented that social processes play into proofs and that proofs may be written differently depending on the anticipated reader of the proof, which echoes Balacheff, who suggests that proving is very much a cultural artifact and defines proof as an explanation accepted by a given community at a given time (as quoted in Weber, 2003, p. 1).

“To create, using some consistent set of accepted axioms, an undeniable argument of some sort, which is readily interpreted by the intended audience.”

**Understanding**

Many students (8 on the pre-assessment, 11 on the post-assessment) spoke about proof as a tool for understanding concepts in mathematics. Although this category is similar to the explanation category, the main difference is that quotes in this category speak to a more holistic sense of what mathematics is about, whereas the explanation category, as defined by Hanna (1990), refers to insights into why a particular theorem or formula is valid. However, these themes are related, since we often need explanations for understanding, and some quotes were classified under both themes. This was a somewhat broad category and incorporated different kinds of understanding. For instance, several quotes spoke to a belief that examining a proof helped one to understand the development of concepts in mathematics.
“The purpose is to understand better [where] a formula or concept came from...”

“To understand where concepts in mathematics are derived from.”

“...to understand how theorems are developed.”

While other quotes spoke to a pedagogical function of proof; that is, it helps students to understand material in mathematics on a deeper level.

“...to make students understand deeper about the concepts they have learned before.”

“Even if the concepts are initially "believable", the practice of proofs helps in understanding the harder concepts later on.”

“...to let students understand the mathematical material better.”

Further quotes focused on how constructing proofs helps one embody mathematics and to understand the big picture around how mathematics works.

“To understand the logic of how ideas are constructed and formed for a better understanding into the dynamics of mathematics as a whole.”

“The purpose of proof is to truly understand the process, from start to finish, of solving problems.”

“To [foster] understanding and break away from the notion that math is about numbers.”

Building Mathematics

Although this theme is similar to axiomatizing, there is a distinction between these themes. The comments classified under this theme refer to the hierarchical nature of mathematics,
but do not specifically speak about axioms and definitions as the base level of mathematics. The comments in this theme refer to the function of proof as providing building blocks for understanding mathematics.

“For this it is important that they be formally proved using results that [were] proved earlier.”

“You have to know the basics before moving on to more complex systems.”

“To derive conclusions or further information from earlier information which has been assumed or already proven.”

“To prove there are connections among concepts.”

Logical Skills

Several students (4 on the pre-assessment, 5 on the post-assessment) responded that proof was intricately related to logic. These students identified sound logical reasoning as a necessary skill in proof construction and logic as a foundation of mathematics.

“To establish logical truth that helps construct reasoning to equations, formulas, and life itself.”

“To understand the logic of how ideas are constructed.”

“The purpose of proofs in mathematics is to not only assess the truth value of a statement (theorem, conjecture, etc.), but also to provide a sound mathematical logic by which the statement is assessed.”
Thinking Skills

Some of the students (3 on the pre-assessment, 1 on the post-assessment) spoke about the specialized thinking skills that are required when constructing a proof and recognized that a function of proof reading and proof writing is to cultivate these skills in students. Even though some students did not identify the exact nature of these skills, they still commented that something was unique about the cognition necessary to successfully write a proof.

“Also, to be forced to think more critically and use a part of our mind that isn’t always required in math classes.”

“I believe the purpose of proof in mathematics is to sharpen our problem solving skills.”

“To be able to develop a plan for understanding, breaking down and solving mathematical problems.”

As can be seen from these descriptions, several of these themes present proof as having a pedagogical function. These would be the functions of illustrating techniques, understanding, thinking skills and logical skills. Each time one of these codes was applied, the student’s comments related to their interpretation of why they were learning proof. As we will discuss in the next sections, students from the treatment and comparison group had responses that fell under differing theme headings and the comparison group was more likely to identify these pedagogical functions of proof than the treatment group.
5.1.2.2 Treatment Group Qualitative Results

From analysis of the written responses from the pre and post-assessments of students in the treatment group, ten themes were present. Seven of these themes were consistent with the nine functions of proof from the literature: verification, explanation, illustration of techniques, axiomatization, discovery, intellectual challenge, and communication. Table IX shows the frequency of comments under each theme, separated by comments on the pre and post-assessment. The functions of proof for justification for a definition was not identified by any students in the treatment group. This is not surprising as proofs that constructed or used complicated definitions did not appear much during the collaborative revision course or the comparison course. Additionally, proof as a way of providing autonomy to students was also not identified by any participant in this study. Again this is perhaps expected since students may not yet see themselves as ‘provers’ after taking only one course on proof.

Consistent with other studies (Almeida, 1995; Knuth, 2002a; Coe and Ruthven, 1994) on undergraduate students and in-service secondary teachers, verification does comprise a large number of the comments. As shown in Table IX, comments about the verification role of proof constitute over 40% of the total comments in the treatment group suggesting that it is a primary component of this group’s beliefs about proof function. On the post-assessment, there was a slight decrease in the identification of verification as a function of proof, with it making up only about 30% of the comments.

However, unlike the previous studies mentioned above, these results show that students in the treatment group have a robust and varied understanding of the functions of proof.
An additional 40% of the students' comments addressed one of the proof functions from the literature that was not verification. Students identified functions of proof on the pre-assessment, such as axiomatizing and discovery, that are at a high level and inconsistent with the hypotheses of other studies (Weber, 2002; Moore, 1994) that students in a transition to proof course may only understand the verification and explanation functions of proof. Also, the average number of codes applied to a students' response was two, which shows that students were able to recognize that proof could serve multiple functions in mathematics.

The themes consistent with a pedagogical conceptualization of proof were illustrating techniques, understanding, thinking skills and logical skills. Only about a quarter of the comments on both the pre and post-assessment referred to one of these pedagogical categories. Thus,
students in the treatment group felt that proof had a function that was distinctly for learning purposes. Indeed there was not much of a shift between frequency of comments in each category from the pre to the post-assessment; the students’ responses covered mostly the same themes on both assessments.

<table>
<thead>
<tr>
<th>Student</th>
<th>Codes on Pre-Assessment</th>
<th>Codes on Post-Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>—</td>
<td>E, U, L</td>
</tr>
<tr>
<td>2</td>
<td>V, D, IT</td>
<td>V, U</td>
</tr>
<tr>
<td>3</td>
<td>V, V</td>
<td>E, V</td>
</tr>
<tr>
<td>4</td>
<td>E, V</td>
<td>U</td>
</tr>
<tr>
<td>5</td>
<td>B</td>
<td>B, D</td>
</tr>
<tr>
<td>6</td>
<td>E, V, L</td>
<td>A</td>
</tr>
<tr>
<td>7</td>
<td>E</td>
<td>CO, A</td>
</tr>
<tr>
<td>8</td>
<td>V</td>
<td>V, B</td>
</tr>
<tr>
<td>9</td>
<td>V, D</td>
<td>V</td>
</tr>
<tr>
<td>10</td>
<td>V, IT, C, A, IT, V</td>
<td>V, L</td>
</tr>
<tr>
<td>11</td>
<td>A</td>
<td>V</td>
</tr>
<tr>
<td>12</td>
<td>V</td>
<td>V, B</td>
</tr>
<tr>
<td>13</td>
<td>V, U</td>
<td>E, D, U</td>
</tr>
<tr>
<td>14</td>
<td>V</td>
<td>V</td>
</tr>
<tr>
<td>15</td>
<td>L, U</td>
<td>L, U</td>
</tr>
</tbody>
</table>

**TABLE X**

**CODES APPLIED FOR EACH STUDENT IN THE TREATMENT GROUP ON THE PRE AND POST-ASSESSMENT**

Table X shows the codes applied to each student’s response. Note that all but two students changed or added a theme in their response from the pre to the post-assessment. Also, if a
student identified verification as a function of proof on the pre-assessment, they were much more likely to also identify verification on the post-assessment, with seven of the ten students who identified verification as a function of proof on the pre-assessment also naming it on the post-assessment.

5.1.2.3 **Comparison Group Qualitative Results**

<table>
<thead>
<tr>
<th>Theme</th>
<th>Pre-Assessment</th>
<th>Post-Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Understanding</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Discovery</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>Building Mathematics</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Verification</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Illustrating Techniques</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Thinking Skills</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Logical Skills</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Intellectual Challenge</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Explanation</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Communication</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Axiomatization</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>26</strong></td>
<td><strong>22</strong></td>
</tr>
</tbody>
</table>

**TABLE XI**

FREQUENCY OF COMMENTS ON BOTH ASSESSMENTS UNDER EACH PROOF FUNCTION THEME FOR THE COMPARISON GROUP

The comparison group identified ten themes, including six of the themes from the literature: verification, explanation, illustration of techniques, discovery, intellectual challenge, and communication. The frequency of comments under each theme on the pre and post-assessment
is shown in Table XI. Again we see that none of these students identified the proof functions of justification of a definition or providing autonomy, similar to the treatment group. However, none of the students in the comparison group spoke of proof for axiomatization on either assessment, which contrasts with the comments from the treatment group about this proof function on both the pre and post-assessment.

This group did not identify proof for verification the most, but instead chose proof for understanding the most on both assessments. As can be seen in Table XI, students in the comparison group did not identify axiomatization, communication or explanation as a function of proof on the pre-assessment. Rather, these students heavily referred to the pedagogical functions of proof that we classified earlier: illustrating techniques, understanding, thinking skills and logical skills. The majority of the comments were in these four categories comprising about 54% of the total comments on the pre-assessment. This persisted on the pre-assessment with 50% of the comments in one of these pedagogical categories. Overall, there was not much of a shift between categories identified in the student responses from the pre to post-assessment.

A summary of the codes applied to each comparison group student’s response is given in Table XII. All of the students in this group changed or added a theme in their response from the pre to the post-assessment. Unlike the treatment group, there is little correlation between a given students response on the pre-assessment and on the post-assessment. The students in this group also identified around two functions of proof, on average, on each assessment. This again suggests that students in a transition to proof course acknowledge that proof can serve several different roles in mathematics.
TABLE XII

CODES APPLIED FOR EACH STUDENT IN THE COMPARISON GROUP ON THE PRE AND POST-ASSESSMENT

5.1.2.4 Between Groups Qualitative Results

We now contrast the qualitative results between the treatment and comparison groups. As previously stated, we consider several of the proof function themes to have a more pedagogical nature than the others. These themes are illustrating techniques, understanding, thinking skills and logic skills and each one refers to proof as a teaching and learning tool of some sort. Thus, we first inspect the themes identified by students in each group that fall into these pedagogical themes. Then we discuss parallels in the average number of proof functions identified by students in each group and finally we give a comprehensive view of what these results tell us about students’ views on proof function.
If themes are separated into pedagogical (illustrating techniques, understanding, thinking skills and logic skills) and non-pedagogical (verification, explanation, discovery, building mathematics, axiomatizing, intellectual challenge and communication) categories, as shown in Table XIII, then we can gain an improved overview of the types of comments in each group. Notice that on the pre-assessment, students in the comparison group had twice as many comments of a pedagogical nature than students in the treatment group. Although the gap was narrowed slightly on the post-assessment, the treatment group still had 70% of the comments in one of the non-pedagogical categories. This suggests that students in the treatment group recognized more of the functions of proof that from the literature (verification, explanation, discovery, building mathematics, axiomatizing, intellectual challenge and communication) than students in the comparison group and these beliefs persisted to the end of the semester. Students in both groups identified, on average, two proof functions per student on each of the pre and post-assessment, inferring that both the treatment and comparison group acknowledged that proof can serve several purposes in mathematics. As we will discuss in Section 5.1.3.3,

<table>
<thead>
<tr>
<th></th>
<th>Group</th>
<th>Pre-Assessment</th>
<th>Post-Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pedagogical Functions</strong></td>
<td>Treatment</td>
<td>7 (26%)</td>
<td>8 (30%)</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>14 (54%)</td>
<td>11 (50%)</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td>Treatment</td>
<td>20 (74%)</td>
<td>19 (70%)</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>12 (46%)</td>
<td>11 (50%)</td>
</tr>
</tbody>
</table>

**TABLE XIII**

FREQUENCY OF COMMENTS FOR EACH PROOF FUNCTION TYPE ON THE EACH ASSESSMENT
there were no significant differences in the average number of proof functions identified between groups on either the pre or post-assessment.

These results also show that many students do appreciate that proof has functions beyond that of verification. This contradicts the hypothesis of Weber (2002), who claimed that students may not appreciate proof for anything other than verification and explanation. This study found that students in both the treatment and comparison group readily identified other uses of proof; even some high-level functions of proof that are not illustrated in a transition to proof course, such as proof for axiomatization.

5.1.3 Quantitative Results

In this section a report of the results from quantitative analysis of the Likert scale ratings from the assessment is given. The eight statements that students were to rate from 1, indicating strong disagreement, to 5, indicating strong agreement, are shown in Table XIV. For ease of reporting, the topic of each statement is identified and we will hence refer to that statement by its topic (however, order of statements will be kept the same in all reporting).

Since there were two assessments given (pre and post) and two groups (treatment and comparison) we present the results for a comparisons within groups and between groups. The presentation of this section is results from the treatment group from the pre and post-assessment followed by results from the comparison group. Then we report on similarities and differences between results in both groups.
Statement | Topic
---|---
1. I believe that proofs are very important in the field of mathematics. | Importance
2. I believe that proofs are only done to verify whether a given statement is true or false. | Verification
3. A proof can explain why a given statement is true. | Explanation
4. I enjoy writing proofs because they are intellectually challenging. | Challenge
5. I believe that proofs are an important communication tool in mathematics. | Communication
6. I believe that proving can be used to discover new mathematics. | Discovery
7. Proofs are important in organizing mathematical knowledge. | Axiomatizing
8. I want to learn proof construction so I can understand the language of mathematics and create my own proofs. | Autonomy

**TABLE XIV**

MEANS AND STANDARD DEVIATIONS OF EACH LIKERT ITEM ON THE PRE-ASSESSMENT FOR ALL STUDENTS

5.1.3.1 **Treatment Group Quantitative Results**

Statistical tests were performed to compare students’ responses between the pre and post-assessment to gain more specific information to determine how (or if) appreciation of proof functions shifted during the semester for students in the treatment group. The results of the before and after ratings for each statement are presented and discussed to answer our first research question about the possible impacts of collaborative revision on proof function beliefs. Then, the results from the calculation of the total function score (TFS) are presented and interpreted.

First, we wished to determine which beliefs of students in the treatment group about proof function stayed consistent from the pre to the post-assessment and which beliefs about proof
function were fluid. To do this, paired-samples $t$-tests were conducted on the mean Likert ratings from each assessment and the results of these tests are summarized in Table XV. Note that students in the treatment group rated every proof function except axiomatizing and verification higher (closer to 5) on the post-assessment than on the pre-assessment indicating a greater appreciation for these varied proof functions after participation in the collaborative revision course. As mentioned earlier, the verification statement was written negatively meaning that a

<table>
<thead>
<tr>
<th>Statement Topic</th>
<th>Pre-Assessment</th>
<th>Post-Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Importance</strong></td>
<td>15 4.73 0.59</td>
<td>15 5.00 0.00</td>
</tr>
<tr>
<td><strong>Verification</strong></td>
<td>15 3.07 1.44</td>
<td>14 2.29 1.38</td>
</tr>
<tr>
<td><strong>Explanation</strong></td>
<td>15 4.40 1.12</td>
<td>15 5.00* 0.00</td>
</tr>
<tr>
<td><strong>Challenge</strong></td>
<td>15 3.87 0.92</td>
<td>15 4.33 0.82</td>
</tr>
<tr>
<td><strong>Communication</strong></td>
<td>15 4.40 0.91</td>
<td>15 4.73 0.59</td>
</tr>
<tr>
<td><strong>Discovery</strong></td>
<td>15 4.53 1.13</td>
<td>15 4.87 0.52</td>
</tr>
<tr>
<td><strong>Axiomatizing</strong></td>
<td>15 4.67 0.62</td>
<td>15 4.53 0.74</td>
</tr>
<tr>
<td><strong>Autonomy</strong></td>
<td>15 4.40 0.91</td>
<td>15 4.67 0.62</td>
</tr>
</tbody>
</table>

* The mean difference is significant ($p = 0.05$) at the $\alpha = 0.05$ level.

**TABLE XV**

MEANS AND STANDARD DEVIATIONS OF LIKERT ITEM RESPONSES ON THE PRE AND POST-ASSESSMENTS OF THE TREATMENT GROUP
higher appreciation of proof functions corresponds with a rating closer to 1, which was the trend from the pre to post-assessment. Though, only one of these mean differences was statistically significant $\alpha = 0.05$ level; the mean response ($M = 5.00, SD = 0.00$) to the explanation statement on the post-assessment was significantly greater ($t(14) = -2.07, p = 0.05$) than the mean response ($M = 4.40, SD = 1.12$) on the pre-assessment. The standardized effect size index, $d$, for the ratings on this statement was 0.54. This indicates that students in the treatment course increased their appreciation of the explanation function of proof throughout the semester. A possible reason for this is that students were often explaining their proofs and arguments while taking part in the collaborative revision process and they may have come to see proof as a tool for understanding why a statement is true. Another possible explanation is that the students in the treatment group were exposed to their classmates proofs and not the proofs of an instructor or textbook. It is plausible that the students may have been writing more explanatory proofs since those may be the ones that made the most sense while they were constructing their proofs.

The mean differences for each statement were computed by taking the mean for that statement on the pre-assessment and subtracting the mean on the post-assessment. Thus, a negative mean difference indicates a higher average rating of a statement on the post-assessment than on the pre-assessment. As discussed above, the mean difference for the statement regarding the explanation function of proof was significant with $p = 0.05$. Correlation coefficients were also computed from the paired-samples data for the pre and post-assessment and are shown in Table XVI. Calculation of the correlation coefficients for each question was desired to determine
TABLE XVI

MEAN DIFFERENCES AND CORRELATION COEFFICIENTS FOR PRE AND POST-ASSESSMENT RESPONSES FOR THE TREATMENT GROUP

<table>
<thead>
<tr>
<th>Question Topic</th>
<th>N</th>
<th>Mean Diff.</th>
<th>Corr. (r)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td>15</td>
<td>-0.27</td>
<td>—</td>
</tr>
<tr>
<td>Verification</td>
<td>14</td>
<td>0.78</td>
<td>0.25</td>
</tr>
<tr>
<td>Explanation</td>
<td>15</td>
<td>-0.60*</td>
<td>—</td>
</tr>
<tr>
<td>Challenge</td>
<td>15</td>
<td>-0.46</td>
<td>0.45</td>
</tr>
<tr>
<td>Communication</td>
<td>15</td>
<td>-0.33</td>
<td>0.34</td>
</tr>
<tr>
<td>Discovery</td>
<td>15</td>
<td>-0.34</td>
<td>0.87**</td>
</tr>
<tr>
<td>Axiomatizing</td>
<td>15</td>
<td>0.14</td>
<td>0.26</td>
</tr>
<tr>
<td>Autonomy</td>
<td>15</td>
<td>-0.27</td>
<td>0.38</td>
</tr>
</tbody>
</table>

* The mean difference is significant ($p = 0.05$) at the $\alpha = 0.05$ level.
** The correlation is significant ($p < 0.001$) at the $\alpha = 0.05$ level.

if the students beliefs were changing during the semester or remaining static. The correlations for the questions on importance and explanation are not given because the post-assessment standard deviation on both questions was zero, making the correlation impossible to compute from the standard formulas. None of the correlations were significant except for the discovery proof function statement ($r = 0.87, p < 0.001$). This implies that treatment group students’ beliefs about proof function were quite fluid and their agreement with each statement increased during the semester.

As an overall measure of proof function appreciation, the TFS was computed for each student in the treatment group by summing their rating for each statement. This TFS, an
acronym for Total Function Score, was necessary to measure a student’s overall appreciation of the proof functions and was computed by summing a student’s Likert scale ratings for each statement. The means and standard deviations of the TFSs on the pre and post-assessment are shown in Table XVII. Two types of tests were performed on this TFS rating to gain information about these students proof function beliefs. First, one-sample \( t \)-tests were executed to determine if the mean TFS on the pre and post-assessments for students in the treatment group was significantly different from a TFS of 24, which corresponds to a neutral rating of 3 on each of the eight statements. The result of these tests showed that the TFS mean on the pre-assessment of 33.93 (\( SD = 5.18 \)) was significantly different from 24 (\( t(14)=7.43, p < 0.001 \)) and the TFS mean on the post-assessment of 36.60 (\( SD = 2.50 \)) was also significantly different from 24 (\( t(14)=19.51, p < 0.001 \)). This indicates that even before completing an introduction to proof course students do have a robust and varied appreciation of the many functions of proof in mathematics. A paired-samples \( t \)-test was also employed to determine if the mean

**TABLE XVII**

MEAN AND STANDARD DEVIATION OF THE TOTAL FUNCTION SCORE (TFS) FOR THE TREATMENT GROUP ON EACH ASSESSMENT

<table>
<thead>
<tr>
<th></th>
<th>( N )</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Assessment</td>
<td>15</td>
<td>33.93</td>
<td>5.18</td>
</tr>
<tr>
<td>Post-Assessment</td>
<td>15</td>
<td>36.60</td>
<td>2.50</td>
</tr>
</tbody>
</table>

* The mean difference is significant (\( p = 0.03 \)) at the \( \alpha = 0.05 \) level.
TFSs on the pre and post-assessment were significantly different from each other. It was found that the TFS mean on the post-assessment was significantly higher than the mean on the pre-assessment \((t(14)=-2.41, p = 0.03)\). Thus students in the treatment group gained a more robust understanding of proof function after participation in the collaborative revision process.

5.1.3.2 Comparison Group Quantitative Results

The same characteristics of proof function understanding were calculated for the comparison group as the ones described above for the treatment group. This section begins with a presentation of comparison group means for ratings on all of the assessment statements. Then, results from tests of mean difference and correlation are discussed. Finally, we display results regarding the mean TFSs for students in this group.

The means and standard deviations for ratings on all of the assessment statements are summarized in Table XVIII. The comparison group, on average, agreed less with all of the functions of proof except discovery on the post-assessment than on the pre-assessment. To determine if these mean differences were significant, paired-samples \(t\)-tests were performed on the means for each of the statements. All of the mean differences were not found to be significant \((p > 0.5)\) at the \(\alpha = 0.05\) level, except for the statement regarding the communication role of proof. The mean rating \((M = 4.58, SD = 0.90)\) to the statement about the communication function of proof on the post-assessment was significantly less \((t(11) = 2.57, p = 0.03)\) than the mean response \((M = 4.08, SD = 1.00)\) on the pre-assessment. The standardized effect size index, \(d\), for this question was 0.74. This finding is consistent with the hypothesis of Hanna and Barbeau (2011) and deVilliers (1990) that the traditional instruction environment in proof
<table>
<thead>
<tr>
<th>Statement Topic</th>
<th>Pre-Assessment</th>
<th>Post-Assessment</th>
<th>$N$</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Importance</strong></td>
<td></td>
<td></td>
<td>12</td>
<td>4.67</td>
<td>0.49</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12</td>
<td>4.67</td>
<td>0.78</td>
</tr>
<tr>
<td><strong>Verification</strong></td>
<td>Pre-Assessment</td>
<td>2.64</td>
<td>11</td>
<td>0.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>2.82</td>
<td>11</td>
<td>1.33</td>
<td></td>
</tr>
<tr>
<td><strong>Explanation</strong></td>
<td>Pre-Assessment</td>
<td>4.67</td>
<td>12</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>4.67</td>
<td>12</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td><strong>Challenge</strong></td>
<td>Pre-Assessment</td>
<td>3.75</td>
<td>12</td>
<td>0.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>3.58</td>
<td>12</td>
<td>1.08</td>
<td></td>
</tr>
<tr>
<td><strong>Communication</strong></td>
<td>Pre-Assessment</td>
<td>4.58</td>
<td>12</td>
<td>0.90</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>4.08*</td>
<td>12</td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td><strong>Discovery</strong></td>
<td>Pre-Assessment</td>
<td>4.42</td>
<td>12</td>
<td>0.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>4.45</td>
<td>12</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td><strong>Axiomatizing</strong></td>
<td>Pre-Assessment</td>
<td>4.58</td>
<td>12</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>4.50</td>
<td>12</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td><strong>Autonomy</strong></td>
<td>Pre-Assessment</td>
<td>4.00</td>
<td>12</td>
<td>1.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>4.00</td>
<td>12</td>
<td>1.04</td>
<td></td>
</tr>
</tbody>
</table>

* The mean difference is significant ($p = 0.03$) at the $\alpha = 0.05$ level.

