

STUDENTS' LOGICAL REASONING AND MATHEMATICAL PROVING OF
IMPLICATIONS

By

KoSze Lee

A DISSERTATION

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

DOCTOR OF PHILOSOPHY

Educational Psychology and Educational Technology

2011

ABSTRACT

STUDENTS' LOGICAL REASONING AND MATHEMATICAL PROVING OF IMPLICATIONS

By

KoSze Lee

Students' difficulties in reasoning with logical implication and mathematical proving have been documented widely (Healy & Hoyles, 2000; Knuth, Choppin, & Bieda, 2009). Review of the educational and cognitive science studies of students' reasoning with logical implications and mathematical proving have revealed that their lack of cognizance of counterexamples might be a crucial factor. This study examined the role of logic training and counterexample in enhancing students' logical reasoning and various aspects of mathematical proving, namely, Proof Construction, Proof Validation and Knowledge of Proof Method. In particular, the study hypothesized that logic training emphasizing counterexamples was better able to improve students' reasoning of logical implications as well as mathematical proving, in comparison to the other two approaches emphasizing rule violations and truth tables.

Using a pretest-intervention-posttest experimental design (3 conditions by 2 test trials), students' written and interview data ($N = 60$) were collected from three Singapore school sites, each over a four-day contact period (including the pretest and posttest administration days). Experimental results showed that logic training emphasizing counterexamples was significantly more effective in improving students' logical reasoning of implication than the other two approaches ($p = .0007$, large effect size). However, logic training was only similarly effective or ineffective in improving some aspects of students' mathematical proving across conditions.

Interview findings from 12 selected students' works on a new proving task conjectured that students improved their use of deductive inferences in all aspects of mathematical proving

after logic training. Moreover, their successes in constructing mathematical proofs were also subjected to two conjectured factors, students' interpretation of implication and mathematical knowledge. These findings suggested the importance of logic training and counterexamples in mathematics education and pointed to further inquiry about the role of students' interpretation of implications and mathematical knowledge in mathematical reasoning and proving.

ACKNOWLEDGEMENTS

From the bottom of my heart, I want to thank the follow people. For without them, completion of this dissertation and the doctoral program is not possible.

I owe my accomplishments to my wife, PohChing Chia, and my two wonderful kids, Johan Lee and Joanne Lee. PohChing is a constant source of love and support to me. She has never doubted my abilities and what I can accomplish. Johan and Joanne have inspired me with their joy and contentment in life. They are the wind beneath my wings.

I am greatly indebted to my advisor and mentor, Dr John (Jack) P. Smith III, for his insights and wisdom in guiding my scholarly journey. In many ways, he is a good model of a university faculty for me. Lunch conversations with him were always enriching and enjoyable. My special thanks also go to committee members, Dr Kelly Mix, Dr Sharon Senk and Dr Raven McCrory. They were committed and supportive in my dissertation, and provided valuable inputs.

I also acknowledge the teachers and students who participated in this study. Their willingness to commit their time and efforts to the training sessions made this study possible.

I also want to thank Leo Chang, Lorraine Males, Aaron Mosier, Sasha Wang and Jungeun Park for their timely suggestions and support for my dissertation preparation. I want to express my gratitude to the colleagues of the Mathematics Learning Research Group (MLRG), uniquely here in Michigan State University.

My heartfelt gratitude to Dan Ouyang and Sumei Wei, and John and Bonnie Bankson, for their generous hospitality and great friendship support while I was back to complete the study.

Most importantly, glory be to God, who has blessed me and my family with His loving kindness during these years in U.S.