**TABLE XVIII**

MEANS AND STANDARD DEVIATIONS OF LIKERT ITEM RESPONSES ON THE PRE AND POST-ASSESSMENTS OF THE COMPARISON GROUP

courses obfuscates the communication aspect of proof. Indeed, Hanna and Barbeau (2011) write that “educators have overlooked to a large extent the role of proof as a bearer of mathematical knowledge in the form of methods, tools, strategies and concepts that are new to the student and add to the approaches the student can bring to bear in other mathematical contexts” (pg. 98). Thus, it is perhaps to be expected that students in this comparison group did not possess a
<table>
<thead>
<tr>
<th>Question Topic</th>
<th>N</th>
<th>Mean</th>
<th>Corr.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Diff.</td>
</tr>
<tr>
<td>Importance</td>
<td>12</td>
<td>0.00</td>
<td>0.63*</td>
</tr>
<tr>
<td>Verification</td>
<td>11</td>
<td>-0.18</td>
<td>0.51</td>
</tr>
<tr>
<td>Explanation</td>
<td>12</td>
<td>0.00</td>
<td>0.25</td>
</tr>
<tr>
<td>Challenge</td>
<td>12</td>
<td>0.17</td>
<td>0.41</td>
</tr>
<tr>
<td>Communication</td>
<td>12</td>
<td>0.50*</td>
<td>0.75**</td>
</tr>
<tr>
<td>Discovery</td>
<td>11</td>
<td>-0.03</td>
<td>0.88***</td>
</tr>
<tr>
<td>Axiomatizing</td>
<td>12</td>
<td>0.08</td>
<td>0.88***</td>
</tr>
<tr>
<td>Autonomy</td>
<td>12</td>
<td>0.00</td>
<td>0.67*</td>
</tr>
</tbody>
</table>

* The mean difference is significant ($p = 0.03$) at the $\alpha = 0.05$ level.

* The correlation is significant ($p = 0.03$) at the $\alpha = 0.05$ level.

** The correlation is significant ($p = 0.005$) at the $\alpha = 0.05$ level.

*** The correlation is significant ($p < 0.001$) at the $\alpha = 0.05$ level.

TABLE XIX

MEAN DIFFERENCES AND CORRELATION COEFFICIENTS FOR PRE AND POST-ASSESSMENT RESPONSES FOR THE COMPARISON GROUP

strong appreciation for this proof function and their ratings for this statement decreased during the semester.

Table XIX shows mean differences (computed as the mean on the post-assessment subtracted from the mean on the pre-assessment) and correlation coefficients for the comparison group’s Likert ratings on each statement. Unlike the results from the treatment group, numerous statements had correlation coefficients between ratings on the pre and post-assessment that were found to be significant. High correlations were found for the statements regarding importance of proof and the proof functions of communication and autonomy. The correlation
coefficients for the statements about the proof functions of discovery and axiomatizing were very high and significant \( (p < 0.001) \), indicating a strong relationship between students’ ratings on the pre-assessment and the post-assessment. These correlation results suggest that students in the comparison group had rigid beliefs about most of the proof functions.

<table>
<thead>
<tr>
<th>Comparison Group</th>
<th>N</th>
<th>Mean (^a)</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Assessment</td>
<td>12</td>
<td>33.75</td>
<td>3.11</td>
</tr>
<tr>
<td>Post-Assessment</td>
<td>12</td>
<td>32.92</td>
<td>5.43</td>
</tr>
</tbody>
</table>

\(^a\) The mean difference is not significant \( (p > 0.4) \) at the \( \alpha = 0.05 \) level.

TABLE XX

MEAN AND STANDARD DEVIATION OF THE TOTAL FUNCTION SCORE (TFS) FOR THE COMPARISON GROUP ON EACH ASSESSMENT

Additional evidence that comparison group students’ beliefs about proof remained static can be seen in the results of the mean TFS, which represents a measure of students’ overall proof function appreciation. The means and standard deviations of the TFSs on the pre and post-assessment are shown in Table XX. It can be seen that the mean TFS for this group did not significantly increase \( (t(11) = 0.77, p = 0.46) \) from the pre-assessment to the post-assessment (in fact, it slightly decreased). The results of the one-sample \( t \)-tests for this group do show that the TFS mean on both the pre-assessment \( (M=33.75, SD = 3.11) \) and on the post-assessment
(M = 32.92, SD = 5.43) were significantly different from the all-neutral score of 24 ($t(11) = 10.87$, $p < 0.001$; $t(11) = 5.68$, $p < 0.001$). However, contrary to the treatment group, a paired-samples $t$-test found that the TFS means on the pre and post-assessment were not significantly different ($t(11) = 0.77, p = 0.45$) from each other. Thus, students in the comparison group do recognize some of the proof functions, but these beliefs remain largely unchanged after a transition to proof course.

5.1.3.3 Between Groups Quantitative Results

To address the research question about any impacts that collaborative revision may have on students’ beliefs about proof function, a contrast between the treatment and comparison groups is provided here. First, we provide statistics on the number of proof functions identified on the pre and post-assessments for students in both the treatment and comparison group. Then, the Likert rating means presented above are compared between groups to determine if differences are present in the proof functions appreciated by students in each group. Finally, we offer a comparison of the total function scores in each group and interpret these results to answer this research question.

To determine if students in either group, on average, recognized several different functions of proof an analysis of codes applied to students’ comments to the open response question on the pre and post-assessment was undertaken. The qualitative analysis results regarding the nature of these responses can be found in Section 5.1.2, however we present the quantitative analysis
<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-Assessment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>14</td>
<td>1.86</td>
<td>1.17</td>
</tr>
<tr>
<td>Comparison</td>
<td>12</td>
<td>2.17</td>
<td>0.94</td>
</tr>
<tr>
<td><strong>Post-Assessment</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>14</td>
<td>1.79</td>
<td>0.70</td>
</tr>
<tr>
<td>Comparison</td>
<td>12</td>
<td>1.83</td>
<td>0.94</td>
</tr>
</tbody>
</table>

\(^a\) None of the mean differences are significant \((p > 0.4)\) at the \(\alpha = 0.05\) level.

**TABLE XXI**

AVERAGE NUMBER OF PROOF FUNCTIONS IDENTIFIED PER STUDENT ON EACH ASSESSMENT

Results here. The number of distinct codes per student\(^1\) was calculated and the mean values for each assessment are given in Table XXI. An independent-samples \(t\)-test \((t(24) = -7.37, p > 0.4)\) showed no significant difference between the treatment group mean of 1.86 \((SD = 1.17)\) and the comparison group mean of 2.17 \((SD = 0.94)\) at the \(\alpha = 0.05\) level on the pre-assessment. An independent-samples \(t\)-test \((t(24) = -1.48, p > 0.8)\) also showed no significant difference between the treatment group mean of 1.79 \((SD = 0.70)\) and the comparison group mean of 1.83 \((SD = 0.94)\) at the \(\alpha = 0.05\) level on the post-assessment. Thus, students in both groups identified, on average, two different functions of proof on both the pre and post-assessment. This suggests that most students recognized that proof has multiple functions in mathematics,

\(^1\)Since it was possible to assign the same code more than once to a single student if more than one comment occurred under a theme, we did not double count these codes when tallying the number of codes per student.
but that there is no difference between the number of proof functions recognized by students in the treatment group and the comparison group.

We now turn to the Likert ratings of the eight statements regarding proof function on the pre-assessment. Table XXII shows the means and standard deviations for each of the eight statement ratings on the pre-assessment, separated by group. None of the mean differences were significant between groups and the table shows that the two groups had very similar means for

### Table XXII

<table>
<thead>
<tr>
<th>Statement Topic</th>
<th>$N$</th>
<th>Mean$^a$</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td>Treatment</td>
<td>15</td>
<td>4.73</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>12</td>
<td>4.67</td>
</tr>
<tr>
<td>Verification</td>
<td>Treatment</td>
<td>15</td>
<td>3.07</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>11</td>
<td>2.64</td>
</tr>
<tr>
<td>Explanation</td>
<td>Treatment</td>
<td>15</td>
<td>4.40</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>12</td>
<td>4.67</td>
</tr>
<tr>
<td>Challenge</td>
<td>Treatment</td>
<td>15</td>
<td>3.87</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>12</td>
<td>3.75</td>
</tr>
<tr>
<td>Communication</td>
<td>Treatment</td>
<td>15</td>
<td>4.40</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>12</td>
<td>4.58</td>
</tr>
<tr>
<td>Discovery</td>
<td>Treatment</td>
<td>15</td>
<td>4.53</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>12</td>
<td>4.42</td>
</tr>
<tr>
<td>Axiomatizing</td>
<td>Treatment</td>
<td>15</td>
<td>4.67</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>12</td>
<td>4.58</td>
</tr>
<tr>
<td>Autonomy</td>
<td>Treatment</td>
<td>15</td>
<td>4.40</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>12</td>
<td>4.00</td>
</tr>
</tbody>
</table>

$^a$ None of the mean differences are significant ($p > 0.3$) at the $\alpha = 0.05$ level.

MEANS AND STANDARD DEVIATIONS OF EACH LIKERT ITEM FOR ALL STUDENTS ON THE PRE-ASSESSMENT, SEPARATED BY GROUP
<table>
<thead>
<tr>
<th>Statement Topic</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Importance</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>15</td>
<td>5.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Comparison</td>
<td>12</td>
<td>4.67</td>
<td>0.78</td>
</tr>
<tr>
<td>Verification</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>14</td>
<td>2.29</td>
<td>1.38</td>
</tr>
<tr>
<td>Comparison</td>
<td>12</td>
<td>2.82</td>
<td>1.33</td>
</tr>
<tr>
<td>Explanation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>15</td>
<td>5.00**</td>
<td>0.00</td>
</tr>
<tr>
<td>Comparison</td>
<td>12</td>
<td>4.67</td>
<td>0.49</td>
</tr>
<tr>
<td>Challenge</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>15</td>
<td>4.33*</td>
<td>0.82</td>
</tr>
<tr>
<td>Comparison</td>
<td>12</td>
<td>3.58</td>
<td>1.08</td>
</tr>
<tr>
<td>Communication</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>15</td>
<td>4.73*</td>
<td>0.59</td>
</tr>
<tr>
<td>Comparison</td>
<td>12</td>
<td>4.08</td>
<td>1.00</td>
</tr>
<tr>
<td>Discovery</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>15</td>
<td>4.87</td>
<td>0.52</td>
</tr>
<tr>
<td>Comparison</td>
<td>11</td>
<td>4.45</td>
<td>1.04</td>
</tr>
<tr>
<td>Axiomatizing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>15</td>
<td>4.53</td>
<td>0.74</td>
</tr>
<tr>
<td>Comparison</td>
<td>12</td>
<td>4.50</td>
<td>1.00</td>
</tr>
<tr>
<td>Autonomy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>15</td>
<td>4.67*</td>
<td>0.62</td>
</tr>
<tr>
<td>Comparison</td>
<td>12</td>
<td>4.00</td>
<td>1.04</td>
</tr>
</tbody>
</table>

* The mean difference is significant ($p < 0.05$) at the $\alpha = 0.05$ level.

** The mean difference is significant ($p < 0.025$) at the $\alpha = 0.05$ level.

TABLE XXIII

MEANS AND STANDARD DEVIATIONS OF EACH LIKERT ITEM ON THE POST-ASSESSMENT

each statement on the pre-assessment. This indicates that students in the treatment group had roughly the same conceptions about proof as students in the comparison group at the beginning of the semester.

However, the results show that the beliefs about proof function in the treatment and comparison group diverge on the post-assessment. Table XXIII shows the means and standard deviations of ratings for each of the eight statements on the post-assessments. The treatment
group’s average rating for each statement was closer to 5\(^1\), indicating a higher level of appreciation for all the functions of proof presented on the assessment. We see that students in the treatment course were more likely to agree with several of the proof functions than students in the comparison course. Namely the mean differences of the responses from the students in the treatment course to questions about the intellectual challenge, communication and autonomy roles of proof were all significantly different (\(p < 0.05\)) that those of the comparison group at the \(\alpha = 0.05\) level. Moreover, the mean difference of the responses from the students in the treatment course to the question about the explanation role of proof was significant (\(p < 0.025\)) at the \(\alpha = 0.05\) level. In fact, all students in the treatment course responded that they strongly agreed to the statements about the importance of proof in mathematics and the explanation function of proof. These results indicate that collaborative revision may have an impact on students beliefs about proof; specifically, that students who participate in collaborative revision are more likely to appreciate that proof can be used as an explanation tool. It is maybe not surprising that students in the treatment course strongly responded to proof functioning as a communication tool because verbalizing arguments to one another occurred very frequently during the collaborative revision process. Furthermore, students in the treatment course were working on proofs on their own before bringing them to class, which could have contributed

\(^1\)The aforementioned statement regarding the verification function of proof was written negatively, so a rating closer to 1 indicates that the student believes that proof has more functions that just that of verification.
to students in the treatment course agreeing more with the statement about proof providing autonomy to students.

As mentioned in Section 5.1.1.2, a total function score (TFS) was computed for all students by adding their ratings for each statement on the assessment. Since there were eight statements to be rated from 1 to 5, a student’s TFS score could range from 5 to 40 if all statements were responded to. As illustrated in Table XXIV, the average total function score for students in both groups was quite high (close to 40), showing that students in a transition to proof course do have a high level of appreciation for many different proof functions even at the beginning of the course. A two-tailed, independent samples $t$-test was performed to determine if the mean differences in the TFS total for the pre and post-assessment were significantly different. The

1This was done only after reordering the ratings on statement 2 since it was written negatively, implying that a lower rating indicated a higher appreciation of the many function of proof. Thus, ratings of 1 and 5 were interchanged as were ratings of 2 and 4.

### TABLE XXIV

MEAN AND STANDARD DEVIATION OF TOTAL PROOF FUNCTION SCORES ON EACH ASSESSMENT SEPARATED BY GROUP

<table>
<thead>
<tr>
<th></th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Pre-Assessment</strong></td>
<td>Treatment</td>
<td>15</td>
<td>33.93</td>
<td>5.18</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>12</td>
<td>33.75</td>
<td>3.11</td>
</tr>
<tr>
<td><strong>Post-Assessment</strong></td>
<td>Treatment</td>
<td>15</td>
<td>36.60$^*$</td>
<td>2.50</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>12</td>
<td>32.92</td>
<td>5.43</td>
</tr>
</tbody>
</table>

* The mean difference is significant ($p = 0.03$) at the $\alpha = 0.05$ level.
mean TFS on the pre-assessment for the treatment group of 33.93 ($SD = 5.18$) was not found to be significantly different ($t(25)=0.11$, $p = 0.91$) than the mean TFS for the comparison group of 33.75 ($SD = 3.11$). We conclude that there was no difference at the beginning of the semester between the overall proof functions recognized by the treatment and comparison groups. Though, an independent-samples $t$-test comparing the TFS means between the two groups on the post-assessment found the treatment group mean of 33.60 ($SD = 2.50$) to be significantly different ($t(25)=2.34$, $p = 0.03$) than the comparison group mean of 32.92 ($SD = 5.43$). Thus, these results imply that students’ participation in the collaborative revision process does seem to have an impact on the number of proof functions that they appreciate at the end of a transition to proof course.

5.1.4 Synthesis of Results on Students’ Proof Function Beliefs

The results of the quantitative and qualitative analysis highlight that students in a transition to proof course have a richer appreciation of the diverse functions of proof in mathematics than found in previous studies. Collectively, participants in this study identified seven of the nine functions of proof from the literature that provided a framework for this study when asked an open response question. Furthermore, These results are not consistent with the outcomes of studies by Almeida (1995) who found that undergraduate students did not recognize many functions of proof beyond explanation and verification, or Coe and Ruthven (1994), who found that some undergraduate students did recognize the function of proof in axiomatizing statements, but many cited proof as merely for explanation or convincing purposes. This study also expands on these previous studies, which only examined up to five separate proof functions, by
using nine proof functions identified in the literature as a framework for data analysis. Students in both the treatment and comparison groups in this study were found to have strong agreement with statements addressing functions of proof distinct from verification and wrote about other functions of proof when asked an open response question about proof function. The fact that students in both groups wrote responses that identified, on average, two functions of proof gives affirmation that these students recognize the varied roles of proof in mathematics and don’t see proof as having just one function.

However, these results also provide evidence that collaborative revision does have an impact on students beliefs about proof function. The treatment group rated the explanation function of proof significantly higher on the post-assessment which is most likely due to the explanatory nature of the collaborative revision process. On the other hand, the comparison group did not rate any proof functions significantly higher on the post-assessment, but rated the communication function of proof significantly lower. This is probably due to the learning environment of the comparison classroom, which was lecture-based (see Section 4.5 for a further description of this learning environment). Furthermore, the treatment group rated the proof functions of explanation, intellectual challenge, communication and providing autonomy significantly higher than the comparison group on the post-assessment despite no significant differences on the pre-assessment. Each of these proof functions is implicit in the collaborative revision process; students spend the majority of class time explaining arguments and illustrating why a statement would be true to each other, which relates to the explanation and communication functions of proof, and students work on proofs of statements outside of class, which emphasizes the pro-
viding autonomy and intellectual challenge aspects of proofs. This suggests that collaborative revision may be a way to enhance students’ understanding of some of the functions of proof outlined in the literature.

These results also illustrate whether students beliefs about the function of proof persist or evolve during an introduction to proof course. The qualitative and quantitative results combined show that students in the treatment course left with a greater appreciation of the functions of proof from the literature than students in the comparison course. Additionally, the correlation coefficients for the proof function statements suggest that the treatment group shifted towards a greater agreement with the proof function statements, indicating a more sophisticated appreciation of the functions of proof at the conclusion of the course. The comparison group had high correlations of ratings on many of the statements given about proof function. This indicates comparison group students had an less flexible appreciation of the proof functions and a transition to proof course had little impact on these beliefs. The total function scores confirm this, with no significant difference in the TFS of students in the comparison course from the pre to post-assessment.

5.2 Students’ Proof Validation Skills

The four proof validation tasks on each of the pre and post-assessments were analyzed for evidence of any impact collaborative revision may have on students’ proof validation skills. The assessment given to students is contained in Appendix A. Each proof validation task asked students for a judgement about the validity of the proof, as well a self-report of understanding of the proof and how certain they were of their classification (valid or invalid), each on a three-
point Likert scale. This section first presents an overview of the quantitative data analysis techniques used on data from the pre and post-assessments given to students in the treatment and comparison groups. Then, we discuss results from the treatment group about students’ accuracy on the proof validation tasks from the assessments and students’ reported understanding and confidence in classifying each proof. An analogous report is given for students in the comparison group. The final part of this section reports on the qualitative data analysis of the students that were interviewed about their responses on the proof validation tasks and a summary of the proof schemes held by each student and how the schemes exhibited on the pre-assessment compare to those on the post-assessment.

5.2.1 Quantitative Data Analysis

<table>
<thead>
<tr>
<th>Statement: The product of any three consecutive integers is a multiple of 6.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Proof 1:</strong></td>
</tr>
<tr>
<td>A multiple of 6 must have factors of 3 and 2.</td>
</tr>
<tr>
<td>If you have three consecutive numbers, one will be a multiple of 3 as every third number is 3 times a whole number.</td>
</tr>
<tr>
<td>Also, at least one number will be even and all even numbers are multiples of 2.</td>
</tr>
<tr>
<td>Thus, If you multiply the three consecutive numbers, the answer will have at least one factor of 3 and at least one factor of 2.</td>
</tr>
</tbody>
</table>

Figure 5. The informal proof on the assessment given to students.
Of the four arguments given for students to evaluate on the assessment each employed a different type of argument. As can be seen on the assessment given to students (found in Appendix A), the proofs were numbered as Proof 1, Proof 2, Proof 3 and Proof 4, based on their presented order. Proof 1, which we subsequently refer to as the informal proof, used correct reasoning, but was written in an informal, descriptive style. This informal argument, shown in Figure 5, was included to determine if students are convinced by a proof that uses correct reasoning, but is written colloquially without rigor.

Proof 2, henceforth referred to as the empirical proof, checked several examples, but did not prove anything in general. The empirical proof shown in Figure 6 employs an argument that a student would be convinced by if they held what Harel and Sowder (1998) would term an inductive proof scheme. This proof was included to determine if students are convinced by inductive proofs at the beginning of an introduction to proof course and if this changes during the semester.

\[
\begin{align*}
\text{Statement:}& \quad \text{The product of any three consecutive integers is a multiple of 6.} \\
\text{Proof 2:} & \\
\text{Note that:} & \\
1 \cdot 2 \cdot 3 &= 6, \text{ which is a multiple of 6.} \\
2 \cdot 3 \cdot 4 &= 24, \text{ which is a multiple of 6.} \\
3 \cdot 4 \cdot 5 &= 60, \text{ which is a multiple of 6.} \\
6 \cdot 7 \cdot 8 &= 336, \text{ which is a multiple of 6.} \\
\text{This pattern continues so that statement is true.}
\end{align*}
\]

Figure 6. The empirical proof on the assessment given to students.
Statement: The product of any three consecutive integers is a multiple of 6.

Proof 3:

Let $x$ be any whole number.
The product of three consecutive numbers is $x \cdot (x + 1) \cdot (x + 2) = (x^2 + x) \cdot (x + 2) = x^3 + x^2 + 2x = x^3 + 3x^2 + 2x$.

Looking at the coefficients, we have $1 + 3 + 2 = 6$, so the product is a multiple of 6.

Figure 7. The invalid deductive proof on the assessment given to students

Proof 3 on the assessment, which is subsequently referred to as the invalid deductive proof, had a general, deductive argument, but used an incorrect statement to draw the conclusion. This argument, shown in Figure 7, is deductive and proceeds in general, using $x$ to represent any integer, however it contains an incorrect statement making it invalid. The unsound argument is on the last line, where the coefficients of the polynomial summing to 6 is used as justification that the polynomial is divisible by 6 for all values of $x$. Note that this happens to work here because it is a true statement, but if we consider a modified version of this polynomial, say $x^3 + 3x^2 + x$, then the coefficients sum to 5, but the polynomial is not divisible by 5 for, say, $x = 2$ or $x = 3$. Thus, this argument is not true in general and amounts to merely a coincidence in this case.

Proof 4, henceforth referred to as the valid deductive proof, used a general, deductive argument where all statements were correct. This proof, which is shown in Figure 8, was presented to determine if students could compare it with the invalid deductive proof to make their determination about the validity. Both proofs use a general argument starting with a
Statement: The product of any three consecutive integers is a multiple of 6.

Proof 4:
Of the three consecutive numbers, the first number is either even, so it can be written as $2a$ for some whole number $a$, or odd, so it can be written as $2b - 1$ for some whole number $b$.
If the first number is even, then we have $2a \cdot (2a + 1) \cdot (2a + 2)$, which is even. Then if $a$ is a multiple of 3, we are done. Otherwise, if $a$ is not a multiple of 3, then $2a$ is not a multiple of 3, but either $(2a + 1)$ or $(2a + 2)$ is a multiple of 3 and we’re done.
If the first number is odd, then we have $(2b - 1) \cdot (2b) \cdot (2b + 1)$, which is even. Then if $b$ is a multiple of 3, we are done. Otherwise, if $b$ is not a multiple of 3, then $2b$ is not a multiple of 3, but either $(2b - 1)$ or $(2b + 1)$ is a multiple of 3 and we’re done.

Figure 8. The valid deductive proof on the assessment given to students

variable representing the first of the three consecutive integers, but then proceed in different
directions.

To determine if collaborative revision has any impact on students proof function skills,
data from the from the proof validation tasks on the assessment was analyzed. The proof
validation tasks were quantitatively analyzed in two ways: first, we considered the accuracy
(i.e. percentage of proofs classified correctly as valid or invalid) of students in the treatment and
comparison groups; second, we considered students’ accuracy on the different types of proofs
(i.e. informal, empirical, valid deductive and invalid deductive). Looking at each of the proofs
separately, was done to determine what types of arguments are convincing to students and if
a change is seen from the beginning of the semester to the end for students who participated
in collaborative revision and those who did not. Additionally, students were asked to rate how
well they felt they understood each argument from 1(not at all) to 3 (completely) and how
confident they were that their classification (valid or invalid) was correct from 1 (not at all) to 3 (completely).

For the purposes of the data analysis in this study, we consider the last proof as the only one that is valid. The empirical argument does not constitute a proof under the definition of proof we are using because it is not written in general. The incorrect deductive argument has a flawed statement inside the proof, which also goes against our definition of proof.

The informal proof, although it is a valid argument, does not “proceed from specific and accepted premises” (Hanna, 1990, pg. 8) and thus we do not consider it as a valid proof based on our proof definition. It was deemed that, although this proof presents a valid argument, it is not a valid proof. One of the main goals of an introduction to proof course and a main goal of the collaborative revision course in this study was to aid students in transitioning from informal, colloquial arguments into rigorous proofs. Since this informal argument uses the ideas of divisibility by 2 and divisibility by 3 without actually defining what is meant by these concepts, this ‘proof’ would not be considered to have enough rigor if turned in for a homework assignment. Thus, we consider it invalid to and it is used to test the strength of narrative arguments to convince students and determine if the convincing power is weakened after a transition to proof course or after a collaborative revision course.

Using this criteria for validity, descriptive statistics were computed for the percentages of correct proof classifications on both the pre and post-assessment for students in the treatment and comparison groups. Each student was assigned a percentage calculated by how many proofs they correctly classified on each assessment divided by four, which was the total number
5. I believe that proofs are an important communication tool in mathematics.

6. I believe that proving can be used to discover new mathematics.

7. Proofs are important in organizing mathematical knowledge.

8. I want to learn proof construction so I can create understand the language of mathematics and create my own proofs.

Part II.

Determine if each proof of the statement is valid or invalid by circling your answer. Then rate how well you understand the proof and how certain you are of your answer.

Statement: The product of any three consecutive integers is a multiple of 6.

Proof 1: A multiple of 6 must have factors of 3 and 2. If you have three consecutive numbers, one will be a multiple of 3 as every third number is 3 times a whole number. Also, at least one number will be even and all even numbers are multiples of 2. Thus, if you multiply the three consecutive numbers, the answer will have at least one factor of 3 and at least one factor of 2.

Proof 1: Valid Invalid

How well do you understand this proof? not at all 1 2 3 completely

How certain are you that you correctly classified the proof?
not at all 1 2 3 completely

Figure 9. The questions regarding understanding and confidence given to students on the assessment.

We also wished to determine any changes in students’ understanding of the arguments during the semester. Students were presented with the questions shown in Figure 9 after each of the proofs shown in Figures 5–8 above. The first question posed to students, “How well do you understand this proof?”, was meant to capture their self-reported comprehension of the argument being made. We refer to this as a student’s understanding for the rest of this report. The second question asked of students, “How certain are you that you correctly classified the proof?” was given to gauge a student’s confidence about choosing the proof to be valid or invalid. We refer to this henceforth as a student’s confidence about their classification (valid or invalid).
Descriptive statistics were calculated for the students’ Likert ratings on each question and paired-samples $t$-tests were used to determine if any of the mean differences of the understanding and confidence ratings were significant from the pre-assessment to the post-assessment for the treatment group or the comparison group. Also, information was desired about any differences between the treatment and comparison group, so independent-samples $t$-tests were used to observe significant differences in the mean differences of the understanding and confidence ratings between the two groups (i.e. treatment and comparison).

5.2.2 Quantitative Results

The quantitative results from the treatment group are presented first in this section. Results regarding mean percentages of correct classifications of proofs on the pre and post-assessments, types of arguments that convince students in the treatment group and confidence and understanding ratings for proofs are discussed. Then, analogous results from the comparison group in this study are reported. Finally, a comparison of the two groups is given for each of these topics.

5.2.2.1 Treatment Group Quantitative Results

After assigning a total correct to each student on the pre and post-assessment, the mean percentage correct on each assessment was calculated by dividing the total correct by the total number of proofs, which was four. As mentioned previously, we consider only the fourth proof on the assessment (shown in Appendix A) to be the only valid one. On the pre-assessment, the average percentage of correct proof classifications was $60\%$ for the treatment group. These results are consistent with the studies of Selden and Selden (2003), Alcock and Weber (2005),
and Brandt and Rimmasch (2012) in that students on the pre-assessment classified proofs correctly less than half the time. On the post-assessment, the treatment group did not improve in their classifications and still had an average correct percentage of 60%. It is somewhat surprising that the treatment group did not make gains on the post-assessment since they had a great deal of practice determining the validity of arguments during the collaborative revision process. In each of the studies cited above, the students showed great progress in evaluating valid arguments after discussion with an interviewer or with other students. Perhaps if the students had been able to confer with each other, much like they do during collaborative revision, these percentages would have increased. The correlation coefficient between a student’s correct percentage on the pre-assessment and the correct percentage on the post-assessment was found to be $r = 0.14$ and not found to be significant ($p = 0.60$). Thus, there was a high amount of variation in treatment group students’ classifications of each proof from the pre to the post-assessment.

<table>
<thead>
<tr>
<th>Proof Type</th>
<th>Pre-Assessment</th>
<th>Post-Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informal</td>
<td>5 (33%)</td>
<td>2 (13%)</td>
</tr>
<tr>
<td>Empirical</td>
<td>11 (73%)</td>
<td>13 (87%)</td>
</tr>
<tr>
<td>Invalid Deductive</td>
<td>9 (60%)</td>
<td>7 (47%)</td>
</tr>
<tr>
<td>Valid Deductive</td>
<td>11 (73%)</td>
<td>14 (93%)</td>
</tr>
</tbody>
</table>

**TABLE XXV**

FREQUENCY AND PERCENTAGE OF STUDENTS IN THE TREATMENT GROUP WHO CORRECTLY CLASSIFIED EACH PROOF
Analysis of the proofs separated by proof type gives a more comprehensive account of what types of proofs are convincing to students at the beginning and end of a collaborative revision course. The frequency and percentage of students in the treatment group that classified each type of proof correctly on each assessment is shown in Table XXV. Notice that correct evaluations increased dramatically on the the empirical and valid deductive proofs on the post-assessment to close to 100% correct. However, correct classifications on the informal proof and the incorrect deductive proof show a decrease in correct classifications. The informal proof, shown in Figure 5, was a valid argument but was not a rigorous proof given to assess student’s level of conviction when faced with proofs written in everyday language instead of formal mathematics. On the pre-assessment only one-third of the students in the study classified this proof as invalid, while on the post-assessment this number decreased drastically, with only two students (13%) classifying this as an invalid proof. This is somewhat perplexing because although the argument is sound, it is not written in the style of the proofs given to students in class or in the textbook. The approach of the textbook was much more rigorous and would have presented a more algebraic proof, much like the valid deductive proof given to students on this assessment. This suggests that students in the treatment group were more likely to be persuaded by a valid argument that lacks appropriate rigor at the end of the course. It is possible that this is due to the way proofs are often presented in the course, with a lot of formalism at the beginning of the semester and slowly tapering off into more informal arguments at the end of the semester as students get more used to constructing proofs. This could lead students to accept more narrative proofs if the arguments presented are sound.
The second ‘proof’ given on the assessment, shown in Figure 6, was an empirical proof meant to determine if students are convinced merely by examples. On the pre-assessment, about a quarter of the students in the treatment course responded that this proof was valid. This is not surprising as empirical arguments can be appealing to students and are similar to the types of arguments they may see in other science courses (Recio and Godino, 2001). However, on the post-assessment only 2 (13%) of the students still considered this a valid argument. Furthermore, one of these two students had also classified this as a valid proof on the pre-assessment. This suggests that this student began the semester with an empirical proof scheme (Harel and Sowder, 1998) and this proof scheme did not evolve during the course. However, it is encouraging that very few students identified this as a valid proof on the post-assessment, as the realization that a proof must take the form of general statements is one of the skills that is highly emphasized in the collaborative revision course.

The invalid deductive proof, shown in Figure 7, seemed to be the most troubling for these students. Roughly half the students in the treatment group classified this proof correctly on both the pre and post-assessment. Since there is an incorrect statement in the argument, it is possible that students were confused by that statement and did not know whether it was valid or invalid. More evidence for this hypothesis comes from the interview data (see Section 5.2.3). All the students that were interviewed mentioned the incorrect statement and said that they were unsure about that statement and that if affected the way that they classified the proof. The invalid statement in this proof could have easily be checked by plugging in numbers and
looking at other examples, but there is no evidence that students did this before they made their classification.

Even though only three-quarters of the students in the treatment group correctly classified the valid deductive proof, shown in Figure 8, as valid on the pre-assessment, all but one student correctly classified it on the post-assessment. Additionally, as will be discussed below, these students reported significantly higher confidence of their classifications and understanding of this proof on the post-assessment. Thus, it seems that students in the treatment group are convinced by deductive proofs in general, but are unable to check and evaluate a statement when they are unsure of its validity.

Students were also asked to rate how well they felt they understood each of the four arguments on a three-point Likert scale. The means and standard deviations for the treatment group of these ratings for each proof is shown in Table XXVI. Paired-samples t-tests revealed that the treatment group reported a significant increase ($t(13) = -2.12, p = 0.05$) in their understanding of the informal proof between the two assessments. The treatment group also reported significantly greater ($t(13) = -2.48, p = 0.03$) understanding of the valid deductive proof on the post-assessment compared to the pre-assessment. All of these students gave a rating of 3, indicating complete understanding, to the empirical proof on the post-assessment. Thus, the treatment group made gains in their reported understanding of the arguments given on the assessment during the semester. However, not much improvement was made in understanding the invalid deductive proof and, as discussed above, the percentage of correct classifications for this proof actually decreased on the post-assessment.
Students in the treatment group made even greater gains on their self-reported confidence about their classifications. The assessment required students to rate on a three-point Likert scale how certain they were that they correctly classified each argument. The means and standard deviations of the ratings from the treatment group can be found in Table XXVII. Paired-samples $t$-tests found significant increases in means on the post-assessment in reported confidence on the informal, invalid deductive, and valid deductive arguments. Additionally, the mean on the empirical was much higher, but the difference was not found to be significant. This is perplexing since the mean percentage of correct classifications did not increase at all, yet reported confidence and understanding increased significantly for most of the proofs.
<table>
<thead>
<tr>
<th>Proof Type</th>
<th>Treatment Group Confidence</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Informal</td>
<td>Pre-Assessment</td>
<td>14</td>
<td>2.36</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>14</td>
<td>2.79*</td>
<td>0.43</td>
</tr>
<tr>
<td>Empirical</td>
<td>Pre-Assessment</td>
<td>15</td>
<td>2.53</td>
<td>0.74</td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>15</td>
<td>2.93</td>
<td>0.26</td>
</tr>
<tr>
<td>Invalid Deductive</td>
<td>Pre-Assessment</td>
<td>15</td>
<td>2.20</td>
<td>0.68</td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>15</td>
<td>2.60**</td>
<td>0.51</td>
</tr>
<tr>
<td>Valid Deductive</td>
<td>Pre-Assessment</td>
<td>14</td>
<td>2.14</td>
<td>0.66</td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>14</td>
<td>2.71***</td>
<td>0.47</td>
</tr>
</tbody>
</table>

* The mean difference is significant ($p = 0.05$) at the $\alpha = 0.05$ level.

** The mean difference is significant ($p = 0.03$) at the $\alpha = 0.05$ level.

*** The mean difference is significant ($p < 0.01$) at the $\alpha = 0.05$ level.

TABLE XXVII

MEANS AND STANDARD DEVIATIONS OF STUDENTS’ SELF-REPORTED CONFIDENCE ABOUT CORRECTLY CLASSIFYING EACH TYPE OF PROOF FOR THE TREATMENT GROUP

5.2.2.2 Comparison Group Quantitative Results

The average percentage of correct proof classifications for the comparison group was 45% on the pre-assessment. This is similar to the Selden and Selden (2003) study, which found that undergraduate students in a transition to proof course correctly judged valid proofs less than half the time. Unlike the treatment group, the comparison group made modest gains on the post-assessment, classifying proofs correctly at a rate of 54%. However, this increase on the post-assessment was not found to be significant ($t(11) = -1.30, p = 0.2$). The correlation coefficient between a student’s correct percentage on the pre-assessment and the correct percentage on
the post-assessment was found to be $r = 0.64$, which is significant ($p = 0.03$). So, while the comparison group did improve slightly in their classifications, there was a high correlation between students’ percentages on each assessment and, thus, students in this group were unlikely to change their classifications between assessments.

Again, an analysis by proof type was done to determine which of the proofs were convincing to students in the comparison group. The comparison group exhibited similar trends as the treatment group on the post-assessment about classifications by proof type as can be seen in Table XXVIII. The number of correct classifications decreased on the informal proof and the invalid deductive proof on the post-assessment, while the number of correct classifications increase on the empirical and valid deductive proofs. One-third of students in the comparison group also classified the informal proof as invalid and on the post-assessment this number decreased, with only one student remarking that it was invalid. Thus, students in the comparison group
group, like those in the treatment group, believe a narrative argument to be valid as long as the reasoning used is correct.

Students in the comparison group increased in the number of correct classifications on the empirical proof, but one student still marked this proof as valid (and one student did not respond). On the pre-assessment, about a quarter of the students in the comparison course responded that this proof was valid. It was again the case that the student who responded that the empirical proof was valid on the post-assessment had also marked it valid on the pre-assessment. The comparison group showed very similar results to the treatment group on the invalid deductive proof. Again, about half the students in the comparison group classified this proof correctly on both the pre and post-assessment. Students in the comparison group also brought up the incorrect statement in this proof in interviews and reported unchanged confidence of classification and understanding of this proof on the post-assessment when on all the other proofs the ratings in both these categories were higher on the post-assessment.

There was a marked increase in the correct classifications of the valid deductive proof for the comparison group; only three students classified it correctly on the pre-assessment, which increased to ten students on the post-assessment. However, while comparison group students reported significantly higher understanding of this proof, there was no significant increase in confidence of their classification on the post-assessment. Therefore, it seems that the comparison group was strongly convinced by symbolic proofs at the end of the course, but also could not carefully check statements in proofs that they were unsure about.
The means and standard deviations of the comparison group’s ratings about understanding for each proof is shown in Table XXIX. Paired-samples $t$-tests were conducted to determine if any of these mean differences were significant. The comparison group did report a significant increase ($t(8) = -2.53, p = 0.04$) in understanding of the valid deductive proof between the two assessments, but all other understanding mean differences were not found to be significant. The means and standard deviations of ratings about confidence on each proof for the comparison group is shown in Table XXX. There were no significant differences between means of students’ reported confidence of classification on any of the proofs. Thus, the comparison group did increase slightly in their average percentage of correct classifications, but at the same time they
Comparison Group Confidence

<table>
<thead>
<tr>
<th>Proof Type</th>
<th></th>
<th>N</th>
<th>Mean(^a)</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informal</td>
<td>Pre-Assessment</td>
<td>12</td>
<td>2.33</td>
<td>0.78</td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>12</td>
<td>2.50</td>
<td>0.52</td>
</tr>
<tr>
<td>Empirical</td>
<td>Pre-Assessment</td>
<td>11</td>
<td>2.82</td>
<td>0.60</td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>11</td>
<td>2.91</td>
<td>0.30</td>
</tr>
<tr>
<td>Invalid Deductive</td>
<td>Pre-Assessment</td>
<td>11</td>
<td>1.82</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>11</td>
<td>1.82</td>
<td>0.60</td>
</tr>
<tr>
<td>Valid Deductive</td>
<td>Pre-Assessment</td>
<td>10</td>
<td>2.10</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>Post-Assessment</td>
<td>10</td>
<td>2.20</td>
<td>0.63</td>
</tr>
</tbody>
</table>

\(^a\) None of the mean differences are significant (\(p > 0.5\)) at the \(\alpha = 0.05\) level.

TABLE XXX

MEANS AND STANDARD DEVIATIONS OF STUDENTS' SELF-REPORTED CONFIDENCE ABOUT CORRECTLY CLASSIFYING EACH TYPE OF PROOF FOR THE COMPARISON GROUP

reported their understanding of the proofs and confidence about classifying the proof largely unchanged.

5.2.2.3 Quantitative Differences Between Groups

Table XXXI shows the means and standard deviations for students in the treatment group and comparison group on the pre and post-assessment, as well as the correlations between students’ correct classification percentage. Looking at the means we see that the treatment group had a higher percentage of correct evaluations on the pre-assessment than the comparison group, but did not improve at all on the post-assessment. A paired-samples \(t\)-test was performed on the mean percentage for each group on the pre and post-assessment. The mean differences
TABLE XXXI

MEANS AND STANDARD DEVIATIONS FOR PRE AND POST-ASSESSMENT RESPONSES FROM BOTH GROUPS

<table>
<thead>
<tr>
<th></th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-Assessment</td>
<td>Treatment</td>
<td>15</td>
<td>0.60</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>12</td>
<td>0.45</td>
<td>0.28</td>
</tr>
<tr>
<td>Post-Assessment</td>
<td>Treatment</td>
<td>15</td>
<td>0.60</td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>Comparison</td>
<td>12</td>
<td>0.54</td>
<td>0.23</td>
</tr>
</tbody>
</table>

between the treatment and comparison groups were not significant on either the pre-assessment or post-assessment at the $\alpha = 0.05$ level. Although the comparison group did improve slightly in their classifications on the post-assessment, this mean was not significantly higher than the pre-assessment mean and was still lower than the mean for the treatment group. When we consider the correlations between percentages on the pre and post-assessment, students in the treatment group have a low correlation of responses, while students in the comparison group have responses that are highly correlated between the assessments. This suggests that students in the comparison group were less likely to change their classifications from the pre-assessment to the post-assessment. The frequency distributions of correct percentages for the pre and post-assessment is shown in Figure 10 for both groups.

The frequency and percentage of students in the treatment and comparison group that classified each proof type correctly is shown in Table XXXII for each of the pre and post-assessments. Independent samples $t$-tests showed that the differences in the correct percentages for each proof type between the groups were not significant for the informal, empirical or invalid
Figure 10. Frequency distributions of correct percentages on the pre-assessment (left) and post-assessment (right).

deductive proof types on either the pre or post-assessment. There was a significant difference ($p = 0.02$) between the correct percentages for the valid deductive proof on the pre-assessment; the treatment group had a significantly higher correct percentage for the valid deductive on the pre-assessment, but the gap between the groups on this proof type disappeared on the post-assessment. Thus, these results imply that students in both the treatment and comparison group were able to identify that empirical proofs were invalid after participation in this transition to proof course. These students were also able to identify a valid deductive proof on the post-assessment. However, it seems that students were unable to test statements in proofs for their validity, such as in the invalid deductive proof. The percentage of correct classifications on this proof type was about the same as guessing and decreased in each group on the post-assessment.
Pre-Assessment

<table>
<thead>
<tr>
<th>Proof Type</th>
<th>Treatment</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informal</td>
<td>5 (33%)</td>
<td>4 (33%)</td>
</tr>
<tr>
<td>Empirical</td>
<td>11 (73%)</td>
<td>8 (67%)</td>
</tr>
<tr>
<td>Invalid Deductive</td>
<td>9 (60%)</td>
<td>7 (58%)</td>
</tr>
<tr>
<td>Valid Deductive</td>
<td>11 (73%)*</td>
<td>3 (25%)</td>
</tr>
</tbody>
</table>

Post-Assessment

<table>
<thead>
<tr>
<th>Proof Type</th>
<th>Treatment</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informal</td>
<td>2 (13%)</td>
<td>1 (9%)</td>
</tr>
<tr>
<td>Empirical</td>
<td>13 (87%)</td>
<td>10 (83%)</td>
</tr>
<tr>
<td>Invalid Deductive</td>
<td>7 (47%)</td>
<td>5 (42%)</td>
</tr>
<tr>
<td>Valid Deductive</td>
<td>14 (93%)</td>
<td>10 (83%)</td>
</tr>
</tbody>
</table>

* The percentage difference is significant ($p = 0.02$) at the $\alpha = 0.05$ level.

TABLE XXXII

FREQUENCY AND PERCENTAGE OF STUDENTS IN BOTH GROUPS WHO CORRECTLY CLASSIFIED EACH PROOF TYPE

from the pre-assessment. Additionally, students in both groups were more convinced by a proof written informally after participation in the course. This may be due to the progression of how proofs are presented in the comparison course; the proofs at the beginning of the course were very rigorous in order to teach students about proof construction and slowly became less rigorous toward the end of the course as students became more comfortable with the proving process. Students that participated in collaborative revision may have also come to appreciate narrative proofs at the same level as rigorous proofs since the collaborative revision process tended to emphasize statements that promoted understanding over symbolic statements.
<table>
<thead>
<tr>
<th>Proof Type</th>
<th>Pre-Assessment Understanding</th>
<th>Post-Assessment Understanding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>Mean</td>
</tr>
<tr>
<td>Informal</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>15</td>
<td>2.53</td>
</tr>
<tr>
<td>Comparison</td>
<td>12</td>
<td>2.17</td>
</tr>
<tr>
<td>Empirical</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>15</td>
<td>2.73</td>
</tr>
<tr>
<td>Comparison</td>
<td>11</td>
<td>2.64</td>
</tr>
<tr>
<td>Invalid Deductive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>15</td>
<td>2.40</td>
</tr>
<tr>
<td>Comparison</td>
<td>11</td>
<td>2.00</td>
</tr>
<tr>
<td>Valid Deductive</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Treatment</td>
<td>14</td>
<td>2.29</td>
</tr>
<tr>
<td>Comparison</td>
<td>10</td>
<td>1.90</td>
</tr>
</tbody>
</table>

* The mean difference is significant ($p = 0.05$) at the $\alpha = 0.05$ level.

** The mean difference is significant ($p = 0.02$) at the $\alpha = 0.05$ level.

TABLE XXXIII

MEANS AND STANDARD DEVIATIONS OF STUDENTS' SELF-REPORTED UNDERSTANDING OF EACH TYPE OF PROOF
### Pre-Assessment Confidence

<table>
<thead>
<tr>
<th>Proof Type</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informal Treatment</td>
<td>15</td>
<td>2.40</td>
<td>0.63</td>
</tr>
<tr>
<td>Comparison</td>
<td>12</td>
<td>2.33</td>
<td>0.78</td>
</tr>
<tr>
<td>Empirical Treatment</td>
<td>15</td>
<td>2.53</td>
<td>0.74</td>
</tr>
<tr>
<td>Comparison</td>
<td>11</td>
<td>2.82</td>
<td>0.60</td>
</tr>
<tr>
<td>Invalid Deductive</td>
<td>15</td>
<td>2.20</td>
<td>0.68</td>
</tr>
<tr>
<td>Comparison</td>
<td>11</td>
<td>1.82</td>
<td>0.75</td>
</tr>
<tr>
<td>Valid Deductive</td>
<td>14</td>
<td>2.14</td>
<td>0.66</td>
</tr>
<tr>
<td>Comparison</td>
<td>10</td>
<td>2.10</td>
<td>0.88</td>
</tr>
</tbody>
</table>

### Post-Assessment Confidence

<table>
<thead>
<tr>
<th>Proof Type</th>
<th>N</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Informal Treatment</td>
<td>14</td>
<td>2.79</td>
<td>0.43</td>
</tr>
<tr>
<td>Comparison</td>
<td>12</td>
<td>2.50</td>
<td>0.52</td>
</tr>
<tr>
<td>Empirical Treatment</td>
<td>15</td>
<td>2.93</td>
<td>0.26</td>
</tr>
<tr>
<td>Comparison</td>
<td>11</td>
<td>2.91</td>
<td>0.30</td>
</tr>
<tr>
<td>Invalid Deductive</td>
<td>15</td>
<td>2.60*</td>
<td>0.51</td>
</tr>
<tr>
<td>Comparison</td>
<td>11</td>
<td>1.82</td>
<td>0.60</td>
</tr>
<tr>
<td>Valid Deductive</td>
<td>15</td>
<td>2.67*</td>
<td>0.49</td>
</tr>
<tr>
<td>Comparison</td>
<td>11</td>
<td>2.18</td>
<td>0.60</td>
</tr>
</tbody>
</table>

* The mean difference is significant ($p = 0.03$) at the $\alpha = 0.05$ level.