CONTENTS

LIST OF TABLES	ix
LIST OF FIGURES	xi
LIST OF EXCERPTS	xiii
CHAPTER 1 INTRODUCTION	1
Background	3
<i>Mathematical Implications as Generalized Conditional</i>	4
<i>Students' Difficulty with Logical Implication – a crucial barrier</i>	6
<i>Effects of Instructions on Students' Reasoning of Logical Implications</i>	10
<i>Effects of Logic Training on Students' Mathematical Proving</i>	12
<i>Students' Difficulty with Mathematical Proving</i>	13
<i>Aim of this study: In search for an effective logic training</i>	15
CHAPTER 2 THEORETICAL FRAMEWORK	17
Mathematical Implication, its Logical Variants and Counterexamples	17
<i>Converse, Contrapositive and Negation of Implication</i>	18
Conceptualization of Mathematical Proving Ability	19
Generalized Conditional and Mathematical Proof and Proving	20
<i>Definition of Mathematical Proofs</i>	21
<i>Proof Construction</i>	22
<i>Proof Validation</i>	22
<i>Knowledge of Proof Methods</i>	23
Leveraging the use of Counterexamples in Reasoning and Proving of Mathematical Implications	24
<i>Constrained Example Generation</i>	25
<i>Enhancing CEG through task formulations</i>	26
The Research Questions	27
CHAPTER 3 METHOD	30
Subjects	30
Design	31
Materials	34
<i>Pre-test and Post-test instruments</i>	34
<i>Training materials</i>	38
<i>Training Materials used in the Control condition</i>	39
<i>Training Materials used in the PO condition</i>	41
<i>Training Materials used in the W condition</i>	44
<i>Students' reflection of learning</i>	47
<i>Materials for Post Study Interview</i>	47
Procedure	48

Data Scoring and Coding.....	50
<i>Scoring of selection task items.....</i>	<i>51</i>
<i>Coding schemes for Students' responses to proof items.....</i>	<i>51</i>
<i>Coding for Deductive-proof Construction.....</i>	<i>52</i>
<i>Coding Proof-by-counterexamples Construction for mathematically false implications.....</i>	<i>54</i>
<i>Coding scheme for Invalidation of empirical proof.....</i>	<i>56</i>
<i>Coding scheme for considering the logical equivalence between implication and converse.....</i>	<i>57</i>
<i>Coding scheme for considering the logical equivalence between implication and contrapositive.....</i>	<i>58</i>
<i>Coding scheme for the validation of Proof-by-contradiction.....</i>	<i>59</i>
<i>Coding of Post Study Interview Data.....</i>	<i>61</i>
Analysis.....	62
<i>Preliminary Analyses.....</i>	<i>62</i>
<i>Overview of the analyses plan</i>	<i>66</i>
CHAPTER 4 RESULTS	70
Effect of training on Students' logical reasoning	71
Effects of training on Students' Proof Construction.....	74
<i>Deductive-proof Construction</i>	<i>75</i>
<i>Deductive-proof Construction of Item related to Elementary Number Theory</i>	<i>76</i>
<i>Deductive-proof Construction of Item related to Quadratics.....</i>	<i>78</i>
<i>Students' Indirect approach in Deductive-proof Construction</i>	<i>79</i>
<i>Proof-by-counterexample Construction.....</i>	<i>81</i>
<i>Proof-by-counterexample Construction of Item related to Elementary Number Theory</i>	<i>81</i>
<i>Proof-by-counterexample Construction of Item related to Quadratics.....</i>	<i>82</i>
<i>Students' False Deductive Proofs with Inadequate Mathematical Considerations</i>	<i>84</i>
Effects of training on Students' Proof Validation.....	85
<i>Invalidation of Empirical Proof.....</i>	<i>87</i>
<i>Validation of Proof-by-contradiction</i>	<i>89</i>
<i>Students' Understanding of Proof-by-Contradiction by Counterexample Elimination</i>	<i>91</i>
Effects of training on Students' Proof Knowledge.....	93
<i>Logical non-equivalence between an implication and its converse.....</i>	<i>94</i>
<i>Logical equivalence between implication and its contrapositive</i>	<i>96</i>
Correlation between Logical Reasoning and Various aspects of Mathematical Proving.....	98
Summary of the Experimental Results	100
Preliminary Discussion	102
<i>Students' Improved Aspects of Mathematical Proving: Practice Effect or Training Effect.....</i>	<i>103</i>
CHAPTER 5 POST-STUDY INTERVIEW RESULTS.....	105
Controlled and uncontrolled factors of the experimental design.....	106
Theoretical constructs: Logical reasoning and Mathematical Proving.....	107
Analysis of Interview Data	109
<i>Background of the interviewees.....</i>	<i>109</i>
The Interview Task and Data Coding.....	111
<i>Coding of Students' Proving Attempts.....</i>	<i>112</i>
<i>Coding of Students' Modification of Implication.....</i>	<i>114</i>