** The mean difference is significant ($p = 0.001$) at the $\alpha = 0.05$ level.

TABLE XXXIV

MEANS AND STANDARD DEVIATIONS OF STUDENTS’ SELF-REPORTED CONFIDENCE ABOUT CORRECTLY EVALUATING EACH TYPE OF PROOF
As shown in Table XXXIII, on the pre-assessment there were no significant differences in self-reported understanding ratings between the treatment and comparison group for any of the proof types. Both groups rated the valid deductive proof lowest in understanding on the pre-assessment. This is meaningful since the valid deductive proof had the most correct classifications on the post-assessment, which is indicative of significant gains in understanding this proof type. On the post-assessment, the treatment group had significantly higher ratings for all proof types except the empirical proof. However, even though the treatment group reported a greater understanding on each proof type (and every student reported that they completely understood the empirical proof), the percentage of correct classifications did not increase for this group.

The results for students’ self-reported confidence about their classifications are similar to what they reported for understanding and are given in Table XXXIV. No significant differences were found for any of the four proof types on the pre-assessment correct percentages between the treatment and comparison group. Students rated both the valid and invalid deductive proofs lowest in terms of their certainty about their classifications. However, the treatment group rated both of these deductive proofs significantly higher than the comparison group on the post-assessment. Thus, while no significant gains were made in either group in correctly classifying proofs as either valid or invalid on the assessments, the treatment group did report greater confidence about their classifications and greater understanding of each of the proof types. This implies that while the collaborative revision process may inspire certainty in students, it does not seem to have an impact on students’ ability to identify valid proofs.
5.2.3 Interview Data Analysis

Four students, two from the treatment group and two from the comparison group, were interviewed to gather initial information about what aspects of a proof are convincing to students and if this changes after a transition to proof course or after participation in a collaborative revision course. Individual interviews were conducted twice with each student; once immediately after the pre-assessment was administered and once immediately after the post-assessment was administered. Each student was asked about their responses on the proof validation tasks and to elaborate on their responses and give reasons about why the classified each proof as valid or invalid. The interviews were semi-structured and a guide for each interview can be found in Appendix C.

All student interviews were transcribed and then coded using the proof schemes of Harel and Sowder (1998) as a framework. The proof scheme being employed during each proof validation task was determined from the student’s verbalization of their thought processes. For example if a student remarked that they chose a particular proof to be valid because it looked like something that would be in a math book (as one student did; see Section 5.2.4.1) then this would correspond to the external conviction category of proof schemes and more specifically it would be classified as a ritual proof scheme. If another student brought up a specific proof type that they thought should have been used in a particular proof, this would be an example of an internalized proof scheme. Harel and Sowder (1998) note that it is possible for students to exhibit different proof schemes when presented with different proofs and this was taken into account during the data analysis.
5.2.4 Treatment Group Interview Results

Two students from the treatment group were interviewed after each of the assessments were given. The results of students interviews and what they reveal about the proof schemes being employed is given in the next two sections. Comments that the students made during the interview are presented to provide evidence for how the proof scheme each student was using was assessed. Parts of the quotes below are bolded to highlight to the reader the comments or phrases deemed most important when classifying students proof schemes. Each student is given a pseudonym when reporting the results as to conceal their identity.

5.2.4.1 Stephanie’s Proof Schemes

Stephanie began the semester with a very strong ritual proof scheme. This was evident in the comments that she made on the first interview regarding each of the four proofs and why she classified them as she did. When asked why she classified the informal proof as invalid, she gave a telling response about her proof scheme on this proof validation task.

Stephanie: So, I would say the first statement is true. And then, if you have three consecutive numbers one will be a multiple of 3 because every third number is 3. I'm going to say I think that’s also true too because, like, I mean, every third number will be a multiple. Also, at least one number will be even and all even numbers are a multiple of 2. I wanna say that’s true too because like any two numbers are...at least one or two of them will be even. Thus, if you multiply three consecutive numbers the answer will be at least one factor of 3 and at least one factor of 2.

S: Um, to me it seems true but for some reason it doesn't seem like a proof that would be in a math book I guess.

In this excerpt, she goes through the argument and relays that it is convincing to her yet she is hesitant to classify it as valid because it was not in a format similar to math books that she
has seen. Her comments mirror almost exactly the example that Harel and Sowder (1998) give when defining the ritual proof scheme.

When pressed further about if she believed this was a convincing proof, she said the following:

Emily: Ok. So you’re saying that maybe it convinces you but you don’t think it would be convincing maybe to other people.
S: Well, I, it convinces me but it just doesn’t sound like something really scarily complex in a math book. Because I notice math proofs are like really, maybe I have read a ton of math proofs and it doesn’t seem math, math ‘proofy’ but I don’t know how exactly.

Again, even though she went through the proof line by line to begin with, she informed her decision about validity solely on surface features of the proof, namely that it doesn’t look mathematical enough. Surprisingly, this was at odds with the level of conviction she herself obtains from the proof as evident in the excerpt below.

E. Ok so let me ask you this. Do you think, so in terms of valid and invalid do you think its valid or invalid or still not sure?
S: Well for me I think it would be valid but then I put invalid because I thought I wouldn’t see it in a math book. That’s why I put that because I wasn’t sure. Like, and then how well do you understand this proof, like, I was kinda thinking along the terms of whether it would be in a math book not whether I understood what it said. Because usually stuff I understand it like not stuff that’s in a complex math book.

She acknowledges the fact that the informal proof, the way it’s written, doesn’t look like something she would see in a math book and that is why she classified as invalid. However, she believes that it is a perfectly convincing proof for her. This behavior is consistent with the
results of Segal (2000) showing that students often make a distinction between arguments that convince and arguments that prove.

This proof scheme persists when she is asked about each of the other proofs. When asked about the empirical proof, she again brings up aesthetic aspects of the proof.

S: See this is, like, examples. That’s why I said invalid, I was thinking it’s like examples. Um, and, like, it would convince me personally, but, like, I know for sure I would never see this in a math book for anything as a proof; maybe for an example. When I put invalid I was thinking whether I would see it in a math book for all these.

This is consistent with the study of Weber (2010) that undergraduate students often have a disconnect between what is convincing to them and what they believe will be convincing to others. Stephanie is clearly displaying that in her comments about the informal and empirical proofs; she is personally convinced, but hesitates to classify these proofs as valid because their appearance does not match her strongly held ritual proof scheme.

When asked about the invalid deductive proof, she exhibits the same proof scheme, but this time she classifies this proof as valid because it did look like a proof from a textbook. Even though she spent some time going through the proof line by line her final determination about the proof’s validity was made on the way the proof looks.

S: Well okay, because it’s defining what $x$ is because like $x$ is a whole number, so you’re thinking like 1 to infinity for whole numbers. It’s, like, clearly defining what the product of three numbers is. It’s like what a math book would do and then it, like, factors out and, like, shows you factoring.

S: Um, looking for the coefficients. Um, I wasn’t quite sure if that, maybe, because it seemed kind of confusing. I thought that would be in a math book. Because I’m not sure, like to me, looking at the coefficients...I don’t know how the coefficients...I
mean, sure they do add up to 6 but I’m not sure how they would relate to, like, knowing that any three consecutive integers is a multiple of 6. So I’m like oh that would be something they would put in a math book.

Stephanie strongly exhibited this proof scheme; to the point where she rarely looked at the words and meaning of the proof. In the quote above Stephanie did follow the argument until she got to a statement she wasn’t sure about (and was, in fact, invalid), at which point she appealed again to the format of the proof.

For the valid deductive proof she reported little understanding of the argument and thus made her decision about validity solely on if it looked mathematical to her.

E: Um, and then, let’s just take a look at the last page. So you said this one is valid and is that for sort of the same reasons...like, it looks like a math book proof.
S: Yeah, I hate to say that’s kind of why I decided which ones are valid and invalid.

S: I think because it’s like going into more cases and it gets kind of confusing and wordy and it’s hard to pay attention to. And it just seems like something that would be in a math book.

However, on the second interview her proof scheme had changed dramatically. She was evaluating proofs line by line and she was concerned if the statements logically followed from one another. During this interview she exhibited a transformational proof scheme, which was evident in her comments below when referring to the informal proof.

S: ...Obviously this proof is true I was just wondering because, like, the only thing it didn’t answer was like, okay, you have a 2 and a 3 and it’s a multiple of 6, obviously. Can the third number ever ruin that? Like, you know what I mean? ;
S: Like that’s the only thing the proof didn’t talk about. I mean it doesn’t seem like
it would logically, but you never know. So I was just wondering if that would change it or in that proof, that was my only question.

When she considered the 'third number' here, she displayed a mental construct of the meaning of divisibility by 2 and divisibility by 3 and she performed a transformation on this construct (i.e. What happens when a third number is multiplied? Is the product still divisible by 6?)

She was also able to work through the question that she posed and she showed evidence of reflection on the questions she had posed.

S: Like later I was thinking about, is it ever going to be possible to, like, make it not a multiple of 6? Um, and I was thinking you couldn't really because whenever you multiply something by 6, well 2 and 3 are a multiple. Like it’s never going to happen, so it’s not a concern.

This quote gives evidence that Stephanie is considering the generality aspect of the statement and she is anticipating transformations. Here she is concerned that, even if there is certainly a multiple of 2 and a multiple of 3 in the product, can multiplying be a third number affect divisibility by 6. She is exhibiting higher-level thinking than in the first interview where only surface aspects of the proof were considered.

When asked about why she classified the empirical proof as invalid, Stephanie responded:

S: But this time I’m sure it’s invalid because it’s just examples. Like, there’s not that one it didn’t show. Because you can always come up with examples that are true, but like it might not be true for all of them. And this one is not valid because there’s just examples but not a general proof.

This again provides evidence that she is aware that we must proof this statement in general and, unlike the first interview, this proof is no longer convincing to her. Also, this statement
implies that she has moved beyond an inductive proof scheme since she classifies this as merely examples and knows that it is not considered a proof. Here she shows evidence of understanding that proofs must take on a certain form, such as being written in general, and she understands that all statements must be justified. However, she never specifically talks about any proof techniques (say, induction or contradiction), which implies that she is in the lowest-level of transformational proof schemes.

Thus, Stephanie began the semester with the lowest level proof scheme, a ritual proof scheme, according to Harel and Sowder (1998). She evaluated arguments not based on logical aspects about the statements, but instead by making a determination about whether the proof looked like something that would be in a textbook. Clearly, it is important to move students away from this type of proof scheme and into the higher category of axiomatic proof schemes. In fact, this goal was achieved with Stephanie after participation in collaborative revision; she was able to move past an inductive proof scheme and into a transformational proof scheme. She demonstrated that empirical arguments were not convincing to her by rejecting the empirical proof as “just examples” and she showed evidence of evaluating arguments on their logical progression rather than on aesthetic aspects.

5.2.4.2 Robert’s Proof Schemes

Robert exhibited comments consistent with an internalized proof scheme on the first interview. He was quite insistent that this statement should be proven using induction and he compared all of the arguments to what he anticipated an induction proof would look like. He also went further to try to prove the statement using induction during the interview, but he
was unable to do so. When he was asked about why he responded that the empirical proof was invalid, he brought up induction.

Emily: But, um, so this one you said invalid and you seemed pretty sure about that. So after you read it can you just elaborate a little bit about why you said that.
Robert: So, it doesn’t cover all cases, for every single, uh, trio of consecutive integers something like induction would.

When discussing the invalid deductive proof, he again made mention of using induction.

R: The product of any two consecutive integers is a multiple of 6 so \( x \) is a whole number. Uh, we’re doing the base case for an inductive proof, but, uh...

Robert had a hard time evaluating each of the arguments because of his firm belief that induction was necessary for the proof. This is indicative of an internalized proof scheme. Since the statement to be proven was about all natural numbers (specifically, all sets of three consecutive natural numbers) he expected induction to be used somewhere in the proof and proceeded to compare each proof to an imagined inductive one. Robert also had taken (but ended up withdrawing from) some prior proof-based mathematics classes so it is possible that he was mirroring a proof heuristic he had previously seen.

Robert still favored induction on the second interview, but this time he communicated a more intuitive-axiomatic proof scheme. On the second interview, he was acutely aware that proofs must follow a format and must be built upon other valid statements. In fact, he was the only interviewee that was concerned about the invalid statement in the invalid deductive proof and he spoke about a way that to test the truth of that statement.

E: So the last line is not clear or you’re not sure what it’s referring to?
R: Yeah, it seems like nonsense. That’s not always the case that you can put
a number in a polynomial and have the coefficients add to 6. You can put...uh, maybe 2’s will work, I don’t know. (does some arithmetic) That actually works out.
E: So you’re saying if you were to actually plug in some numbers...
R: Yeah, it probably wouldn’t work.

He proposed a way to determine if this statement is true and, even though he was not able to successfully apply the test, he concluded that the proof must be invalid if that statement is invalid. This showed an appreciation of the logical chains present in proofs and he recognized if one element in the chain is invalid the whole proof is invalid.

He still preferred that this statement be proven by induction and again offers it when asked about the informal proof. However, this time he is able to address the merits of the proof and able to recognize that the narrative structure of the proof does not have the appropriate rigor.

E: What about this proof makes you hesitate a little bit to classify it as valid?
R: So there’s something about a difference between this proof and what’s typically regarded as rigorous in a proof structure. i.e. induction is missing. The core statements of induction, like when one case leads to another, is missing in this. But the logic is completely sound.

The results of Robert’s interview show that he held an internalized proof scheme at the beginning of the semester, which means that he was concerned with the use of a proof heuristic, specifically induction, and was classifying the arguments as valid or invalid based on how they compared to his idea of an inductive proof. At the end of the course, his proof scheme evolved into an intuitive-axiomatic one, meaning that he realized that proofs must originate from definitions and axioms. Though he was aware of the main structure of a proof and that proofs must precede from know statements, he was still very much deciding about proofs according to his own intuition.
5.2.5 Comparison Group Interview Results

From the comparison group, two students were also chosen to be interviewed. A report of the proof schemes these students exhibited on the first interview and on the second interview is given in the next two sections. We present the proof scheme employed by the students as well as comments they made that led to their being classified under that particular proof scheme. Again comments in boldface denote the main evidence that was used in deciding which proof scheme a student was employing. Pseudonyms are given to each of the students to maintain anonymity.

5.2.5.1 James’ Proof Schemes

During the first interview, James spoke often about the need for examples in proofs. It became clear in his responses about the valid deductive proof that he did see a need for some sort of generality in proofs and for proofs to use logical progressions, but he held on to the belief that examples suffice when proving. Thus, James is classified as having an empirical proof scheme on the first interview.

When asked if the informal proof would be convincing to other people, James seemed to think that the only flaw in the proof was that it did not contain sample numbers so the reader could have an example present while trying to follow the proof.

Emily: Ok, great. And do you think...So, seems like this convinced you, do you think it would convince other people?
James: Uh, I mean some people would probably have questions about it unless some people. The way I would probably do it personally, I’d say if you have consecutive numbers, you know. I would probably give an example 61, 62, and 63. Yeah or, you know, something like that and then every third number...and then you do another one so they can see. So people can physically see the numbers
because sometimes some people can’t really, you know. Some people don’t really have the spatial intelligence to see numbers out there unless they see it right in front of them.

This is indicative of the empirical proof scheme mindset; that statements are verified by considering specific cases. It is important to note that James was a secondary mathematics education major, which may be why he felt strongly about presenting examples since he may feel that showing examples is a characteristic of a good teacher.

James also seemed particularly convinced by the empirical proof and added only that he would break the numbers into their prime factorizations so that each product’s divisibility by 6 would be even more apparent. When asked if he thought others would be convinced by the empirical proof, he said the following:

J: Yeah, I mean if they don’t see it, you would have to, I mean. Probably...for this one right here. 3, 4, 5. I would probably you know, since 4 equals 2 times 2 and 3 times 2 equals 6. That’s what I would probably add to something like that. That way people would be able to see it. Like I said, some people can’t spatially see that 2 times 2 equals 4 and so I would do that.

Thus, he felt that this was a fairly convincing proof and only need to show more algebraic steps to be fully convincing.

There is evidence that James did recognize some of the basic aspects of rigorous proof, but he was still convinced when these features were not present in an argument. He classified both the informal and empirical proofs as valid on the pre-assessment. However, when asked why he classified the invalid deductive proof as invalid he remarked that:
J: There’s no theorem, there’s not...that’s one of the reasons, **that’s the main reason I said it was invalid because there’s no theorem.** They didn’t have a proof or a theorem stating this.

Thus, he acknowledges that if a statement is used in a proof it should follow from another statement that we know to be true (such as an axiom or definition) or that has been previously proven. He also rated the valid deductive proof as valid and remarked that “that’s basically the written explanation of what that is,” referring to it as an elaboration of the informal proof. Therefore, in the first interview, James seems to appreciate that proofs should be logical progressions and statements should be explained, yet remains convinced by proofs that are merely examples and do not have these aspects.

On the second interview, James has no longer holds an empirical proof scheme. When asked about the empirical proof and why he marked it invalid on the post-assessment, he responded:

J: The reason why I said it’s invalid because it’s only going up to 6, 7, 8. **He is assuming, but you haven’t really proved it for, you know, another number, for all numbers, all three numbers, all three consecutive numbers.** It’s just basically, okay, well if he had it...if it was, it’s not complete. That’s what I’m trying to say. You’re just assuming, but you haven’t really proved anything.

Hence, he is no longer convinced by this empirical arguments and he realized that proofs must go beyond examples.

While he never explicitly talks about a generality requirement for proofs, when asked about the valid deductive proof he commented:

J: Because you’re showing algebraically that it’s all consecutive numbers. **It’s like you can say any number you choose is going to have a multiple of two and a multiple of three, you know, no matter what.**
This quote showed that James understood that if a statement was written in general then the proof should cover all possible cases. He also alluded to the importance of using known or proven statements to prove other statements when he said:

J: Looking at this proof, you know, the coefficients. **If there was a, you know, using a theorem stating that, you know, I would say yeah that would be valid.** But there’s no, you know, theorem, there was no prior theorem stating that statement. that why my reason is that its invalid.

These quotations indicate a transformational proof scheme, where he is recognizing the importance of generality and building a hierarchy of statements, but he does yet display anything like an internalized proof scheme, which would be the next level. This is because he never mentions any proof techniques by name and he understands only the basics of a valid proof (i.e. it must be general and it must use theorems).

James began the semester with a strong empirical proof scheme and was convinced by proofs that just showed examples. This could be due to the fact that he is a mathematics education major and, without much proof background, he could have approached proof like a secondary mathematics course, where more examples can aid students’ understanding of concepts. However, by the end of the semester he was able to recognize that empirical proofs are not valid because they do not cover all possible cases. So, he held a transformational proof scheme, but, I would argue, a very low-level one. He was able to convey some aspects that are required in a proof, but he was still very convinced by the informal proof, calling it “more straightforward.”
5.2.5.2 Francine’s Proof Schemes

Of the four interviewees, Francine said the least about what was convincing to her during the interviews, even when asked directly. It is clear that she was beyond an empirical proof scheme at the beginning of the course because she chose the empirical proof to be invalid and she said that it was not convincing to her and she did not think it would be convincing to others either. When asked about the empirical proof, she commented:

Francine: Because it’s just kinda like, like this proof they just...compared with proof one...but this one, like they just give you, like examples and they say it’s true.
E: So thats not convincing for you necessarily?
Francine: No...this is not precise, I mean, yeah, this proof.

Francine seemed the most comfortable with the informal proof and the argument presented therein.

F: Yeah, I chose it’s valid because, well I think to prove the statement it is pretty straightforward. It says, yeah a multiple of 6 must have factors three and two.

This implies that while she recognized the basic structure of a proof, that it must have some aspect of generality and it must be a hierarchy of statements, but she did not yet have the skills to evaluate arguments when they were based on mathematical definitions and theorems. Thus, we concluded that she did not find empirical arguments convincing and does recognize that proofs have certain characteristics, which is indicative of a transformational proof scheme.

On the second interview, Francine was less convinced by the informal proof. She still classified it as valid, but recognized that it was not the most rigorous way to prove the statement.
F: Today, like, after I read this and, uh, I sort of compared with the second one [referring to the empirical proof] it’s sort of the same. But, this one, like, explain more like, um, instead of the examples...Well I think it’s valid, but it’s not like that precise, so yeah.

During this interview, she identified the valid deductive proof as the most straightforward one, whereas on the first interview she said it was the informal proof. When discussing the valid deductive proof, she said:

F: It’s pretty straightforward. I mean the, the proof. Yeah and he talks about like the two parts, the first part is if the number is even and if its odd and if it’s even it could be like still like another 2. I mean like two possibilities because it is divisible by 3 or its not.  
E: Why do we need to do the two cases?  
F: Because it covers all the possibilities for the numbers, the integers.

This implies that Francine is aware of the generality aspect of proofs and that she is comfortable evaluating the mathematical statements used in a proof. However, there does not seem to be many differences from her earlier interview and so she is again classified as having a transformational proof scheme.

Although Francine was less convinced by the informal proof and more convinced by the valid deductive proof on the post-interview than the pre-interview, there was no evidence that her proof scheme had changed. She never made mention of axioms, definitions or theorems as a necessary part of a proof, indicating she had not adopted an axiomatic proof scheme. Nor did she discuss any proof techniques or heuristics, meaning she was not using an internalized or interiorized proof scheme. Thus, it seems that her proof scheme remained transformational throughout the semester.
5.2.6 Between Groups Interview Results

The interviews conducted with the two students from each group (i.e. treatment and comparison) illuminated the types of proof schemes being employed by the students. Table XXXV shows the proof schemes exhibited by each of the students interviewed on the first interview and on the second interview. From this table, it seems that students in the treatment group were able to transition into higher-level proof schemes from the first to the second interview, while students in the comparison group remained basically at the same level. This is clear if we consider again the proof schemes hierarchy tree shown in Figure 4. Although the proof schemes are not organized in a strict hierarchy, it can be said that external conviction proof schemes are at a lower level than inductive proof schemes which are in turn at a lower level than analytical proof schemes. Additionally, under analytical proof schemes, transformational proof schemes are at a lower level than axiomatic proof schemes.

One student in the treatment group, Stephanie, began the semester with the lowest-level proof scheme, a ritual proof scheme. After the post-assessment, however, she was no longer

<table>
<thead>
<tr>
<th>Student</th>
<th>First Interview</th>
<th>Second Interview</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Treatment</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stephanie</td>
<td>Ritual</td>
<td>Transformational</td>
</tr>
<tr>
<td>Robert</td>
<td>Internalized</td>
<td>Intuitive-Axiomatic</td>
</tr>
<tr>
<td><strong>Comparison</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>James</td>
<td>Empirical</td>
<td>Transformational</td>
</tr>
<tr>
<td>Francine</td>
<td>Transformational</td>
<td>Transformational</td>
</tr>
</tbody>
</table>

TABLE XXXV

PROOF SCHEMES EXHIBITED BY STUDENTS IN EACH GROUP DURING EACH INTERVIEW
evaluating arguments on appearances and seemed to have elevated to a transformational proof scheme. This means that she ended the semester with an analytical proof scheme and has graduated up two ‘steps’. Also, Robert from the treatment group began the semester with an internalized proof scheme, which was at a higher level than the other interviewees. He still was able to improve on the second interview and was found to hold an intuitive-axiomatic proof scheme. This is a jump from a transformational proof scheme into an axiomatic one, which is another ‘step’ up in analytical proof schemes.

However, students in the comparison course largely embraced the proof schemes they had entered the course with. The comparison group student James had an empirical proof scheme at the beginning of the course and was convinced by inductive arguments, although he recognized some aspects of a transformational proof scheme. At the conclusion of the course, James had improved slightly to a transformational proof scheme, where he was no longer convinced by empirical arguments, yet did not have a full grasp about all the requirements of a proof. He recognized that proofs require generality and was aware that theorems are often used, but still had naive notions of validity. Meanwhile, Francine did not make any improvement in her proof scheme during the semester. She began the course unconvinced by empirical proofs (and thus not employing an inductive proof scheme) yet appreciating only the basics about the components of a proof. She, like James on the second interview, was aware that proofs should be written in general, but did not mention any proof techniques or anything regarding axioms or theorems.
5.2.7 Synthesis of Results on Students’ Proof Validation Skills

The results presented in this section illustrate the similarities and differences between the treatment and comparison group. The treatment group had a proof classification rate at the beginning of the semester (60%) that was not much better than guessing on each proof. On the post-assessment the average percentage of correct classifications for students in the treatment group did not increase at all. This is consistent with the studies on undergraduate students of Selden and Selden (2003), Alcock and Weber (2005), and Brandt and Rimmansch (2012) that students have poor proof validation skills even after a transition to proof course. However, it is surprising in the context of this study as these students were participating in determining the validity of each other’s proofs throughout the entire semester and so it was hypothesized that their proof validation skills would improve on the post-assessment. Students in the comparison group did not fare much better when evaluating proofs. Beginning the semester with an average rate of 45%, the comparison group was only able to improve to 54%, which is still lower than the average for the treatment group. Although there was increase, it was not found to be significant.

Even though the treatment group did not show an increase in average percentage of proofs correctly classified from the pre to the post-assessment, they did report a significantly better understanding than the comparison group on almost every proof. Additionally, the treatment group reported significantly higher confidence of their classification than the comparison group on two out of four proofs. So, collaborative revision has an impact on students’ ability to understand a written proof and their certainty about classifying that proof. This is surprising
because even though confidence and understanding increased for the treatment group, the number of correct classifications did not.

Though it seems that collaborative revision does not assist students in identifying valid proofs, there is encouraging evidence from the interviews that it may impact the way that students gain conviction about an argument. When we consider the proof schemes of Harel and Sowder (1998) that students were employing at the beginning and end of the semester, the treatment group students that were interviewed displayed a greater transformation in their proof schemes than students in the comparison group. The treatment group students increased several steps up the hierarchy while students in the comparison group mainly stayed at the same levels. This is important because students’ struggles with proof construction could be linked to their misunderstandings of what constitutes a proof. A student with a higher-level proof scheme will be unconvinced by an empirical or non-sensical symbolic argument and possibly become a better proof-writer.

5.3 Students’ Proof Construction Skills

In order to address our final research question about any impacts that participation in collaborative revision may have on students’ proof construction skills, data from the proof construction task on the pre and post-assessment were analyzed. In the next section, the data analysis techniques and the use of a coding scheme developed by Andrew (2009) are discussed. The use of this coding scheme provided a framework for data analysis of errors that students made while attempting a proof of the statement on the assessment. Results from the proof construction task for the treatment group are given, including the number of valid proofs
written by this group on each assessment, an analysis of the errors made on proof attempts and an overview of the different proof techniques that were used. Comparison group results are also given and then a contrast between the two group is made.

5.3.1 Data Analysis

To aid in determining any impact of collaborative revision on students proof construction skills, the proof construction task on the assessments were analyzed to determine validity of proofs and types of errors made. The statement that students were asked to write a proof of was “For any positive integer \( n \), if \( n^2 \) is divisible by 3, then \( n \) is divisible by 3.” The entire assessment given to students can be found in Appendix A.

In order to evaluate how student proof-writing progressed, errors made on the proof construction task on the pre and post-assessment were coded using the Proof Error Evaluation Tool (PEET) developed by Andrew (2009). This coding scheme was designed for university instructors to use when grading student proofs, as well as for students to use while constructing proofs (Andrew, 2009). Errors are divided into two categories, errors in proof structure (S) and errors in understanding (U), and then divided further into the specific error categories given in Figure 11. Andrew defines structure errors as those affecting the format and flow of the proof, whereas understanding errors betray a lack of comprehension of the concepts needed for a particular proof. Note that a single code may cover up to three related errors, which was done intentionally to allow professors to give a code to student proofs that will hint at where an error occurs and what type of error it is, but not give away the exact nature of the error to challenge students to work through their errors.
<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>Introduced variables without defining them or performed operations that</td>
<td>Proof setup</td>
</tr>
<tr>
<td></td>
<td>were undefined.</td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>The approach taken in proving a statement will not work.</td>
<td></td>
</tr>
<tr>
<td>S1</td>
<td>The proof was to be completed using a specific method, but this method</td>
<td></td>
</tr>
<tr>
<td></td>
<td>was not used.</td>
<td></td>
</tr>
<tr>
<td>S2</td>
<td>Made a false assumption somewhere in the proof.</td>
<td>Correct assumptions</td>
</tr>
<tr>
<td>S3</td>
<td>Didn’t proceed through the proof in a linear fashion, and ideas were not</td>
<td>Linear/Sequential Order</td>
</tr>
<tr>
<td></td>
<td>in logical order.</td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td>The proof contained extraneous details or steps that did not really</td>
<td>Stray detail/Conciseness</td>
</tr>
<tr>
<td></td>
<td>contribute to the proof.</td>
<td></td>
</tr>
<tr>
<td>S4</td>
<td>The length of the proof was unnecessarily long and thus extremely difficult</td>
<td></td>
</tr>
<tr>
<td></td>
<td>to follow.</td>
<td></td>
</tr>
<tr>
<td>S5</td>
<td>The write-up was illegible at times, making it difficult to read and/or</td>
<td>Neat presentation</td>
</tr>
<tr>
<td></td>
<td>understand.</td>
<td></td>
</tr>
<tr>
<td>S6</td>
<td>Relied too much on calculator or computer-generated information in one</td>
<td>Technology’s place</td>
</tr>
<tr>
<td></td>
<td>step of the proof.</td>
<td></td>
</tr>
<tr>
<td>S7</td>
<td>Needed to show ( p \rightarrow q ), but did not show directly, or by</td>
<td>Proof type</td>
</tr>
<tr>
<td></td>
<td>((\sim q) \rightarrow (\sim p)) or by contradiction.</td>
<td></td>
</tr>
<tr>
<td>S7</td>
<td>Only gave an example to establish the truth of a mathematical statement.</td>
<td></td>
</tr>
<tr>
<td>S8</td>
<td>Used nonstandard or confusing notation.</td>
<td>Correct use of symbols or</td>
</tr>
<tr>
<td></td>
<td></td>
<td>notation</td>
</tr>
<tr>
<td>U1</td>
<td>Wrote a statement that was not justified, explained, or verified.</td>
<td>Sufficient details</td>
</tr>
<tr>
<td>U2</td>
<td>Wrote a statement or paragraph that was ambiguous, confusing, and/or</td>
<td>Clarity</td>
</tr>
<tr>
<td></td>
<td>unnecessarily complex.</td>
<td></td>
</tr>
<tr>
<td>U3</td>
<td>Failed to include an illustrative picture that would make the proof easier</td>
<td>Pictures in proof</td>
</tr>
<tr>
<td></td>
<td>to understand</td>
<td></td>
</tr>
<tr>
<td>U3</td>
<td>Relied too much on a picture to prove something was true.</td>
<td></td>
</tr>
<tr>
<td>U4</td>
<td>Did not sufficiently justify a crucial step in the proof.</td>
<td>Crucial step/main idea</td>
</tr>
<tr>
<td>U4</td>
<td>An error caused important parts of the proof to be left unaddressed.</td>
<td></td>
</tr>
<tr>
<td>U5</td>
<td>Made a false statement or incorrect computation in the proof.</td>
<td>Correct implications and</td>
</tr>
<tr>
<td></td>
<td></td>
<td>statements</td>
</tr>
<tr>
<td>U5</td>
<td>Incorrectly claimed that one statement implied or equaled another statement.</td>
<td></td>
</tr>
<tr>
<td>U6</td>
<td>Included some cases but not others (which were not trivial).</td>
<td>All cases present</td>
</tr>
<tr>
<td>U6</td>
<td>Did not address one aspect of the problem.</td>
<td></td>
</tr>
</tbody>
</table>

Figure 11. The Proof Error Evaluation Tool (PEET) (Andrew, 2009).
There are a number of reasons that this particular coding scheme has been chosen. First of all, this error evaluation tool was developed using an iterative process, generating categories by looking at the researcher’s work as a student, then refining and adding categories by applying the tool to student work and finally applying the tool to proofs from the literature to finalize the categories. This ensures that the errors that encountered in students proofs analyzed for this research fell into one of the categories given in the tool. Secondly, in his paper, Andrew (2009) applied the PEET to the three incorrect proofs given in Selden and Selden (2003), all of which are proofs of the same statement for student proof construction used on the assessment in this study. An example of a student proof of the statement “for any positive integer \( n \), if \( n \) square is a multiple of 3, then \( n \) is a multiple of 3” and the error coding from Andrew (2009) using the PEET tool is given in Figure 12. Since proofs of the same statement used on the assessment in this study were analyzed in the paper, there were sufficient examples to compare with to ensure coding reliability during the analysis. Finally, this is a fairly simple coding scheme and allowed for easy identification of the errors that appear in the proofs to be analyzed. The student proofs from the third part of the assessment were coded using this scheme and errors were compared between the experimental and comparison courses and between the pre and post-assessment to determine the impact that completion of either course has on students’ abilities to construct valid proofs.

It is important to note that a proof can be considered valid in this study while still containing several setup or understanding errors from the PEET. Valid proofs often had variables that were not properly defined (S1) or statements written that were not fully justified (U1).
THEOREM. For any positive integer \( n \), if \( n \) squared is a multiple of 3, then \( n \) is a multiple of 3.

"Proof (d)". PROOF. Let \( n \) be a positive integer such that \( n^2 \) is a multiple of 3. Then \( n = 3m \) where \( m \) is a positive integer. So \( n^2 = (3m)^2 = 9m^2 = 3(3m^2) \). This breaks down into 3\( m \) times 3\( m \), which shows that \( m \) is a multiple of 3.

Moreover, different codes can be applied to the same proof if the student made more than one error. In the same manner, the same proof can have the same code applied to it in different places. For example, a student could write several statements throughout the proof that were not justified, explained or verified, which would result in the code U1 being applied several times to the same student’s proof. Also, the same statement was given to be proved on the post-assessment as given on the pre-assessment.

Several examples of how I analyzed proofs collected in this study using the PEET will now be presented. Figure 13 shows a student’s proof from the pre-assessment and how it was analyzed using the PEET tool. This proof was considered invalid and was assigned three codes:

---

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>S4</td>
<td>This statement is not used anywhere in the proof.</td>
</tr>
<tr>
<td>S7</td>
<td>Is attempting to prove ( q \geq p ).</td>
</tr>
<tr>
<td>U2</td>
<td>This statement is confusing because it was already shown that ( n^2 = 3(3m^2) ), clearly a multiple of 3, but then it is stated that ( n^2 ) is a multiple of 3 because it “breaks down into 3( m ) times 3( m ).”</td>
</tr>
</tbody>
</table>

---

Figure 12. The PEET tool (Andrew, 2009) used to analyze errors from a student proof in Selden and Selden (2003).
For any positive integer \( n \), if \( n^2 \) is divisible by 3, then \( n \) is divisible by 3.

For any positive integer \( n \),
so \( 3n \) can be divisible by 3, \( S1 \)
so \( (3n)^2 = 9n^2 \).
\( 9 \) can be divisible by 3, \( S4 \)
so \( 9n^2 \) can be divisible,
then \( 3n \) can be divisible. \( U4 \)

Hence: Since \( n^2 \) is divisible by 3, \( n \) is divisible by 3

Figure 13. A student proof from a pre-assessment and the coding applied.

from the PEET. The first code was \( S1 \) applied to the second line of the proof since the student began by considering the integer \( n \) and then immediately transitioned to considering \( 3n \) without explanation. This student introduced \( 3n \) without justifying why this should be done and, in fact, it was not needed in the proof. The second code applied was \( S4 \) on line 4 of the proof. Here the student claims that \( 9 \) is divisible by 3, which is true, but is entirely extraneous information. The last code applied to this proof was \( U4 \) on lines 5-6. The student claims that \( 9n^2 \) is divisible by 3 implies that \( n \) is divisible by 3. In this proof, that argument leads directly to the conclusion of the statement and it is not sufficiently explained, nor is it true.

Another student proof analyzed using the PEET tool is given in Figure 14. This proof was considered valid and was also assigned three codes from the PEET. The first code, applied to the proof as a whole, was \( S3 \) since the proof was not linear in any way. This student used
proof by contradiction and then proof by cases, but that is very hard to see from this proof since the cases are not delineated at all and there is a lot of arrows and text in many different directions. The code S8 was applied because the student defines $n$ as an integer, but then considers $3n, 3n + 1$ and $3n + 2$. The standard mathematical notation would be to define $n$ as an integer that can be written as either $3k, 3k + 1$ and $3k + 2$, where $k$ is an integer. It is confusing in this proof to use $n$ in two different contexts. Finally, the code U1 was applied because the breaking up of this proof into the three cases ($3n, 3n + 1$ and $3n + 2$) is not explained, nor is the ‘if and only if’ used in the conclusion of this proof.
Besides identifying errors in students’ proofs using the PEET, student proofs were also evaluated for validity and proof approach. While evaluating the proofs it was decided that it was insufficient to classifying proofs as merely valid, invalid or incomplete. Several students wrote proofs that were not completely valid, yet used correct reasoning albeit in a confusing or careless manner, which were classified in this study as ‘weak valid’. This classification is consistent with Coe and Ruthven (1994) who categorized high school students’ proofs in their study as strong deductive, weak deductive or empirical. An example of a proof that was classified as weak valid is given in Figure 15. In this proof, the student is using Euclid’s theorem\(^1\) yet not stating it or giving any justification as to why it should be true. Also, they

\[2 | (n \cdot n) \quad \Rightarrow \quad 2k = n_1 \cdot n_2\]

since \(n^2\) is divisible by three, then either \(n_1\) or \(n_2\) must be divisible by 3, and since \(n_1 = n_2\), then \(3 | n\).\(\square\)

\(^1\)Euclid’s theorem states that if \(p\) is a prime and \(p\) divides \(ab\) for integers \(a\) and \(b\), then \(p\) divides \(a\) or \(p\) divides \(b\).
are using confusing notation by renaming $n$ as $n_1$ and $n_2$, which is unnecessary. However, the underlying idea in this proof is valid even though the reasoning given is lacking.

Proofs were considered incomplete if it was clear that a full argument had not been made. Two examples of proofs that were classified as incomplete are shown in Figure 16. An incomplete proof was often indicated to me as the coder with a question mark in the proof, such as in Figure 16(a). However if only cursory information had been written down, such as in Figure 16(b) where only the contrapositive statement has been written but no attempt to prove it has been made, that proof was also considered incomplete.

Once proofs were classified as either valid, weak valid, invalid or not complete, an examination of each student’s proof type on the pre-assessment compared with the post-assessment was done to determine if any improvements were made between the assessments and for which students. A note was made of the approach taken in each completed proof, whether it was valid or invalid if it was different that just a direct proof of the statement. The proof techniques attempted on the proof construction task included induction, use of Euclid’s theorem, contradiction, proof of the contrapositive, and proof by cases. This was to determine if there were any differences between proof approaches taken by students in the treatment group and students in the comparison group.

5.3.2 Results on Students’ Proof Construction Skills

Results from the treatment group from the data analysis described above is now given. A discussion of the number of valid proofs students in the treatment group were able to write at the beginning and end of the collaborative revision course is given, as well as the frequency
For any positive integer $n$, if $n^2$ is divisible by 3, then $n$ is divisible by 3.

\[ 3 \mid n^2 \Rightarrow 3 \mid n \]

Either $n^2 = 3q$, $n^2 = 3q + 1$, or $n^2 = 3q + 2$

Proof by Induction

**Base Case:** $n = 3$
\[ 3^2 = 3q, \quad q = 3q, \quad q = 3 \]
\[ 3 = 3q, \quad q = 1 \]

**Inductive Step:** suppose $n^2 = 3r \Rightarrow n = 3s$ for $r, s \in \mathbb{Z}$ and $n \geq 3$, prove $(n+1)^2 = 3t$

\[ (n+1)^2 = n^2 + 2n + 1 \]
\[ = 3n + 2n + 1 \text{ by ind. hyp.} \]
\[ = 5n + 1 \]

(a) The first student proof.

For any positive integer $n$, if $n^2$ is divisible by 3, then $n$ is divisible by 3.

**Proof.**

\[ n \in \mathbb{Z} > 0 \]

\[ p \to q \]

\[ 7p \to 7q \]

\[ 79 \to 7p \]

Assume $n$ is not divisible by 3

\[ \text{then } n^2 \text{ is not divisible by 3} \]

(b) The second student proof.

Figure 16. Two student proofs that were classified as incomplete.
TABLE XXXVI

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Pre-Assessment</th>
<th>Post-Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid</td>
<td>0</td>
<td>3 (20%)</td>
</tr>
<tr>
<td>Weak Valid</td>
<td>1 (7%)</td>
<td>4 (27%)</td>
</tr>
<tr>
<td>Invalid</td>
<td>8 (53%)</td>
<td>7 (47%)</td>
</tr>
<tr>
<td>Incomplete</td>
<td>6 (40%)</td>
<td>1 (7%)</td>
</tr>
</tbody>
</table>

TYPES OF TREATMENT GROUP STUDENTS’ PROOFS SEPARATED BY ASSESSMENT

and type of errors that students made on each assessment. Additionally, we consider the proof techniques employed by students in each group before comparing results between groups.

5.3.2.1 Treatment Group Results

When analyzing a student’s proof, it was classified as either valid, weak valid, invalid or incomplete. The additional category of ‘weak valid’ allowed for a distinction between proofs that had major flaws and proofs that could be turned into a valid proof by adding justifications or by reformatting. This section will first present the results about the numbers of valid, weak valid, invalid and not complete proofs that were written by students in the treatment group on each assessment. Then a discussion of the various proof techniques used by treatment group students on the pre and post-assessment will be given followed by a report of the results from the use of the PEET on students’ proofs to analyze errors made.

The results of evaluating all of the treatment group students’ proofs as valid, weak valid, invalid or not complete is given in Table XXXVI. Notice that none of the students in the treatment group were able to write a completely valid proof (however, one student did write
a weak valid proof) on the pre-assessment. Another six students were not able to write any proof at all of this statement. This is perhaps unsurprising since for many of these students this course was the first time they had encountered proofs and so their proof construction skills were not developed. On the post-assessment though almost half (47%) of the students were able to write a valid or weak valid proof and only one student was not able to complete a proof of the statement.

When examining the proof techniques that students in the treatment group used in their proof construction on the pre-assessment as compared to the post-assessment it becomes clear that these students really expanded the number of approaches used in the proof. On the pre-assessment, of the nine completed proofs (valid or invalid), seven of the students used a direct proof approach. Each one of these seven students began their proof with some variation of the statement “let $n^2$ be divisible by 3, then $n^2 = 3k$ for some integer $k$.” One student attempted a proof of the converse of the given statement, but she also used direct proof. Only one student used proof by contradiction and he was the only one to write a weak valid proof on the pre-assessment. Again, this is to be expected as most of these students were unfamiliar with other proof techniques at this beginning of this study. However, the proofs on the post-assessment included a wide range of proof techniques, which resulted in many more valid or weak valid proofs. Of the fourteen completed proofs (valid or invalid) on the post-assessment, seven of the proofs used a technique other than direct proof. One student proved the contrapositive of the statement, one used proof by contradiction, two attempted proof by cases, and three others used Euclid’s theorem to simplify the proof; this is summarized in Table XXXVII. Even the
one student who could not complete a proof of the statement wrote down the contrapositive of
the statement, but got stuck when attempting to prove it. Thus, the treatment group students
were able to approach a proof in a variety of ways and not rely only on direct proof, which may
not be simple or even possible in some instances.

The results of the error analysis using the PEET are given in Table XXXVIII. There were
only four codes applied to any of the students proofs on the pre-assessment. The code S7
was applied twice because one student tried to prove the converse, not realizing that it is not
logically equivalent, and another student merely gave several examples as a proof. Two other
students had the code U5 applied to their proofs for incorrectly claiming that one statement
implied another statement. The error code applied most on the pre-assessment was U4, which
means that students did not justify a crucial step in the proof. This code was applied to most of
the invalid proofs since on the pre-assessments students often claimed that a statement implied

<table>
<thead>
<tr>
<th>Proof Approach</th>
<th>Pre-Assessment</th>
<th>Post-Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrapositive</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Induction</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Contradiction</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Proof by Cases</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Euclid’s Theorem</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

TABLE XXXVII
A SUMMARY OF PROOF APPROACHES TAKEN BY STUDENTS IN THE TREATMENT GROUP ON EACH ASSESSMENT
that $n^2$ was divisible by 3 (the conclusion of the statement to be proven) and that claim was either untrue or unjustified.

On the post-assessment, there were many more codes applied, but that is partly a function of the fact that there were fourteen complete proofs to analyze versus only nine on the pre-assessment. This time there were several structure errors, mainly for students failing to properly define variables used in the proof, students proving the converse of the statement or using confusing notation. However, this time the most applied error code was U1, indicating that a student wrote a statement that was not fully explained. Half of these U1 codes were actually applied to valid or weak valid proofs, as many students did not define variables before they used them in their proofs.

The fact that many more errors were applied to proof on the post-assessment and the fact that every valid or weak valid proof written on the post-assessment had at least one error code
applied goes along with a result from the last section. Almost all of the students in the treatment group identified the informal proof as valid on the proof validation task from the assessment. This implies that these students are more concerned with the details of an argument than the fact that the argument is written out rigorously. The results here correlate with that, in that students did not see a real need to write things out carefully in their own proofs, resulting in many structure errors and many statements written without proper justification.

5.3.2.2 Comparison Group Results

Now a report of the proofs written on the assessments by the comparison group and their classifications as valid, weak valid, invalid or incomplete will be given. This is followed by another discussion of proof techniques used and the results from analysis using the PEET.

The types of proofs (valid, weak valid, invalid or incomplete) constructed by the students in the comparison group on the pre and post-assessments are summarized in Table XXXIX. As in the treatment group, no student in the comparison group was able to write a valid proof of the statement on the pre-assessment. Furthermore, eight students in the comparison group were not able to write any proof at all on the pre-assessment. Although there was a little improvement on the post-assessment, still only 25% of the students were able to write a valid proof and 42% of students were not able to complete a proof.

When considering proof techniques used on the pre-assessment and post-assessment by students in the comparison group, we see little change from the pre to post-assessment. The types of approaches employed by students in the comparison group on the pre and post-assessments are given in Table XL. On the pre-assessment, only one student attempted anything other
than direct proof by trying to use induction. Two other students wrote down the base case of induction before becoming stuck and being unable to finish the proof. On the post-assessment, two students used proof by cases and one student proved the contrapositive of the statement. It is important to note that these three students who tried an alternative approach were the only three students that were able to write valid proofs.

<table>
<thead>
<tr>
<th>Proof Approach</th>
<th>Pre-Assessment</th>
<th>Post-Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contrapositive</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Induction</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Contradiction</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Proof by Cases</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Euclid’s Theorem</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

TABLE XL
A SUMMARY OF PROOF APPROACHES TAKEN BY STUDENTS IN THE TREATMENT GROUP ON EACH ASSESSMENT
The frequency of error codes from the PEET applied to students in the comparison group on the proof construction task from the pre and post-assessments are summarized in Table XLI. The error codes on both the pre and post-assessment were split mostly evenly between structure errors and understanding errors. Students in the comparison group had several codes applied that were not present for the proofs in the treatment group. For example, no proofs from the treatment group received the code S2, which is applied when a false assumption is made in the proof. Additionally, the code S3, applied when the proof is not written linearly, and U6, applied when a proof only includes some of the required cases, did not show up in the treatment group but did appear in the comparison group. The codes applied were more or less spread out amongst the categories given with no single code being applied the most.