Overview of students' proving behaviors	115
Students' Considerations of Mathematical Objects	117
Students' Deductive-proof and Proof-by-counterexample Constructions	120
<i>Use of Numerical Representations in Deductive Proofs or Proof-by-counterexample</i>	121
<i>Use of Algebraic Representations in Deductive Proofs or Proof-by-counterexample</i>	122
<i>Alternative use of representations in Deductive Proofs</i>	123
Students' Deductive-proof Construction for modified implications	124
<i>Students' representations and their uses in Proof Construction</i>	128
Students' modifications of the implication	129
<i>Students' consideration of objects and modification of implications</i>	132
Summary of findings from the interview	135
<i>Account of possible Contextual factors affecting Students' Performance</i>	137
 CHAPTER 6 DISCUSSION AND CONCLUSION	 139
Summary of the Study and its Findings	139
<i>Experimental Findings from the Pretest and Posttest</i>	142
<i>Findings from the Post-study interview</i>	143
Discussion	146
<i>Effects of Logic Training on Various Aspects of Mathematical Proving</i>	146
<i>Wason's Task as a Logical Reasoning Indicator</i>	156
<i>Logic Training Emphasizing Counterexamples for Students' logical reasoning</i>	158
Implications for Education and Research	159
<i>The Role of Logic Training in Teaching and Learning of Mathematics</i>	160
<i>The Role of Counterexamples in Mathematical Reasoning and Proving</i>	161
<i>Setting Realistic Expectation of Logic Training</i>	162
<i>Rethinking about Instructions of Logical Reasoning for Mathematics Classrooms</i>	163
<i>Role of Mathematical Knowledge in Mathematical Reasoning and Proving</i>	165
<i>Research implications</i>	165
Limitations and Future Studies	166
<i>Specificity of Singapore Students</i>	167
<i>Sample size and Significance of findings</i>	167
<i>More tasks needed for reliability</i>	168
<i>Conviction and Validation: Personal vs. Social</i>	168
<i>Coding scheme for Proof Constructions</i>	169
<i>Training tasks</i>	170
<i>Laboratory-based Instructions and Classroom Instructions</i>	170
<i>Maintenance of Training Effects and Latent Effects</i>	171
<i>Future studies</i>	172
Conclusion	172
 APPENDICES	 173
Appendix A: Test Set 1	174
Appendix B: Summary of Isomorphic Items in Test Sets 1 and 2	184
Appendix C: Implications used in the Training Materials of all Conditions	186
Appendix D: Training Materials for Conventional Approach (Condition C)	187
Appendix E: Training Materials emphasizing Counterexample (Condition W)	194

Appendix F: Training Materials emphasizing Rule Violation (Condition PO).....	215
Appendix G: Interview Task and Sample of Questions.....	223
Appendix H: Student’ Proving Attempts for the Given Implication	224
Appendix I: Student’ Modification of Implications and Proving Attempts	229
REFERENCES	236

LIST OF TABLES

Table 1: Overview of the design	31
Table 2: Composition of Test instruments.....	34
Table 3: Design of training materials.....	38
Table 4: Descriptors for coding students' Deductive-proof Constructions.....	52
Table 5: Descriptors for coding students' Proof-by-counterexample Constructions.....	55
Table 6: Descriptors for coding students' Invalidation of Empirical Proof.....	57
Table 7: Coding scheme for the implication and its converse	58
Table 8: Coding scheme for the implication and its contrapositive	59
Table 9: Coding scheme for validating the Proof-by-contradiction item	60
Table 10: Inter-rater agreement of coding and Reliability coefficients.....	63
Table 11: Principal component analysis	65
Table 12: Overview of the analyses of students' test responses.....	67
Table 13: Students' performance in logical reasoning by condition.....	72
Table 14: Students' total raw scores in Proof Construction by condition.....	74
Table 15: Students' principal component scores in Proof Construction by condition.....	75
Table 16: Students' performance in Proof Validation by condition	86
Table 17: Students' Consideration of Logical Equivalence	93
Table 18: Spearman correlational matrix of students' gain scores.....	99
Table 19: Summary of ANOVA outcomes and Conclusions	100
Table 20: Post test performance of Interviewees	110
Table 21: Coding scheme for students proving attempts	113