<table>
<thead>
<tr>
<th>Proof Errors</th>
<th>Pre-Assessment</th>
<th>Post-Assessment</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>S2</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>S3</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>S4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>S7</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>S8</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>U1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>U4</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>U5</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>U6</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>10</strong></td>
<td><strong>14</strong></td>
</tr>
</tbody>
</table>

TABLE XLI

FREQUENCY OF COMPARISON GROUP STUDENTS’ PROOF ERRORS ON THE ASSESSMENTS SEPARATED BY ERROR CODE
5.3.2.3 Between Groups Results

The goal of this section is to compare the results between the treatment and comparison
groups and summarize what has been learned about the impact of collaborative revision on
students’ proof construction skills. First, a comparison of the number of valid proofs written
and the different approaches taken in the two groups will be shown. Then, the results of the
error analysis using the PEET will be contrasted between the treatment and comparison group.
Finally, a report of the improvements made on the proof construction task from the pre to the
post-assessment will be given and juxtaposed between the two groups.

As previously discussed, on the pre-assessment, neither group had any students that were
able to write a completely valid proof of the given statement. The treatment group did have
one student who was able to write a weak valid proof using contradiction. This provides further
evidence that at the beginning of the course the two groups were similar in abilities. However,
on the post-assessment almost half of the students in the treatment group (7 out of 15) were
able to write a valid or weak valid proof, while only a quarter of students in the comparison
group (3 out of 12) were able to do the same. Perhaps more concerning is that while only one
student in the treatment group was not able to complete any proof on the post-assessment, five
out of 12 (42%) of the students in the comparison group were not able to write any proof on the
post-assessment. This gives some evidence that participation in collaborative revision does in
fact aid students in becoming better proof writers. These findings suggest that students have
the confidence and ability to at least make an argument, even though it may not be a valid one.
More evidence that collaborative revision has an impact on proof construction skills comes when we look at the varied proof approaches that students in the treatment course attempted on the post-assessment compared to those in the comparison course. A quarter of the students in the comparison course used a proof technique other than direct proof on the post-assessment and these were the same quarter of students that were able to write a valid proof. However, almost half of the students in the treatment course used a proof technique alternative to direct proof. Thus, the treatment group students seem more comfortable with varied proof techniques and tried to use them on this proof construction task, though some were applied unsuccessfully. This is important because the statement given to prove is actually much easier when attempted using either the contrastive or contradiction and is very straightforward if proof by cases is used. It is much harder to prove directly, which could explain the large number of invalid proofs on the post-assessment from students in both groups.

If the codes applied to the proofs in both groups are separated into (S)tructure and (U)nderstanding categories, as they are in Table XLII, then we can observe what types of errors each group is making. Note that both on the pre-assessment and on the post-assessment, both groups made roughly the same amount of structure errors. These errors affect mainly the setup of the proof and do not usually affect the underlying argument. In fact, on the pre-assessment both groups made roughly the same total amount of errors. However, on the post-assessment the treatment group made twice as many errors (28) than the students in the comparison group (14) and most of these were understanding errors. A possible reason is that there were many more completed proofs in the treatment group than in the comparison group.
AN ANALYSIS OF STUDENTS’ PROOF ERRORS MADE ON THE ASSESSMENTS SEPARATED BY ERROR CATEGORY

(14 vs. 7) and since error codes were applied to all proofs, even valid ones, this would affect the total number of error codes applied.

The last difference between groups that will be discussed here is the number of improvements made from the pre to the post-assessment when we analyze student-by-student. An improvement is defined by moving from what we consider a lower classification to a higher one. For instance, if a student wrote an invalid proof on the pre-assessment, but a valid proof on the post-assessment, we would consider that an improvement. Also, if a student wrote an incomplete proof on the pre-assessment but was able to write a invalid proof on the post-assessment that is considered an improvement as well. A summary of the types of improvements made and the frequency for each group is given in Table XLIII. When we look at the number of improvements made in each group it emerges again that collaborative revision is beneficial to students’ proof writing skills. Two-thirds of the students in the treatment group (10 out of 15)
TABLE XLIII

TYPES OF IMPROVEMENTS MADE ON STUDENTS’ PROOFS SEPARATED BY GROUP

<table>
<thead>
<tr>
<th>Improvement</th>
<th>Treatment</th>
<th>Comparison</th>
</tr>
</thead>
<tbody>
<tr>
<td>Incomplete → Invalid</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Incomplete → Weak Valid</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Invalid → Weak Valid</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Incomplete → Valid</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Invalid → Valid</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Weak Valid→ Valid</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>10</strong></td>
<td><strong>5</strong></td>
</tr>
</tbody>
</table>

were able to make an improvement from the pre-assessment to the post-assessment while only 42% of the students in the comparison group (5 out of 12) were able to make any improvement.

Thus, it seems that participation in collaborative revision does have an impact on students’ abilities to write valid proofs. Students in the treatment group were able to write a greater number of valid proofs on the post-assessment than students in the comparison group and tried a wider variety of approaches. Additionally, students in the the treatment group made more improvements from the pre to the post-assessment in the kind of proof they were able to write than students in the comparison group.

5.3.3 Synthesis of Results on Students’ Proof Construction Skills

From the analysis of our data, it was found that no student in either the treatment or comparison course was able to write a fully valid proof on the pre-assessment given in this study. On the post-assessment it was found that about half of the students in the treatment
group were able to write acceptable (valid or weak valid) proofs compared to only a quarter of the students in the comparison group. These results are consistent with other research (Moore, 1994; Dreyfus, 1999; Weber, 2001) that even upon completing an introduction to proof course, the majority of students are still unable to construct valid proofs.

Nevertheless, these results do show learning gains for the treatment group over the comparison group. The treatment group students were not only able to construct a greater number of valid or weak valid proofs on the post-assessment than the comparison group, they also used a wider variety of proof techniques (i.e. contradiction, contrapositive, induction, etc.) to do so. The percentage of treatment group students (67%) that made an improvement on their constructed proof (i.e. from invalid to valid or from incomplete to invalid) was also greater than the percentage of students who improved from the comparison group (42%). Taken together, these results indicate that the answer to our third research question is that collaborative revision does positively impact students’ ability to construct valid proofs.

When the errors in the student proofs were analyzed using the PEET, it was found that, while the two groups made approximately equal numbers of errors on the pre-assessment, students in the treatment group made far more errors on the post-assessment. Of the errors made on the post-assessment by students in the treatment group, the majority were understanding errors. This is most likely due to the fact that the treatment group wrote a much greater number of complete proofs on the post-assessment. Five of the students in the comparison group were unable to write a complete proof on the post-assessment whereas only one student in the
comparison group was unable to. Again, this provides evidence that collaborative revision can aid students’ proof-writing skills.

5.4 Conclusion

This chapter gave the data analysis used and the results obtained to answer each of the three research questions in this study. It was found that collaborative revision may be an effective way to introduce students to some of the other functions of proof besides verification and explanation. Also, collaborative revision may aid students’ in proof construction skills. While collaborative revision group students did report higher understanding of the arguments presented to them and greater confidence about their classifications of arguments, no gains in proof validation skills were found. In the next chapter, we will summarize the results presented in this chapter as well as discuss the limitations of this study. Additionally, we will discuss some implications from this study for classroom teaching and give some directions for future research.
CHAPTER 6

DISCUSSION

This study was conducted to determine to what extent collaborative revision may impact students’ proving skills and beliefs about the function of proof. Information was desired about undergraduate students’ proof validation and proof construction abilities before and after an introduction to proof course and if participation in collaborative revision affected these skills. To achieve this goal, this study and the teaching experiment component was designed to answer the research questions that were introduced in Chapter 1. These questions were:

1. (a) What are undergraduate students beliefs about the functions of proof in mathematics before and after a transition to proof course?
   (b) Are there differences present in undergraduate students’ beliefs about proof function between students who participated in collaborative revision and those who did not? If so, what are they?

2. (a) What are undergraduate students’ proof validation skills before and after a transition to proof course?
   (b) Are there differences present in proof validation skills between students who participated in collaborative revision and those who did not? If so, what are they?

3. (a) What are undergraduate students’ proof construction skills before and after a transition to proof course?
(b) Are there differences present in proof construction skills between students who participated in collaborative revision and those who did not? If so, what are they?

The data collected from assessments, interviews and observations provided a measure of students’ proving skills and allowed for a comparison to be made between students who participated in collaborative revision and students who did not. Student participants were drawn from a treatment and a comparison course and were given a pre and post-assessment and had their classrooms observed. The treatment group participated in a teaching experiment called collaborative revision, which required the students to present proofs that were written outside of class to other students who were responsible for providing notes and comments to aid in revising the proofs.

6.1 Summary of Results

This study expanded on previous studies (e.g. Almeida, 1995; Weber, 2010) and measured students’ appreciation of nine functions of proof identified by mathematics education researchers in the literature. It was found that, overall, the participants in this study displayed a diverse appreciation of the functions of proof even on the pre-assessment, which was administered at the beginning of the transition to proof course. Students also identified more functions of proof, both in their responses to an open response question and by rating their agreement with proof function statements on a Likert scale, on the post-assessment than was found in previous studies (Almeida, 1995; Coe and Ruthven, 1994; Knuth, 2002a) measuring undergraduate students’ and in-service teachers’ beliefs about the function of proof.
At the start of the study, as measured by the pre-assessment, students in the treatment and comparison groups displayed equal levels of appreciation for the functions of proof from the literature, as measured by the pre-assessment, and there were no significant differences between groups on any of the eight Likert items. However, on the post-assessment, there was a divergence in proof function beliefs between the two groups. Students in the treatment group rated the explanation function of proof significantly higher on the post-assessment than the pre-assessment. Since students in the treatment group were very often explaining their proofs and arguments to other students it is plausible that students garnered a greater appreciation for the explanation function of proof directly from their participation in the collaborative revision process. On the other hand, students in the comparison group rated the communication function of proof significantly lower on the post-assessment than they had on the pre-assessment. Again, this may be a result of the type of learning environment they encountered, which is one that utilized direct instruction.

When comparing the two groups, it was found that the treatment group rated the functions of explanation, intellectual challenge, communication and providing autonomy significantly higher on the post-assessment than the treatment group, despite no significant differences on the pre-assessment. These proof functions are emphasized, albeit implicitly, in the collaborative revision process so this result was encouraging. Students in the comparison course had significant correlation coefficients on most of the Likert items when a paired-samples t-test was performed, while the treatment group had a significant correlation coefficient on only one item. This indicates that the beliefs of the comparison group about proof functions were rigid and
largely unchanged at the end of the study yet the treatment group had dynamic beliefs and tended to agree more with each statement on the post-assessment. These results suggest that collaborative revision could be a teaching tool that strengthens students’ understanding of some of the functions of proof from the literature.

When analyzing data from the proof validation task on the pre and post-assessment, it was found that students in both the treatment and comparison group were not proficient at identifying valid proofs, which is consistent with previous studies (Selden and Selden, 2003; Alcock and Weber, 2005; Brandt and Rimmasch, 2012) regarding undergraduate students proof validation skills. Moreover, there were no significant increases in the average percentage of correctly classified proofs from the pre-assessment to the post-assessment in either the treatment or comparison group. An additional concerning result is that, even though the average correct percentage for students in the treatment group did not increase at all from the pre to the post-assessment, the students’ self-reported understanding of each of the proofs did significantly increase as did the self-reported confidence of classification. Thus, students in the treatment group were not making gains in correctly identifying valid proofs yet were reporting a greater certainty about the proofs themselves. This implies that while collaborative revision may increase students’ confidence about approaching proofs it does not seem to enhance their proof validation skills.

Data collected from interviews with students in the treatment and comparison groups yielded results about the proof schemes that these students operated under. Although only a small number of students were interviewed, it seems that the students in the treatment group
were able to ascend to higher-level proof schemes throughout the semester, while students in the comparison group remained at roughly the same level. Both of the students in the treatment group exhibited much higher-level proof schemes during the second interview than they had on the first interview, even one student who began the semester with a fairly high-level proof scheme. Of the students interviewed from the comparison group, one did not transform her proof scheme at all and the other just went one level higher. Thus, while collaborative revision may not aid in developing students abilities to identify valid proofs, it may impact the way in which students gain conviction about an argument. More research is needed to clarify this distinction.

Collaborative revision may impact students’ proof construction skills, however. On the pre-assessment, only one student from the treatment group was able to write an acceptable (valid or weak valid) proof of the given statement. Also, none of the students in the comparison group were able to write an acceptable proof on the pre-assessment. Yet seven students (47%) in the treatment group were able to write an acceptable (valid or weak valid) proof on the post-assessment, compared to only three students (25%) in the comparison group. Also, five of the students in the comparison group were not able to write any proof (valid or invalid) on the post-assessment compared to only one student in the treatment group. Furthermore, students in the treatment group used a much wider variety of approaches (i.e. contradiction, contrapositive, Euclid’s Theorem, induction, etc.) in their proof on the post-assessment than students in the comparison group. Thus, the treatment group displayed a greater competency when writing proofs of this statement on the post-assessment and showed they were comfortable with varied
approaches and proof techniques. Finally, more students in the treatment group improved in their proof writing (i.e. invalid to valid, incomplete to invalid, etc.) from the pre-assessment to the post-assessment indicating that collaborative revision may have an impact on students’ proof construction abilities.

Together these results suggest that collaborative revision may have benefits when teaching students about proof. Namely, students who participate in collaborative revision seem to have stronger proof construction skills and more diverse beliefs about the functions of proof. Further studies are needed to determine to what extent collaborative revision impacts proof validation skills.

6.2 Limitations

The sampling technique used in this study was pragmatic and convenient, which could have several limitations. Using the Emerging Scholars Program course students to form the treatment group meant that students were self-selecting to be in this group by their enrollment. Even though any student can register for any ESP course, it is an additional one-credit course that students take so it is possible that only highly motivated students or students that are already amenable to collaborative work were enrolled. This study addressed this concern by the collection of several types of data from students in the treatment group. First, GPA data and prior mathematics courses taken were collected on all student participants. Independent samples t-tests on this data indicated that there were no significant differences in either the overall GPA or the mathematics GPA of students in the treatment and comparison group. Additionally, roughly the same number of students in each group had taken prior proof-based
mathematics course indicating comparable mathematical maturity in both groups. A survey, found in Appendix E was also administered to the students in the treatment course to determine if they had prior experience in the Emerging Scholars Program and their thoughts about collaborative learning. Out of the 15 students, only two had previously taken an Emerging Scholars course and only one student preferred to do mathematics in a group and two more mentioned they preferred to work in a group setting only after first working on problems by themselves. Thus, we can determine that, as a whole, the students in the treatment group were not more inclined to prefer collaborative learning nor were they superior students in terms of GPA or mathematical maturity. Results from the pre-assessment corroborated these results as students in the treatment and comparison groups were very evenly matched on all parts of the pre-assessment.

The unique format of the treatment course also presents some limitations for this study. First, the twelve students that were enrolled in both the treatment and comparison course effectively had an additional hour and fifty minutes of classroom time per week, making it hard to attribute learning gains solely to collaborative revision. However, it is important to note that this classroom time was spent only minimally on instruction and instead on discourse between students. Thus, these students are not receiving more instruction (any instruction given was designed either as review or to present definitions needed for the proofs) than students in the comparison group, but given the extra time for collaborative revision. Secondly, the majority of the students in the experimental course were also enrolled in the comparison course, so they had exposure to both types of learning environment. Even though students were
exposed to a lecture-based teaching of the material, the exposure to this highly collaborative and communicative method of teaching still affected their beliefs about proof and their ability to write valid proofs.

Another study limitation is that the revision component of the collaborative revision process occurred in a way that was not faithful to the study design. While the students did revise their proofs, this tended to be much more hurried than originally intended. Students often completed their discussion and revisions within a single class period, which resulted in the submittal of far fewer drafts of proof than anticipated. This was mostly a function of a long class period, which allowed the students to have in-depth discussions. However, students were still participating in a form of the collaborative revision process. Thus, learning gains can still be attributed to collaborative revision even though it is possible that the collaboration aspect may have played a larger part than the revision aspect.

The final limitation of this study is its fairly small sample size. Sample size was constrained by the number of students enrolled in the courses and by their willingness to participate in the study. Also, many students that took the pre-assessment did not take the post-assessment due to absence or withdrawal from class. Thus, the data reflected in this report is only from students who participated in the entire study (i.e. both assessments or both interviews). A greater number of student participants would have been preferred to gain a better insight into students’ beliefs about proof functions and proof construction and validation skills. The first interview conducted had six participants, three from each group, however I was unable to schedule second interviews with two of the students (one from each group). It would have been
desirable to have more interview participants to better determine how collaborative revision impacts students’ proof schemes.

6.2.1 **Proof Portfolio Limitations**

To understand the proof-writing skills of the treatment group throughout the semester, each student in the treatment course was required to submit first and subsequent drafts of proofs to me for inclusion into a proof portfolio. The portfolios were designed not only to evaluate if collaborative revision has any impact on the treatment group’s proof construction skills, but also to provide a way to document the process of collaborative revision. However, the proof portfolios as designed in this study did not furnish much meaningful data for assessment of the treatment groups proof-writing skills during the semester. This was due to several unforeseen difficulties that arose when implementing the proof portfolios as an evaluation tool. A discussion of the difficulties encountered and the possible reasons for these are given below.

The first complication that arose in using the proof portfolios is that even though students were instructed to come to class with a proof of the statement already written in order to have something to discuss with the other students this did not always occur. Each week at least one student arrived at class unable to complete a proof of the statement that had been given to him or her. This then made it difficult for that person to effectively participate in the group discussion as they had nothing to present. Therefore, when all the other students in a group had presented their proofs I often instructed groups to assist students who were unable to complete their proof. The student who did not write a complete proof was asked to share their statement with the other members of the group and to explain to them what they had attempted thus
far (i.e. an induction proof, worked out some examples, found an applicable definition, etc.). Then the group as a whole would attempt to work out a proof of the statement.

While the group working together on a proof of a statement was usually able to complete a valid proof, this resulted in poor data for the study. After the group members decided they had proved the statement, it was usually written up in a group manner, with one student writing and the other group members helping him or her with wording. Several times a particularly sharp group member would practically write a whole proof while the group was working together and then when the other student would write it up for his or her proof portfolio it would essentially be copied from that group member. Thus, it became impossible at times to discern which part of a proof was written by the student and which part was written by the group.

The second challenge encountered was that students often had very few comments for other students about their proofs. When a student presented a proof, often the other students would have little to say about it, which did not aid the student in making revisions. This was much more prevalent at the beginning of the semester and required numerous interventions from the instructor. Often if students did not say much after the presentation, I as the instructor would pick a more difficult part of the proof and implore “can you explain this part?” to one of the other students in the group. After the students became more familiar with the process and what was expected of them, this happened less frequently, but as I was unable to observe every group all of the time sometimes students would be left with very few comments to direct their revision and would write up their proof as is.
Third, since the class met for one hour and fifty minutes each time, students were often able to complete their revisions during one class period. This was counter to the original design of the collaborative revision process, in which students were supposed to bring proofs to class, attain comments from the other students, work on revising the proof outside of class, and bring it back in and present it again. Since students were presenting and revising within a single class session resulted in very few separate submissions of a single proof (i.e. first draft, second draft, etc.) to the proof portfolio. In fact, we were only able to collect at most one revision per proof for each student and, thus, it was difficult to assess the revision process in this study. In the majority of the proofs that were submitted to the portfolios the revisions manifested as additions or deletions to the original proof. This was in the form of carats between words to add words or phrases, addition of sentences or phrases to further explain arguments, and crossing out of words or statements. It became impossible to then determine when the revisions were made and to discern where one draft ended and another began.

The data analysis plan set forth in the proposal for this project was to analyze these proof portfolios for instances of errors made in each draft and apply the error codes from the PEET as described in Section 5.3.1. These codes were to be analyzed in two ways to further answer the third research question. First, I would count the raw number of errors for each version of each proof. The purpose of this was to assess the efficacy of the collaborative revision process in boosting students’ proof construction skills by determining if there was a decrease throughout the semester in the number of errors they made when constructing a proof. Codes were also to be separated into error categories of either structural errors or understanding errors. The
rationale for this was to determine if a change in the number of errors of each type was detected throughout the semester. The type of errors made is important and structural errors tend to denote lack of comprehension about how to write a proof, whereas understanding errors betray a lack of comprehension of the underlying logical ideas behind writing a proof. However, this data analysis plan became impossible to enact because of a lack of useable data from the proof portfolios.

The difficulties outlined above highlight the problems with trying to assess the revision process from the proof portfolios. Thus, the way that the proof portfolios had been envisioned to collect data was not successful during this study. However, the process the students were participating in, independent and group proof-writing and constantly evaluating arguments for validity, which we believe is the essence of collaborative revision, was not comprised. Additionally, separate results about the treatment group student’s proof construction skills were obtained from the assessment so we have a measure of how collaborative revision impacts these skills. Future studies of collaborative revision or classroom that wish to use it for assessment purposes need to create a new way to document the process.

6.3 Teaching Implications

The results from this study expose the areas where collaborative revision enhances students’ learning and understanding, which can be useful for university instructors and for designing learning environments. What follows are some implications for classroom teaching that emerged from the results presented in Chapter 5.
As outlined in Chapter 3, it is important for students to understand the diverse functions of proof and not to be misled into believing that proof is only for verification purposes (Hanna, 1990; Weber, 2002; deVilliers, 2012; Almeida, 1995). Understanding the functions of proof is important for students to retain meaning from proof and avoid seeing it as a pedantic activity. deVilliers (2002) writes that “rather than one-sidedly trying to focus on proof only as a means of verification in dynamic geometry (which does not make sense to novice students), the more fundamental function of explanation and discovery ought to be initially utilized to introduce proof as a meaningful activity to students” (pg. 11). The results from Section 5.1 imply that collaborative revision can enhance students’ views about proof possessing an explanation function. Furthermore, Yackel and Cobb (1996) argue that “the development of intellectual and social autonomy is a major goal in the current educational reform movement, more generally, and in the reform movement in mathematics education in particular” (pg. 473). Collaborative revision was shown in this study to enhance students appreciation of proof as giving mathematical autonomy to students and could be used to achieve these reform goals. Thus, collaborative revision, with its proven impacts on students beliefs about proof functions, could be utilized to enhance students’ appreciation of specific proof functions.

Although this study concerns collaborative revision as a teaching method, it is important to note that it was designed to supplement and enhance traditional instruction. In the context of this study, students in the treatment group were receiving direct instruction about different proof methods and mathematical concepts in their lecture-based course and were able to employ collaborative revision in the treatment course. Whereas there was minimal lecture given in the
treatment course it was primarily intended to make sure all the students were aware of the
concepts for that week and had access to any definitions needed to complete the proofs. Thus,
it was relied upon that examples of proof techniques, such as induction and contradiction, as
well as in-depth instruction on concepts needed, such as functions and set theory, were presented
in the comparison course. It would have been almost impossible to introduce this material and
proof techniques as well as allow enough time for the collaborative revision process to take place
in the two hours per week that the treatment group students attended class.

Thus, collaborative revision is intended as a much-need supplement to a lecture-based class-
room. It is envisioned that for a class that meets three days a week, two days a week could be
spent on introducing concepts and methods and the remaining day could be a time to imple-
ment collaborative revision. It would be very hard to arm students with all the tools needed to
attack the proofs without spending some time explaining concepts and giving examples. It is
not meant to be implied that collaborative revision can only be auxiliary to lecture; I believe
it would also fit well into an inquiry-based learning classroom. The principles of collaborative
revision are akin to those in IBL and it would be manageable to implement a collaborative
revision process into an IBL course. Instead of a whole-group discussion about proofs, students
could be divided into small groups for discussion and proofs could be graded only after several
revision iterations have been completed. Supplemental instruction may still be needed; for in-
stance, it is very difficult to have students discover induction independently and some ancillary
instruction is needed even in an IBL classroom (M. Mocanasu, personal communication, March
23, 2013).
Just as collaborative revision could be integrated into a variety of learning environments, it could also be implemented in many different levels. Any course that uses proofs could have a collaborative revision component, including high school geometry. Even in the context of two-column proofs used in geometry, students could have a chance to discuss their arguments and simultaneously learn about what constitutes a valid proof for that level. Collaborative revision could also be incorporated into nearly any proof-based course, such as linear algebra or analysis. Since it is designed to supplement instruction, the instructor could decide how often and for which concepts they believe collaborative revision would be useful and implement it wholly or piece-meal. Explaining arguments and carefully reading and critiquing others’ arguments could be beneficial for any student learning how to prove. Collaborative revision could prove to be especially useful for pre-service teachers. Knuth (2002a) writes that “it seems clear that if teachers are to be successful in enhancing the role of proof in secondary school mathematics classrooms, then their conceptions of proof must be enhanced” (pg. 403).

A final implication for classroom teaching is that collaborative revision, according to these results, aids somewhat in proof construction skills for students and helps to build confidence about proof reading. Students in the treatment group in this study reported a greater understanding of the proof validation task arguments as well as a greater confidence about their classifications than students in the treatment group. Thus, collaborative revision is responsible for students feeling comfortable when approaching proofs. It should be cautioned that separate instruction is probably needed on the aspects of a valid proof as the proof validation skills of students in both the treatment and comparison group were very poor.
6.4 Further Research

As with any study, we are bound by the limitations of time and scope. Additionally, several negative results obtained in this study need to be explored further to determine exactly how collaborative revision is beneficial. This leaves many directions for future research from this study that can be addressed by other researchers. Below we give suggestions for future directions that studies could take as well as some research questions that could be explored.

First, it would be beneficial to determine if individual student’s beliefs about the functions of proof has an impact on their ability to write a valid proof. This study addressed these two components separately, asking students about proof function and then asking them to construct a proof of a statement, but a differently designed assessment could link these attributes. Perhaps if a student has a low opinion of the explanation function of proof, say, that could affect the type of proofs they write. A guiding research question for a future study could be: Do the beliefs that students have regarding the functions of proof impact their ability to write a valid proof or the proof techniques they use?

Second, there is a lot of room for studies that enact collaborative revision in a way different from this study. In the previous section, it is mentioned that collaborative revision could be implemented in a variety of different courses and could comprise the majority of class time or only a small part. Thus, future studies could investigate the impacts on students’ proof construction and validation skills when collaborative revision is integrated into a classroom in diverse ways. Also, the generalizability of these results to students at other levels needs to be studied. For example, it is suggested in the previous section that collaborative revision could
be integrated into a high school geometry course or into a course for pre-service teachers. What are the impacts in these classrooms of collaborative revision and how do they compare with the results presented here for undergraduates? It is important to note that a new definition of proof would have to be operated under in these types of studies since proof in high school geometry and proof for pre-service teachers has a different meaning than proof for undergraduate students in a transition to proof course.

Third, besides implementation into other types of classrooms, future studies could concentrate on using a similar design but a larger sample size. Studies that used a proper control group (perhaps one group receives direct instruction, but participates in collaborative revision once a week, while another group just receives direct instruction) could build on the results in this study and obtain more definitive answers to the research questions presented here. Also, a study that interviewed more students could attain enhanced results about students’ proof schemes and determine if collaborative revision does impact the proof scheme levels of every student.

Fourth, this study focused on the impacts of collaborative revision, specifically on proof construction and validation skills and beliefs about the functions of proof. However, a theoretical investigation of exactly which aspects of the collaborative revision process assist in students’ proving skills in how would be a good companion to this study. Future work could concentrate on determining how collaborative revision affects proving skills and how students interact with one another during the process.
Finally, there are very few studies that focus on revision of proofs in the classroom. Even if
the instructor was giving direction to the students on revisions that need to be made, there is a
need for more studies about how prolonged reflection on proofs by students enhances learning. I
would hypothesize that even minimal revision can be beneficial since, when first learning proof,
it can be difficult to submit a perfect finished product the first time and consideration of one’s
mistakes can augment understanding.

6.5 Conclusion

We conclude with this quote from Alibert and Thomas (1991) that gives insight into the
usefulness of collaborative revision: “The context in which students meet proofs in mathematics
may greatly influence their perception of the value of proof. By establishing an environment in
which students may see and experience first-hand what is necessary for them to convince others,
of the truth or falsehood of propositions, proof becomes an instrument of personal value which
they will be happier to use in the future” (pg. 230). Collaborative revision provides a way to
introduce proofs that may enhance students’ appreciation of the functions of proof, strengthen
their proof construction skills, and help students become comfortable in approaching proofs.


CITED LITERATURE (Continued)


CITED LITERATURE (Continued)


Sfard, A. (2001). There is more to discourse than meets the ears: Looking at thinking as communicating to learn more about mathematical learning. *Educational studies in mathematics* 46, 13–57.


APPENDICES
Appendix A

ASSESSMENT

Part I.
What do you believe is the purpose of proof in mathematics?

Circle the number that corresponds to how strongly you agree or disagree with the given statement. Circle 1 if you strongly disagree, 2 if you slightly disagree, 3 if you are neutral, 4 if you slightly agree and 5 if you strongly agree.

1. I believe that proofs are very important in the field of mathematics.

   strongly disagree  1  2  3  4  5  strongly agree

2. I believe that proofs are only done to verify whether a given statement is true or false.

   strongly disagree  1  2  3  4  5  strongly agree

3. A proof can explain why a given statement is true.

   strongly disagree  1  2  3  4  5  strongly agree

4. I enjoy writing proofs because they are intellectually challenging.

   strongly disagree  1  2  3  4  5  strongly agree
Appendix A (Continued)

5. I believe that proofs are an important communication tool in mathematics.

   strongly disagree  1  2  3  4  5 strongly agree

6. I believe that proving can be used to discover new mathematics.

   strongly disagree  1  2  3  4  5 strongly agree

7. Proofs are important in organizing mathematical knowledge.

   strongly disagree  1  2  3  4  5 strongly agree

8. I want to learn proof construction so I can create understand the language of mathematics and create my own proofs.

   strongly disagree  1  2  3  4  5 strongly agree

Part II.

Determine if each proof of the statement is valid or invalid by circling your answer. Then rate how well you understand the proof and how certain you are of your answer.

Statement: The product of any three consecutive integers is a multiple of 6.

Proof 1:
A multiple of 6 must have factors of 3 and 2.
If you have three consecutive numbers, one will be a multiple of 3 as every third number is 3 times a whole number.
Also, at least one number will be even and all even numbers are multiples of 2.
Thus, If you multiply the three consecutive numbers, the answer will have at least one factor of 3 and at least one factor of 2.

Proof 1: Valid Invalid
How well do you understand this proof? not at all  1  2  3 completely

How certain are you that you correctly classified the proof?
not at all  1  2  3 completely
Appendix A (Continued)

Statement: The product of any three consecutive integers is a multiple of 6.

Proof 2:
Note that:
1 · 2 · 3 = 6, which is a multiple of 6.
2 · 3 · 4 = 24, which is a multiple of 6.
3 · 4 · 5 = 60, which is a multiple of 6.
6 · 7 · 8 = 336, which is a multiple of 6.
This pattern continues so that statement is true.

Proof 2: Valid Invalid
How well do you understand this proof? not at all 1 2 3 completely
How certain are you that you correctly classified the proof?
not at all 1 2 3 completely

Proof 3:
Let x be any whole number.
The product of three consecutive numbers is 
\[ x \cdot (x + 1) \cdot (x + 2) = (x^2 + x) \cdot (x + 2) \]
\[ = x^3 + x^2 + 2x^2 + 2x = x^3 + 3x^2 + 2x. \]
Looking at the coefficients, we have 1 + 3 + 2 = 6, so the product is a multiple of 6.

Proof 3: Valid Invalid
How well do you understand this proof? not at all 1 2 3 completely
How certain are you that you correctly classified the proof?
not at all 1 2 3 completely
Appendix A (Continued)

Statement: The product of any three consecutive integers is a multiple of 6.

Proof 4:
Of the three consecutive numbers, the first number is either even, so it can be written as $2a$ for some whole number $a$, or odd, so it can be written as $2b - 1$ for some whole number $b$.

If the first number is even, then we have $2a \cdot (2a + 1) \cdot (2a + 2)$, which is even. Then if $a$ is a multiple of 3, we are done. Otherwise, if $a$ is not a multiple of 3, then $2a$ is not a multiple of 3, but either $(2a + 1)$ or $(2a + 2)$ is a multiple of 3 and we’re done.

If the first number is odd, then we have $(2b - 1) \cdot (2b) \cdot (2b + 1)$, which is even. Then if $b$ is a multiple of 3, we are done. Otherwise, if $b$ is not a multiple of 3, then $2b$ is not a multiple of 3, but either $(2b - 1)$ or $(2b + 1)$ is a multiple of 3 and we’re done.

**Proof 4:** Valid Invalid

How well do you understand this proof? not at all 1 2 3 completely

How certain are you that you correctly classified the proof?

not at all 1 2 3 completely

**Part III.**

Write a proof of the following statement.

For any positive integer $n$, if $n^2$ is divisible by 3, then $n$ is divisible by 3.
Appendix B

CLASSROOM OBSERVATION PROTOCOL

Classroom Observation Form

Date:

Instructor:

Topic:

Estimate the percentage of class time spent on:

- Lecture
- Group Work
- Whole Class Discussion
- Other

Notes on Distribution of Class Time:

Notes on the Role of the Instructor:
Appendix C

INTERVIEW GUIDES

Guide for Interview I:

1. What year of study are you in?

2. What is your major?

3. Why are you taking this class? Are you taking this class to fulfill requirements for your major or is it an elective?

4. What other math classes have you taken?

5. What do you feel is the importance of proof in mathematics?

6. (Referring to a proof in Part II of the assessment) Why did you think this proof was valid/invalid? (Will ask for several of the proofs in Part II)

7. (Referring to their proof in Part III)?

   Explain to me how your proof works.

   Why did you choose to prove the statement this way?

   How easy is it to explain your proof. Please explain.
Appendix C (Continued)

Guide for Interview II:

1. What grade do you expect to receive in your Introduction to Advanced Mathematics course? Why?

2. What do you feel is the importance of proof in mathematics and have those views changed throughout the semester?

3. What is the purpose of proof in mathematics?

4. (Referring to a proof in Part II of the assessment)
   Why did you think this proof was valid/invalid??

   What percentage of your classmates do you think would agree with your answer?

   (Will ask for several of the proofs in Part II) Why?

5. (Looking at Part II of their previous assessment, if applicable) I see you changed your answer on this question since the last assessment. Please explain why?

6. (Referring to their proof in Part III)
   Explain to me how your proof works.

   Why did you choose to prove the statement this way?

   How easy is it to explain your proof. Please explain.
Appendix D

TREATMENT COURSE SYLLABUS

Special Topics in Mathematics
Math 294, Fall Semester, 2012 – Call #23407, #25803
12:00 – 1:50 PM, Tuesdays or Thursdays in 300 Taft Hall

• Instructor: Emily Cilli-Turner
  Office: 633 SEO; Phone (312) 413-3740
  Email: ecilli2@uic.edu
  Web: http://www.math.uic.edu/~ecilli2
  Office Hours: by appointment
  Please feel free to email me any issues you would like to discuss.

• Description: This is the Emerging Scholars Program (ESP) workshop associated with Math 215: Introduction to Advanced Mathematics. In this course we will work in small groups to construct and revise proofs. You are expected to be prepared and engaged in the course, and to focus on figuring things out for yourselves and explaining them to each other.

• Grading: Your final grade is based on:
  Attendance & Participation: 30%
  Proof Portfolio: 70%

• Attendance Policy: Students are allowed two unexcused absences before it starts to affect their grade. Missing 3 class periods will result in the lowering of the students grade by 10% and missing 4–5 class periods will result in the lowering of the students grade by 20%. Missing any more than 5 class periods will result in failure of the class. Please make sure you attend class and are attentive to your classmates and participating in group discussions. Group work, comments and feedback are essential to this class, so if you are absent you are hurting your group.

• Proof Portfolios: Each student will be responsible for submitting five completely correct proofs throughout the semester. Statements to be proved will be given out by the instructor for the student to work on at home. Each week, students will bring to class a written copy of their proof of the statement and present whatever they have done on their proof to a small group. Each group will discuss each member’s proof and point out passages they don’t understand or they think are incorrect. Then, students will be responsible for revising the proof at home based on their classmates’ comments and bringing in a revised written copy the following week.

• Preparation: You are expected to come to class on time and prepared. Being prepared means having attended your regular Math 215 lecture and having completed your assigned work at home.
Appendix E

TREATMENT GROUP SURVEY

ESP Math 294 ESP Survey Fall 2012

• Name:

• Major:

• Math 215 Teacher:

• List any previous Emerging Scholars Courses you have taken:

• What do you consider as “doing mathematics”?

• How do you like to work on mathematics problems?
Appendix F

PROOF FUNCTION DATA

All of the student responses to the question “What do you believe is the purpose of proof in mathematics?” on the assessments are given in the tables below. They are separated into responses from the treatment on the pre-assessment and the post-assessment and then responses from the comparison group on the pre and post-assessment.

**Treatment Group: Pre-Assessment**

<table>
<thead>
<tr>
<th>#</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>no response</td>
</tr>
<tr>
<td>2</td>
<td>“Without proof something is true, no application of mathematics is possible. A model or a formula is useless if it has not been proven to work for its intended use. To learn how to prove things we already know (because it has been proven by others). We build the tools to discover and prove new things.”</td>
</tr>
<tr>
<td>3</td>
<td>“The purpose of proof in mathematics is to determine if a statement (i.e. equation), is true or false according to its initial assumption. It is a method to validate a problem.”</td>
</tr>
<tr>
<td>4</td>
<td>“To explain why certain statements/operation are true.”</td>
</tr>
<tr>
<td>5</td>
<td>“To give logical steps or theorems to build more math on.”</td>
</tr>
<tr>
<td>6</td>
<td>“Proofs demonstrate the truth of mathematical statements, and are essential for validating the conclusions of all branches of math. Proofs also help investigate the nature of these statements and logical methods, whether true or false.”</td>
</tr>
<tr>
<td>7</td>
<td>“To provide a clear and precise explanation of logical conclusions of patterns and relationships.”</td>
</tr>
<tr>
<td>8</td>
<td>“The purpose of proof in mathematics is that the result is proved for specified data in general and it can be used in various applications.”</td>
</tr>
<tr>
<td>9</td>
<td>“To accept or deny proposed ideas. Proving helps us advance in mathematics, allowing for greater, and more “accurate” knowledge.”</td>
</tr>
</tbody>
</table>
Appendix F (Continued)

10 “Much like a chemical mechanism, a mathematical proof serves to not only justify a statement, but also to explain the (or a possible) method by which the statement is justified. By this, I mean that a proof not only provides information regarding the truth of a statement, but also the reasoning for the assessment. (This component of why is the foundation of mathematics and logical reasoning). Conversely, proofs indicate when a statement is untrue, which in turn can help to define the boundaries and "rules" of mathematics. (Also, proofs are fun, and that’s purpose enough :)”

11 “To incorporate a logical statement into the logical hierarchy that already exists.”

12 “I believe a proof is meant to prove that a mathematical equation is valid other than through the form of an equation but rather through concepts.”

13 “To show that the things we’re working with are actually valid. It is also to understand the basics of mathematics and to see the origin of where things come from.”

14 “Purpose of proof is to show that something is true or false. Important for math as some things might be proved to be false, so it will not be used in the future.”

15 “I believe that proof is way of thinking to build up a understanding of math, especially on logical aspect.”

Treatment Group: Post-Assessment

#  Response

1 “The purpose of proof is to logically explain why and how concepts work in mathematics. Even if the concepts are initially "believable", the practice of proofs helps in understanding the harder concepts later on.”

2 “To verify conjectures we suspect to be true/false and to further understand the reasoning behind them.”

3 “To understand how or why some statement is true or false.”

4 “The purpose of a proof is to go beyond the usual "it works because it just does" and actually understand why math works the way it does.”

5 “To build on past math [and] to make new math.”

6 “I believe proof in mathematics establishes a (frequently long) connection between known or assumed knowledge in the form of axioms and knowledge found to be true dependent on these initial facts or foundations.”

7 “To create, using some consistent set of accepted axioms, an undeniable argument of some sort which is readily interpreted by the intended audience.”
Appendix F (Continued)

8  “For any statement or claim to [be] used in the future, it is necessary for it to be true generally for some conditions. For this it is important that they be formally proved using results that are proved earlier.”

9  “To methodically confirm or deny proposed ideas that could have benefits in various fields.”

10 “The purpose of proofs in mathematics is to not only assess the truth value of a statement (theorem, conjecture, etc.), but also to provide a sound mathematical logic by which the statement is assessed.”

11 “To validate a mathematical expression.”

12 “A proof in mathematics is used to show the validity of a statement by use of theorems and other various laws.”

13 “To understand where concepts in mathematics are derived from and why they work the way that they do. Proofs are also important for finding new mathematical ideas.”

14 “Proofs show that there are solutions to the problems. Without proofs we wouldn’t be able to solve some of the problems.”

15 “To understand the basic idea behind the theorem and other principles. On the other hand, proof can help building logical construction in people’s mind.”

**Comparison Group: Pre-Assessment**

<table>
<thead>
<tr>
<th>#</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>“To understand how to prove or disprove with empirical evidence certain formulas and mathematical theories.”</td>
</tr>
<tr>
<td>2</td>
<td>“To understand the concepts behind the formulas and such that we often take for granted. Also, to be forced to think more critically and use a part of our mind that isn’t always required in math classes.”</td>
</tr>
<tr>
<td>3</td>
<td>“The purpose is to understand better how a formula or concept came from.”</td>
</tr>
<tr>
<td>4</td>
<td>“To establish exactly what we can and cannot know for sure and what conclusions we can draw.”</td>
</tr>
<tr>
<td>5</td>
<td>“Use the methods we know or we don’t know yet to prove the mathematical concepts and other methods. To prove there are connections among concepts. To make students understand deeper about the concepts they have learned before.”</td>
</tr>
<tr>
<td>6</td>
<td>“To gain a better understanding of what we do and don’t know. Also, to gain the ability to logically think about and solve a problem.”</td>
</tr>
</tbody>
</table>
Appendix F (Continued)

<table>
<thead>
<tr>
<th>#</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
<td>“To get a step by step logical conclusion to a problem. Also to develop and expand logic skills.”</td>
</tr>
<tr>
<td>8</td>
<td>“They challenge intellectually. They also expand a knowledge from what the specific statements came from and that there might be some new statement evolved.”</td>
</tr>
<tr>
<td>9</td>
<td>“To understand how theorems are developed, to use known facts and discover new insight into mathematics.”</td>
</tr>
<tr>
<td>10</td>
<td>“To derive conclusions or further information from earlier information which has been assumed or already proven. To further mathematical knowledge and understanding.”</td>
</tr>
<tr>
<td>11</td>
<td>“The purpose of proof is to truly understand the process, from start to finish, of solving problems.”</td>
</tr>
<tr>
<td>12</td>
<td>“To deduce complex rules from simple ones, some complex rules yield useful techniques others yield some desired result, true on the basis of reason for any variables that fit w/in the context, an invaluable tool for consistent abstract thought w/ itself and optionally w/ physical reality.”</td>
</tr>
</tbody>
</table>

Comparison Group: Post-Assessment

<table>
<thead>
<tr>
<th>#</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>“To show that without a reasonable doubt that we know the solution.”</td>
</tr>
<tr>
<td>2</td>
<td>“To establish [the] rules and show the logic that math is built off of. To [foster] understanding and break away from the notion that math is about numbers.”</td>
</tr>
<tr>
<td>3</td>
<td>“The purpose of proofs in mathematics is to understand better a concept taught or simply just prove if something is true or false.”</td>
</tr>
<tr>
<td>4</td>
<td>“To concretely say that a hypothesis is true or untrue based on facts and logic.”</td>
</tr>
<tr>
<td>5</td>
<td>“To let students understand the mathematical material better and grasp the basic knowledge how the math works in this field.”</td>
</tr>
<tr>
<td>6</td>
<td>“To better understand what we believe and try to expand on what we know.”</td>
</tr>
<tr>
<td>7</td>
<td>“To help establish a logical step to show how someone could solve an unknown using what is known.”</td>
</tr>
<tr>
<td>8</td>
<td>“To explain why a given statement is true or false.”</td>
</tr>
<tr>
<td>9</td>
<td>“To understand the foundation of mathematics, also to remove doubt of certain truths such as theorems. And, if you doubt, then proceed to disprove by the many tools that have been used to prove and disprove mathematical statements.”</td>
</tr>
</tbody>
</table>
10 “To show the truth of mathematical statements and explore methods of reasoning. Proofs can push math forward into new fields, and can often be beautiful.”

11 “Explain reasoning and evidence why certain theorems/processes work.”

12 “A language difficult for both humans and computers to understand but it’s somehow fundamental – therefore useful so we bother.”

The following two tables show how the responses were coded into the categories. Each code is given along with the sentence or phrase from students that had that code applied to it. If a sentence or phrase has a (T) next to it that was a comment from a student in the treatment group and if it has a (C) next to it that was a comment from the comparison group.