Table 22: Sample of coded transcript	113
Table 23: Sample of Modified Implications	115
Table 24: Overview of Students' proving of the impromptu task.....	116
Table 25: Students' considerations of mathematical objects for proving implication	117
Table 26: Students' reasons for choosing a set of numbers	119
Table 27: Types of representations, Conclusions and Level of Proof	120
Table 28: Students use of Mathematical Representations by groups	125
Table 29: Modifications of Implication by Students.....	129
Table 30: Isomorphic implications in Test Set 1 and 2.....	184
Table 31: Implications used across all Conditions.....	186
Table 32: Attempts made by Students to prove or falsify the Implication.....	224
Table 33: Modifications of Implications made by Students.....	229

LIST OF FIGURES

Figure 1: A Proof Construction item in the test set	36
Figure 2: A Proof Construction practice problem and its solution (Control Condition).....	40
Figure 3: Practice problem involving obligation situation.....	42
Figure 4: Solution for Why Violations were impossible.....	43
Figure 5: Practice Problem in Wason Condition	45
Figure 6: Solution for Why Counterexamples were impossible	46
Figure 7: Pretest and Posttest mean scores by conditions.....	73
Figure 8: Students' Deductive-proof Construction (Elementary number theory)	77
Figure 9: Students' Deductive-proof Construction (Quadratics).....	78
Figure 10: Sample of Indirect Proof approach in Deductive-proof Construction (This figure was hand drawn by: Alex).....	80
Figure 11: Students' Proof-by-counterexample Construction (Elementary Number Theory)	82
Figure 12: Students' Proof-by-counterexample Construction (Quadratics).....	83
Figure 13: Students' Proof-by-counterexample Construction with inadequate consideration (This figure was hand drawn by: Brady).....	85
Figure 14: Students' Invalidation of Empirical Proof.....	88
Figure 15: Students' Validation of Proof-by-contradiction	90
Figure 16: Student's justification for the Validity of Proof-by-contradiction (This figure was hand drawn by: Carl).....	92
Figure 17: Students' Consideration of Logical non-equivalence of Converse	95

Figure 18: Students' Consideration of Logical equivalence of Contrapositive.....	97
Figure 19: Conjecture Factors affecting Students' Mathematical Proving performance .	147
Figure 20: Graph of Quadratic Equation.....	178
Figure 21: Truth table of "<i>If P then Q</i>"	187

LIST OF EXCERPTS

Excerpt 1: Average is the ‘Center’ number of a ‘balanced’ set	124
Excerpt 2: S6’s Proof Constructions prior to modification	132
Excerpt 3: S5’s first modification of the implication.....	134

CHAPTER 1 INTRODUCTION

The role of students' logical reasoning in learning mathematical proving has regained educational attention recently (Durand-Guerrier, 2003; Epp, 2003; Inglis & Simpson, 2004; Selden & Selden, 2003). Educational studies of the effects of training logical reasoning on students' abilities in mathematical proving did not produce strong desirable benefits (Deer, 1969; Durand-Guerrier, 2003; Epp, 2003; Mueller, 1975; J. L. Platt, 1967). Observed improvements in proving are limited to students with better mathematics ability. Cognitive studies of the effects training in logic instructions had on improving students' logical reasoning also often revealed little to moderate benefits (Cheng, Holyoak, Nisbett, & Oliver, 1986; Leighton, 2006). On the whole, logic instructions and students' logical reasoning seems unproductive for their ability in mathematical proving.

However, a closer analysis of the difficulties faced by students in both logical reasoning and mathematical proving suggests that a common source of difficulty lies in the students' inclination for empirical verifications and their lack of cognizance of the possible counterexamples. To date, discussion in educational literature concerning the use of counterexamples to help students overcome this inclination is still in its exploratory stage and limited to instructional studies focusing on mathematical proof and proving (A. Stylianides & Stylianides, 2009a; Zazkis & Chernoff, 2008). Recognizing that students' use of logical implications is foundational to their deductive proving (Harel & Sowder, 1998; Healy & Hoyles, 2000; Hoyles & Küchemann, 2003; G. Stylianides & Stylianides, 2008), this study intends to examine how logic instructions with an explicit emphasis on counterexamples might help students improve their reasoning of logical implications and ability in proving.

Progress in educational and cognitive science studies of students' use of logical implications warrants this study's approach based on counterexamples. Recent cognitive science studies have devised ways of improving students' logical reasoning through eliciting students' cognizance of possible counterexamples (Cheng, et al., 1986; Griggs & Cox, 1982; R. Platt & Griggs, 1993; Stenning & Lambalgen, 2004). Educational studies of students' mathematical proving had also developed finer conceptual frameworks of mathematical proving, beyond proof writing ability in geometry, for better distinction of the different aspects of students' ability in mathematical proving. These additional aspects included students' evaluation of mathematical proofs and logical understanding of different proof approaches (Alcock & Weber, 2005; Antonini & Mariotti, 2008; Balacheff, 1988; Harel & Sowder, 1998; Selden & Selden, 2003; A. Stylianides & Stylianides, 2009a; Weber, 2001). The aforementioned studies reporting limited effects of logic training, however, did not explore the use of counterexamples as an alternative of logic training nor the effects of such training on students' validation of mathematical proofs and logical knowledge of different proof approaches (Deer, 1969; Durand-Guerrier, 2003; Epp, 2003; Mueller, 1975; J. L. Platt, 1967). Yet, advocacy of the benefits of logic training and which better training approach to adopt in mathematics classrooms had been made frequently based on anecdotal evidences and theoretical speculations (Epp, 1994, 2003; Selden & Selden, 2003; G. Stylianides & Stylianides, 2008). This study thus addresses this research gap by inquiring whether better alternatives in training students' logical reasoning and students' mathematical proving exist, and how these logic training approaches impact other aspects of students' mathematical proving, in addition to students' proof productions. In particular, this study aims to find out, through a pre-post intervention design, whether students' construction of possible

counterexamples can have beneficial effects on their reasoning of mathematical implications and mathematical proving.

In sum, the purpose of this study is thus important in a few ways. First, with the research advancements made in the area of mathematical reasoning and proving, aspects of students' performance in proof and proving are no longer limited to proof productions as studied in the past (Alcock & Weber, 2005; Antonini & Mariotti, 2008; Deer, 1969; Epp, 2003; Mueller, 1975; Selden & Selden, 2003). While logic training has been increasingly advocated as central to these additional aspects (Epp, 2003; Selden & Selden, 2003), an empirical inquiry of the role of logic training is in need. Without a clear understanding of the extent of logic training with regards to these additional aspects, the instructional theory and goals of developing students' logical reasoning and mathematical proving in classrooms remained as individual teachers' pedagogical beliefs. Second, the role of counterexamples had only gained emerging research attention for the purpose of mathematics learning and still at the stage of theory-building via case study methods (A. Stylianides & Stylianides, 2009b; Zazkis & Chernoff, 2008). This study clarified further the role of counterexamples in developing students' mathematical reasoning and proving in classrooms through experimental methods. Third, logic training had typically begun with truth tables followed by proof practices. This study aimed to explore the feasibility of other training approaches that used counterexamples, as implied by documented empirical studies (Cheng, et al., 1986; G. Stylianides & Stylianides, 2008).

Background

Before the reviewing the background of this study, I will first have to present an overview of the notions of logical implications and proving with the disciplines of logic and

mathematics, akin to a crash course on logic of implications, to inform the readers the nature of the mathematical reasoning and proving tasks situated within the scope of the study's problem. Next, I will provide a literature review of the findings of students' reasoning with logical implications from cognitive science studies and mathematical educational studies. I will also review findings about students' difficulty in mathematical proving and the effects of logic training on students' reasoning and proving, before I state and justify the problem pursued by this study.

Mathematical Implications as Generalized Conditional

Mathematical implication is often expressed in the form of a conditional statement of the sentence form "If [statement P] then [statement Q]", which relates the antecedent (statement P) and the consequent (statement Q). The antecedent P and consequent Q are mathematical propositions concerning mathematical concepts and properties.