**Pre-Assessment Codes**

<table>
<thead>
<tr>
<th>Code Applied</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>• “Without proof something is true, no application of mathematics is possible.” (T)</td>
<td></td>
</tr>
<tr>
<td>• “…determine if a statement (i.e. equation), is true or false.” (T)</td>
<td></td>
</tr>
<tr>
<td>• “It is a method to validate a problem.” (T)</td>
<td></td>
</tr>
<tr>
<td>• “Proofs demonstrate the truth of mathematical statements…” (T)</td>
<td></td>
</tr>
<tr>
<td>• “…the result is proved for specified data in general and it can be used in various applications.” (T)</td>
<td></td>
</tr>
<tr>
<td>• “To accept or deny proposed ideas.” (T)</td>
<td></td>
</tr>
<tr>
<td>• “To understand how to prove or disprove with empirical evidence certain formulas and mathematical theories.” (C)</td>
<td></td>
</tr>
<tr>
<td>• “…a mathematical proof serves to not only justify a statement, but also to explain the (or a possible) method by which the statement is justified.” (T)</td>
<td></td>
</tr>
<tr>
<td>• “To establish exactly what we can and cannot know for sure and what conclusions we can draw.” (C)</td>
<td></td>
</tr>
<tr>
<td>• “To gain a better understanding of what we do and don’t know.” (C)</td>
<td></td>
</tr>
<tr>
<td>• “I believe a proof is meant to prove that a mathematical equation is valid…” (T)</td>
<td></td>
</tr>
<tr>
<td>• “To show that the things we’re working with are actually valid.” (T)</td>
<td></td>
</tr>
<tr>
<td>• “Purpose of proof is to show that something is true or false.” (T)</td>
<td></td>
</tr>
<tr>
<td>• “Conversely, proofs indicate when a statement is untrue, which in turn can help to define the boundaries and &quot;rules&quot; of mathematics.” (T)</td>
<td></td>
</tr>
</tbody>
</table>

• “Proofs also help investigate the nature of these statements and logical methods, whether true or false.” (T)
Appendix F (Continued)

(E)xplanation
- “To provide a clear and precise explanation of logical conclusions of patterns and relationships.” (T)
- “To explain why certain statements/operation are true.” (T)

(D)iscovery
- “We build the tools to discover and prove new things.” (T)
- “Proving helps us advance in mathematics, allowing for greater, and more accurate knowledge.” (T)
- “They also expand a knowledge from what the specific statements came from and that there might be some new statement evolved.” (C)
- “…to use known facts and discover new insight into mathematics.” (C)
- “To derive conclusions or further information from earlier information which has been assumed or already proven.” (C)
- “To further mathematical knowledge…” (C)

Intellectual

(C)hallenge
- “Also, proofs are fun…” (T)
- “They challenge intellectually.” (C)

(A)xiomatizing
- “Conversely, proofs indicate when a statement is untrue, which in turn can help to define the boundaries and "rules" of mathematics.” (T)
- “To incorporate a logical statement into the logical hierarchy that already exists.” (T)

(I)llustrating

(T)echniques
- “To learn how to prove things we already know…” (T)
- “Use the methods we know or we don't know yet to prove the mathematical concepts and other methods.” (C)
- “…some complex rules yield useful techniques…” (C)
- “…a mathematical proof serves to not only justify a statement, but also to explain the (or a possible) method by which the statement is justified.” (T)
- “To understand how to prove or disprove with empirical evidence certain formulas and mathematical theories.” (C)
- “To get a step by step logical conclusion to a problem.” (C)

(U)nderstanding
- “To understand the concepts behind the formulas and such that we often take for granted.” (C)
- “The purpose is to understand better how a formula or concept came from.” (C)
- “To make students understand deeper about the concepts they have learned before.” (C)
- “It is also to understand the basics of mathematics and to see the origin of where things come from.” (T)
- “I believe that proof is way of thinking to build up a understanding of math…” (T)
- “To further mathematical knowledge and understanding.” (C)
- “The purpose of proof is to truly understand the process, from start to finish, of solving problems.” (C)
Appendix F (Continued)

- "To understand how theorems are developed..." (C)

(T)hinking (S)kills
- "Also, to be forced to think more critically and use a part of our mind that isn't always required in math classes." (C)
- "To understand how to prove or disprove with empirical evidence certain formulas and mathematical theories." (C)
- "Also, to gain the ability to logically think about and solve a problem." (C)

(B)uilding Mathematics
- "To give logical steps or theorems to build more math on." (T)
- "To derive conclusions or further information from earlier information which has been assumed or already proven." (C)
- "To deduce complex rules from simple ones..." (C)
- "To prove there are connections among concepts." (C)
- "They also expand a knowledge from what the specific statements came from and that there might be some new statement evolved." (C)

(L)ogical Skills
- "Proofs also help investigate the nature of these statements and logical methods, whether true or false." (T)
- "Also, to gain the ability to logically think about and solve a problem." (C)
- "Also, to develop and expand logic skills." (C)
- "I believe that proof is way of thinking to build up a understanding of math, especially on logical aspect." (T)

Post-Assessment Codes

<table>
<thead>
<tr>
<th>Code Applied</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>(V)erification</td>
<td>&quot;To validate a mathematical expression.&quot; (T)</td>
</tr>
<tr>
<td></td>
<td>&quot;The purpose of proofs in mathematics is to... prove if something is true or false.&quot; (C)</td>
</tr>
<tr>
<td></td>
<td>&quot;To concretely say that a hypothesis is true or untrue based on facts and logic.&quot; (C)</td>
</tr>
<tr>
<td></td>
<td>&quot;A proof in mathematics is used to show the validity of a statement by use of theorems and other various laws.&quot; (T)</td>
</tr>
<tr>
<td>(T)</td>
<td>&quot;To verify conjectures we suspect to be true/false...&quot; (T)</td>
</tr>
<tr>
<td></td>
<td>&quot;To understand how or why some statement is true or false.&quot; (T)</td>
</tr>
<tr>
<td></td>
<td>&quot;For any statement or claim to [be] used in the future, it is necessary for it to be true generally for some conditions.&quot; (T)</td>
</tr>
<tr>
<td></td>
<td>&quot;To methodically confirm or deny proposed ideas that could have benefits in various fields.&quot; (T)</td>
</tr>
<tr>
<td></td>
<td>&quot;To show that without a reasonable doubt that we know the solution.&quot; (C)</td>
</tr>
<tr>
<td></td>
<td>&quot;The purpose of proofs in mathematics is to not only assess the truth value of a statement...&quot; (T)</td>
</tr>
</tbody>
</table>
Appendix F (Continued)

• “Proofs show that there are solutions to the problems.” (T)
• “To show the truth of mathematical statements...” (C)
• “...also to remove doubt of certain truths such as theorems.” (C)

(E)xplanation
• “The purpose of proof is to logically explain why and how concepts work in mathematics.” (T)
• “To understand how or why some statement is true or false.” (T)
• “To explain why a given statement is true or false.” (C)
• “To understand where concepts in mathematics are derived from and why they work the way that they do.” (T)
• “Explain reasoning and evidence why certain theorems/processes work.” (C)

(D)iscovery
• “...to try to expand on what we know.” (C)
• “Proofs are also important for finding new mathematical ideas.” (T)
• “Proofs can push math forward into new fields...” (C)
• “To build on past math [and] to make new math.” (T)

(C)ommunication
• “…which is readily interpreted by the intended audience.” (T)
• “A language difficult for both humans and computers to understand...” (C)

(A)xomatizing
• “I believe proof in mathematics establishes a (frequently long) connection between known or assumed knowledge in the form of axioms and knowledge found to be true dependent on these initial facts or foundations.” (T)
• “To create, using some consistent set of accepted axioms, an undeniable argument of some sort...” (T)

(I)llustrating
• “And, if you doubt, then proceed to disprove by the many tools that have been used to prove and disprove mathematical statements.” (C)
• “To help establish a logical step to show how someone could solve an unknown using what is known.” (C)

(T)echniques
• “Even if the concepts are initially "believable", the practice of proofs helps in understanding the harder concepts later on.” (T)
• “To verify conjectures we suspect to be true/false and to further understand the reasoning behind them.” (T)
• “The purpose of a proof is to go beyond the usual "it works because it just does" and actually understand why math works the way it does.” (T)
• “To [foster] understanding and break away from the notion that math is about numbers.” (C)
• “The purpose of proofs in mathematics is to understand better a concept taught...” (C)

(U)nderstanding
• “To let students understand the mathematical material better...” (C)
• “…and grasp the basic knowledge how the math works in this field.” (C)
• “To better understand what we believe...” (C)
• “To understand where concepts in mathematics are derived from...” (T)
Appendix F (Continued)

- “To understand the foundation of mathematics...” (C)
- “To understand the basic idea behind the theorem and other principles.” (T)

<table>
<thead>
<tr>
<th>(T)hinking (S)kills</th>
<th>“...explore methods of reasoning.” (C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(B)uilding Mathematics</td>
<td>“For this it is important that they be formally proved using results that are proved earlier.” (T)</td>
</tr>
<tr>
<td></td>
<td>“To establish [the] rules and show the logic that math is built off of.” (C)</td>
</tr>
<tr>
<td></td>
<td>“...show the validity of a statement by use of theorems and other various laws.” (T)</td>
</tr>
<tr>
<td></td>
<td>“To build on past math [and] to make new math.” (T)</td>
</tr>
</tbody>
</table>

| (L)ogical Skills                          | “The purpose of proof is to logically explain why and how concepts work in mathematics.” (T) |
|                                            | “To establish [the] rules and show the logic that math is built off of.” (C) |
|                                            | “The purpose of proofs in mathematics is to not only assess the truth value of a statement (theorem, conjecture, etc.), but also to provide a sound mathematical logic by which the statement is assessed.” (T) |
|                                            | “To concretely say that a hypothesis is true or untrue based on facts and logic.” (C) |
|                                            | “On the other hand, proof can help building logical construction in people’s mind.” (T) |
Appendix G

PROOFS GIVEN TO STUDENTS

Sample Proof Set 1

- If \(m\) and \(n\) are both odd, then \(m + n\) is odd.
- For integers \(a, b, c\), if \(a|b\), then \(a|bc\).
- If \(n\) is odd, then \(n^3\) is odd.
- If \(a, b\) are real numbers, then \(a^2 + b^2 \geq 2ab\).
- If \(a|b\) and \(k\) is any integer, then \(a|(b + ak)\).
- If \(a|b\) and \(a|c\) and \(x, y\) are any integers, then \(a|(bx + cy)\).
- If \(a|b\) and \(b|a\), then \(a = \pm b\).
- If \(a|(4n + 3)\) and \(a|(2n + 1)\) for some integer \(n\), then \(a = \pm 1\).
- Prove for all real numbers \(a\) and \(b\) with \(0 \leq a < b\), that \(a^2 < b^2\).

Sample Proof Set 2

- Prove that for every odd integer \(n\), \(n^2 - 1\) is divisible by 8.
- Prove that for any two real numbers \(a\) and \(b\), \(||a| - |b|| \leq |a - b|\). Hint: How many cases are there to consider? Can you narrow down the number of cases by using the triangle inequality?
- Prove that for any two real numbers \(a\) and \(b\), \(\max(a, b) = \frac{1}{2}(a + b + |a - b|)\). Find a similar formula for \(\min(a, b)\).
- Prove by contradiction that if \(1 \leq x\), then \(\sqrt{x} \leq x\).
- Prove that for every positive integer \(n\), \(n^3 + 5n\) is divisible by 3.
- Prove that for every positive integer \(n\), \(n^3 + (n + 1)^3 + (n + 2)^3\) is divisible by 9.
Appendix G (Continued)

• Prove that if \(m\) and \(n\) are odd integers, then \(n^2 - m^2\) is divisible by 8.
• Assume \(a\) and \(b\) are real numbers. Prove or disprove the following statement: If \(4ab < (a + b)^2\), then \(a < b\).

Sample Proof Set 3

• Prove that for every positive integer \(n\), \(n^3 + 5n\) is divisible by 3.
• Use the fact that \(\frac{d}{dx}x = 1\) and the product rule to prove by induction the power rule, \(\frac{d}{dx}x^n = nx^{n-1}\) for all \(n \geq 1\).
• Prove that if \(m\) and \(n\) are odd integers, then \(n^2 - m^2\) is divisible by 8.
• Prove that \(\sqrt{3} - \sqrt{2}\) is irrational.
• Prove that there does not exist a smallest positive number.
• Prove that for all integers \(n \geq 2\), we have \(n^3 + 1 > n^2 + n\).
• Prove that for all \(x > 0\) and all nonnegative integers \(n\), \((1 + x)^n \geq 1 + nx\).
• Prove that every integer \(n \geq 12\) can be written in the form \(n = 3m + 7l\), where \(m\) and \(l\) are nonnegative integers.

Sample Proof Set 4

• Determine if the statement is true. If the double implication fails, determine whether one or the other implications holds. If an equality fails, determine whether the statement becomes true if the equals is replaced by \(\subset\) or \(\supset\).
  \(A \subset B\) and \(A \subset C \iff A \subset (B \cup C)\)
• Determine if the statement is true. If the double implication fails, determine whether one or the other implications holds. If an equality fails, determine whether the statement becomes true if the equals is replaced by \(\subset\) or \(\supset\).
  \(A \subset B\) or \(A \subset C \iff A \subset (B \cup C)\)
• Determine if the statement is true. If the double implication fails, determine whether one or the other implications holds. If an equality fails, determine whether the statement becomes true if the equals is replaced by \(\subset\) or \(\supset\).
  \(A \subset B\) and \(A \subset C \iff A \subset (B \cap C)\)
Appendix G (Continued)

- Determine if the statement is true. If the double implication fails, determine whether one or the other implications holds. If an equality fails, determine whether the statement becomes true if the equals is replaced by $\subseteq$ or $\supseteq$.
  
  $A \subseteq B$ or $A \subseteq C \iff A \subseteq (B \cap C)$

- Determine if the statement is true. If the double implication fails, determine whether one or the other implications holds. If an equality fails, determine whether the statement becomes true if the equals is replaced by $\subseteq$ or $\supseteq$.
  
  $A - (A - B) = B$

- Determine if the statement is true. If the double implication fails, determine whether one or the other implications holds. If an equality fails, determine whether the statement becomes true if the equals is replaced by $\subseteq$ or $\supseteq$.
  
  $A \cap (B - C) = (A \cap B) - (A \cap C)$

- Determine if the statement is true. If the double implication fails, determine whether one or the other implications holds. If an equality fails, determine whether the statement becomes true if the equals is replaced by $\subseteq$ or $\supseteq$.
  
  $(A \cap B) \cup (A - B) = A$

- Determine if the statement is true. If the double implication fails, determine whether one or the other implications holds. If an equality fails, determine whether the statement becomes true if the equals is replaced by $\subseteq$ or $\supseteq$.
  
  $A - (B \cup C) = (A - B) \cap (A - C)$

- Determine if the statement is true. If the double implication fails, determine whether one or the other implications holds. If an equality fails, determine whether the statement becomes true if the equals is replaced by $\subseteq$ or $\supseteq$.
  
  $A - (B \cap C) = (A - B) \cup (A - C)$

### Sample Proof Set 5

- Prove that $A \times B = B \times A$ if and only if $A = B$.
- Let $A$ have $n$ elements and $B$ have $m$ elements. Show that $A \times B$ has $mn$ elements.
- Show $A \times (B \cap C) = (A \cap C) \times (B \cap C)$.
- Show $A \times (B \cup C) = (A \times B) \cup (A \times C)$.
- Show $A \times (B - C) = (A \times B) - (A \times C)$.
- Show $(A - B) \times (C - D) = (A \times C - B \times C) - (A \times D)$.
- Show if $A \subseteq B$, then $A \times C \subseteq B \times C$. 

Appendix G (Continued)

- If $A, B \neq \emptyset$, then $A \times B \subseteq C \times D$ if and only if $A \subseteq C$ and $B \subseteq D$.
- Show $(A \times B)^c = (A^c \times B^c) \cup (A^c \times B) \cup (A \times B^c)$.

Sample Proof Set 6

- Suppose $f : X \to Y$ and let $A \subseteq X$ and $B \subseteq X$. Prove:
  (a) $f(A \cap B) \subseteq f(A) \cap f(B)$. Give an example where equality fails.
  (b) $f(A - B) \supseteq f(A) - f(B)$. Give an example where equality fails.
- Suppose $f : X \to Y$ and let $C \subseteq Y$ and $D \subseteq Y$. Prove:
  (a) If $C \subseteq D$, then $f^{-1}(C) \subseteq f^{-1}(D)$.
  (b) $f^{-1}(C \cup D) = f^{-1}(C) \cup f^{-1}(D)$.
- Suppose $f : X \to Y$ and let $C \subseteq Y$ and $D \subseteq Y$. Prove:
  (a) $f^{-1}(C \cap D) = f^{-1}(C) \cap f^{-1}(D)$.
  (b) $f^{-1}(C - D) = f^{-1}(C) - f^{-1}(D)$.
- Suppose $f : X \to Y$ and let $A \subseteq X$ and $B \subseteq Y$. Then:
  (a) If $f$ is one-to-one, then $f^{-1}[f(A)] = A$.
  (b) If $f$ is onto, then $f[f^{-1}(B)] = B$.
- Suppose $f : X \to Y$ and $g : Y \to Z$ are functions.
  (a) If $g \circ f$ is onto, prove that $g$ must be onto.
  (b) Find an example where $g \circ f$ is onto, but $f$ is not onto.
- Suppose $f : X \to Y$ and $g : Y \to Z$ are functions. If $f$ and $g$ are one-to-one, then so is $g \circ f$.
- Suppose $f : X \to Y$ and $g : Y \to Z$ are functions. If $f$ and $g$ are onto, then so is $g \circ f$.

Sample Proof Set 7
Appendix G (Continued)

1. A function \( f \) is a convex function if and only if for all \( x, y \in C \) and every real number \( t \) with \( 0 \leq t \leq 1 \), it follows that \( f(tx + (1 - t)y) \leq tf(x) + (1 - t)f(y) \). Prove that if \( f, g \) are convex functions, then \( f + g \) is a convex function.

2. Prove that if \( m, n \) are odd integers, then the equation \( x^2 + 2mx + 2n = 0 \) has no rational root.

3. Prove that at a party of \( n \geq 2 \) people, there at least two people who have the same number of friends at the party.

4. Prove that if \( c \) is an odd integer, then the equation \( x^2 + x - c = 0 \) has no integer solutions. [Hint: Consider the contrapositive.]

5. Prove that if \( x > 2 \) is a real number, then there is a unique real number \( y > 0 \) such that \( x = \frac{2y}{1 + y} \).

6. For all integers \( n \geq 2 \), show \( \prod_{k=2}^{n} \left( 1 - \frac{1}{k^2} \right) = \frac{n+1}{2n} \).

7. Suppose that \( a, b \) are real numbers and \( f : \mathbb{R} \to \mathbb{R} \) is defined by \( f(x) = ax + b \).
   a) Find a condition on \( a \) so that \( f \) is injective.
   b) Find a condition on \( a \) so that \( f \) is surjective.
   c) State and prove propositions for your claims in (a) and (b).

8. Let \( S \subseteq \{1, 2, \ldots, 2n\} \) such that \( |S| = n + 1 \). Show that \( S \) contains a pair of consecutive integers.

9. Prove that the product of any \( n \) consecutive positive integers is divisible by \( n! \).

Sample Proof Set 8

- Let \( x, y \in \mathbb{Z} \). Prove that if \( 3 \mid x \) and \( 3 \mid y \), then \( 3 \mid (x^2 - y^2) \).
- Let \( a, b \in \mathbb{Z} \). Show that if \( a \equiv 5 \pmod{6} \) and \( b \equiv 3 \pmod{4} \), then \( (4a + 6b) \equiv 6 \pmod{8} \).
- Prove that \( 2 \mid (n^4 - 3) \) iff \( 4 \mid (n^2 + 3) \).
- Let \( n, m \in \mathbb{Z} \). Prove that if \( n \equiv 1 \pmod{2} \) and \( m \equiv 3 \pmod{4} \), then \( (n^2 + m) \equiv 0 \pmod{4} \).
- If \( m \in \mathbb{Z}^+ \) and if \( a \equiv b \pmod{n} \), then \( a^m \equiv b^m \pmod{n} \).
- If \( a \equiv b \pmod{n} \) and \( c \equiv d \pmod{n} \), then \( ac \equiv bd \pmod{n} \).
- If \( a \equiv b \pmod{n} \), then \( a \equiv (b + kn) \pmod{n} \).
- If \( a \equiv b \pmod{n} \) and \( c \equiv d \pmod{n} \), then \( a + c \equiv b + d \pmod{n} \).
VITA
Education

July 2013  Doctor of Arts, Mathematics, University of Illinois at Chicago (UIC), Chicago, IL.
This degree is designed to train educators and researchers for undergraduate instruction and
research at the post-secondary level. Research is conducted in methodologies and
techniques that lead to successful teaching of undergraduate mathematics.

May 2006  M.A., Mathematics, University of Colorado, Boulder, CO.
Master's Thesis Talk: Ramsey Theory
Advisor: Dr. J. Donald Monk

May 2004  B.S., Mathematics, Colorado State University, Fort Collins, CO.
Undergraduate Thesis: An Introduction to Electrical Impedance Tomography and Usage
of the Bessel Function
Advisor: Dr. Jennifer Mueller

Doctoral Thesis

Title  Proof Construction and Collaborative Revision in Undergraduate Mathematics
Advisor  Dr. Mara Martinez
Description  The purpose of this study is to examine the effects of collaborative revision on undergraduate students’ proof writing and proof validation skills. Additionally, the study will collect data on the beliefs held about proof by students in a transition to proof course. Implications from this study can inform teaching practices by identifying how students think about proof in mathematics and by the use of a new teaching approach to help students learn proof construction.

Teaching Experience

Pre-Service Teachers

Fall 2012–Present  Teaching Assistant, Department of Mathematics, Statistics, and Computer Science (MSCS), UIC, Chicago, IL.
Assisting in a class of pre-service teachers with classroom activities and group work.
• Arithmetic and Algebraic Structures

Undergraduate Mathematics

Fall 2010–Present  Teaching Assistant, Emerging Scholars Program, MSCS, UIC, Chicago, IL.
This program offers an opportunity to work on challenging problems with classmates through innovative techniques of cooperative learning. Created and facilitated small group work with students on problems specifically chosen to challenge students.
• Introduction to Advanced Mathematics (Fall 2012), Calculus III (Spring 2012, Fall 2011), Calculus I (Spring 2010), Precalculus (Fall 2010)

2011–2012  Adjunct Mathematics Instructor, Tribeca Flashpoint Media Arts Academy, Chicago, IL.
Taught college mathematics class covering linear programming, geometry, statistics, and exponential/logarithmic functions, assigned and graded homework, created and graded quizzes and projects, responsible for all lesson plans and class activities.
• General Math (Spring 2012, Fall 2011)
Summer 2011, 2012 **Lecturer, Summer Enrichment Workshop, MSCS, UIC, Chicago, IL.**
Designed, organized, and taught all lessons, created and graded all mid-term exams, co-created final exam, held office hours, about 15 students enrolled.
- Beginning Algebra (Summer 2012), Intermediate Algebra (Summer 2011)

2009–2010 **Teaching Assistant, MSCS, UIC, Chicago, IL.**
Planned, organized, and led all discussion sessions, created and graded quizzes, graded exams, held tutoring hours for discussion.
- Calculus I (Spring 2010), Intermediate Algebra (Fall 2009)

2006–2009 **Full-Time Tenured Professor, Olympic College, Bremerton, WA.**
Taught 3-4 courses per quarter, created syllabi and course materials, served on college and course committees, faculty advisor and founder of the Mathematics Club, developed learning assessments, chose textbooks and course material, served on hiring committees and advised students.

2004–2006 **Teaching Assistant, Department of Mathematics, University of Colorado, Boulder, CO.**
Instructor of record for one course per semester, responsible for developing all course materials and syllabus, grading, tutoring, proctoring exams.
- Quantitative Reasoning & Mathematical Skills, Calculus I, Precalculus

**Publications & Reports**


*Cilli-Turner, E.* (2013). Effects of collaborative revision on beliefs about proof function and validation skills. *Proceedings of the 16th annual Conference on Research in Undergraduate Mathematics Education (RUME), Denver, CO.*


**Presentations**

Nov 2013 **Presenter**, “Effects of Collaborative Revision on Undergraduate Students’ Proof Validation Skills”, 34th Annual Meeting of PME-NA, Chicago, IL.


Feb 2013 **Presenter**, “Effects of Collaborative Revision on Beliefs About Proof Function and Validation Skills”, 16th Annual Conference on RUME, Denver, CO.

Jan 2013 **Presenter**, “Proof Construction and Collaborative Revision in Undergraduate Mathematics”, Joint Mathematics Meetings, San Diego, CA.


May 2012 **Presenter**, “Student Thinking about Proof Construction and Validation”, Washington Mathematical Association of Two-Year Colleges (WAMATYC) Conference, Wenatchee, WA.
Mar 2012  **Presenter**, "Cooperative Learning Workshops: Discussion on Organization and Methods", Chicago Symposium Series, Northwestern University, Chicago, IL.

### Conferences & Workshops Attended

**Mar 2013**  **Chicago Symposium Series**, *Excellence in Teaching Mathematics and Science: Research and Practice*, Loyola University, Chicago, IL.

**Feb 2013**  **Chicago Symposium Series**, *Excellence in Teaching Mathematics and Science: Research and Practice*, University of Illinois at Chicago, Chicago, IL.

**Oct 2012**  **AMS Sectional Meeting**, *University of Akron*, Akron, OH.

**June 2012**  2nd *Transforming Research in Undergraduate STEM Education (TRUSE) Conference*, *University of St. Thomas*, St. Paul, MN.

**May 2012**  **Chicago Symposium Series**, *Excellence in Teaching Mathematics and Science: Research and Practice*, Northeastern Illinois University, Chicago, IL.

**Feb 2012**  15th *Annual Conference on RUME*, Portland, OR.

### Fellowships & Grants

**June 2013**  **Project NExT Fellow.** A professional development program of the Mathematical Association of America for recent graduates starting a university teaching position.

**Sept 2012**  **AMS Travel Grant.** For travel to the AMS sectional meeting at University of Akron.

**Summer 2006**  **National Science Foundation (NSF) Fellowship.** Grant provided through East Asia and Pacific Summer Institute at the NSF, whose goals are to help students initiate scientific relationships that will better enable future collaboration with foreign counterparts. Funding was given to perform mathematical research in China on the topic of game theory.  
  *Advisor: Dr. Yan Guiyin, Academy of Mathematics and System Sciences, Beijing, China*

### Service

**2013**  **Reviewer.**


**2012-2013**  **Teaching Assistant Coordinator**, MSCS, UIC, Chicago, IL.

**2011-Present**  **Graduate Employees Organization**, Secretary (2011–12) & Treasurer (2012–13), UIC, Chicago, IL.

**2010-2011**  **Mathematics Graduate Student Association President**, MSCS Department, UIC, Chicago, IL.

**May 2008**  **Hiring Committee Member**, Olympic College, Bremerton, WA.

**2007-2009**  **Mathematics Club Faculty Advisor**, Olympic College, Bremerton, WA.

### Memberships

**2009-Pres**  **Association for Women in Mathematics.**

**2010-Pres**  **American Mathematical Society.**

### Computer Skills

Mathematica, Matlab, \LaTeX, Beamer, SPSS