Various notions of logical implications have been proposed to define the truth values of implications as 'True' or 'False' (Quine, 1950). Of particular interest to this study is the notion of *generalized conditional*, which postulates a logical implication as a conditional relationship between sets of mathematical objects satisfying the antecedent and the consequent (Tarski, 1956). The implication is considered logically and mathematically true when no mathematical instance that satisfies the antecedent P but not the consequent Q can be found. The implication is falsified when its complementary statement "a mathematical instance satisfying the antecedent but not the consequent can be found" is true, i.e., a *counterexample to the implication* exists (Durand-Guerrier, 2003). In other words, logical implication is characterized as a statement which is falsified only by the counterexample instantiating the statement P is true and the

statement Q is false. In essence, a mathematical implication relates a consequent Q as a logical consequence of the antecedent P bounded by mathematical properties which are relevant to the antecedent and the consequent.

As an immediate consequence, a logical implication is true when sets of mathematical object are defined by the antecedent P and the consequent Q and at the same time, none of the defined objects constitute a falsifying counterexample of P and not Q. Inevitably, defining a set of objects involves the quantification of the set using “all”, “some” and “none.” This notion of *generalized conditional* turns out to be congruent with the logico-mathematical criterion of justifying or rejecting conditional statements in which quantifications of the set of object are expressed, either implicitly or explicitly (Durand-Guerrier, 2003, 2008). Related to the implication statements are also other logically related statements which will be introduced when I frame the inquiry in the next chapter.

A subtle difference between the use of the terms, logical implication and mathematical implication, in this study is warranted here. Reference to logical implication foregrounds the logical character of an implication, i.e., a statement is assigned ‘True’ or ‘False’ according to the logical criterion of whether a counterexample exists. It bears no criterion for what sets of objects are being considered and how that counterexample came about. Reference to mathematical implication foregrounds the mathematical character of an implication, i.e., mathematical rules and laws are used as the criterion for establishing whether a counterexample is mathematically possible. For example, “If a number is less than 1, then the square of the number is less than itself” is a logical implication with respect to the dependence of its truth value on the possibility of counterexample but also a mathematical implication with respect to the possibility of counterexample subject to mathematical laws of ‘squaring’ a number. Note that the number -1 is

less than 1 but its square, 1, is more than itself by mathematical laws. By logical criterion, the implication is false due to the existence of a counterexample of -1.

Few if any mathematical results can be established without the use of logical implications. By connecting mathematically meaningful chains of logical implications, mathematical conclusions can be proven with certainty and stand robust to the possibility of mathematical counterexamples (Benacerraf & Putnam, 1964; Jahnke, 2008). However, the understanding of the mathematical certainty and robustness underlying mathematical proof originates from an understanding of logical implications as assertions of mathematical relationship which does not admit counterexamples (Durand-Guerrier, 2003). Hence, students' ability to use and understand logical implication is essential to understand proofs as well as to validate mathematical conjectures and construct formal or informal mathematical proofs.

Students' Difficulty with Logical Implication – a crucial barrier

Unfortunately, students often exhibited little competence in their understanding and use of logical implication in mathematical proving (Coe & Ruthven, 1994; Durand-Guerrier, 2003; Hoyles & Küchemann, 2003; Knuth, et al., 2009; Recio & Godino, 2001). Hoyles and Küchemann (2003) carried out a large-scale one-year longitudinal study to find out how the United Kingdom students' understanding of the logical implications evolved over time. Students were presented with a mathematical implication, "if the sum of two numbers is even, then the product is odd" and its converse "if the product of two numbers is odd, the sum is even". The former was mathematically valid where the latter was not. For the latter mathematical implication, 36% of the students were able to falsify it by providing counterexamples – 8% used a specific counterexample while 28% used more generic counterexamples. As for the former

valid statement "if the product of two numbers is odd, the sum is even", 24% of students regarded the rule as correct but provided justifications based on empirical verification of examples. Only 9 % of students engaged in logical implications in justifying the rule, providing cryptic arguments like "they must both be odd" (Durand-Guerrier, 2003; Hoyles & Küchemann, 2003). At the beginning of the study, 71% of the students treated the implication and its converse as mathematically equivalent implications. After a year of middle school mathematics with emphasis in mathematical proving, over 60% of students still maintained that the mathematical implication is logically equivalent to its converse.

Knuth, Choppin, & Bieda (2009) surveyed the proofs constructed by 40 middle school students after a year-long of a reform-oriented curriculum. Of the six assessment items given to the students, three were posed as implications concerning number properties (e.g. 36% of 6th graders, 30% of 7th graders, and 31% of 8th graders still generated proofs based on specific numerical examples). Instead of producing a logical proof that showed no counterexamples were mathematically possible, they started with a given number and showed that it satisfied the antecedent and the consequent.

Students' difficulty with logical implications seemed to persist despite going through a mathematics curriculum that emphasized proof and proving (Hoyles & Küchemann, 2003). Further analysis of students' interview data showed that most students could only understand logical implications as an implication for the case of the antecedent is true, i.e., when the antecedent is instantiated.

College students' understanding of logical implications also has been shown to be problematic. Durand-Guerrier (2003) surveyed a group of 273 new students and 92 repeating students in a logic and proof course in college mathematics to find out students' lack of

understanding about logical implications. Students' performance indicated that they failed to recognize when valid inferences can be made from logical implications under different circumstances. When the consequent is satisfied or the antecedent is not satisfied, inferences made based on the logical implications are invalid. When given the statement "In a rhombus, the diagonals are perpendicular" and asked to respond to the question "The diagonals of a quadrilateral (A, B, C, D) is a perpendicular. Is it (the quadrilateral) a rhombus?" About 62% of the new students made an invalid inference of a definite "yes" (22%) or "no" (40%) response. Only about 30% of the students gave an indefinite response, noting that some of the quadrilaterals may be rhombus. Students repeating the course did not show any significantly better performance as well – only about 27% gave an indefinite response. They seemed to be unaware of the possible counterexamples that could invalidate their inferences which were derived from the consequent of logical implications. Furthermore, repeating the course in mathematical logic did not benefit the students' reasoning of logical implications.

In sum, students faced obstacles in their reasoning with logical implications that have negatively influenced their performance in mathematical proving. They make invalid inferences by assuming logical equivalence between logical implications and its converse, and by providing empirical verifications to reason about the truth of logical implications. Their lack of ability to engage in deductive reasoning of logical implications hampered their abilities to construct or validate proofs for mathematical statements.

Conditional Reasoning in Selection tasks – Wason's or Other versions

Cognitive science studies have shown similar difficulties with conditional reasoning, i.e., reasoning of logical implications, in arbitrary and abstract contexts (Johnson-Laird & Byrne,

1991; Stenning & Lambalgen, 2004; Wason, 1968). These studies were mostly conducted based on Wason's (1968) or other versions of selection tasks, in which people were asked to reason about an abstract conditional statement. In this type of choice-response task, the subjects are presented with four cards and a conditional statement "If P then Q." The conditional statement usually states its antecedent P and its consequent Q in the form of "If there is ... on one side of a card, then there is ... on the other." The four cards are pictorially presented with each of their upper faces showing either one positive or one negative instance of the specific p or q mentioned in the statement. Thus, the four cards instantiate p , $not\ p$, q and $not\ q$ exhaustively as a whole. In the Wason's version, the statement "If there is a vowel on one side of a card, then there is an even number on the other" and cards showing "A", "K", "4", "7" on the upper faces are used (see Appendix E, practice item 1 for the presentation of task).

The subjects are told that the other invisible side contains information about the counterpart of the conditional statement to the antecedent or consequent shown on the upper face. They are then asked to indicate which cards they must turn over to look for evidence connecting the antecedent and consequent that logically support or reject the conditional rule. In accordance to the notion of logical implication, a conditional rule is falsified by a counterexample of p and $not\ q$. The correct response thus consists of a combined choice of two cards, one instantiating p and the other instantiating $not\ q$, which may possibly be counterexamples to the statement "If p then q ." In the rule used in the Wason's version, a card having a vowel on one side and an odd number on the other would constitute falsifying examples to the conditional statement. The possible card choices are limited to the cards "A" and "7," one being an instance of a vowel and the other an instance of an odd number. The subjects who are aware of this logical criterion recognized that they have to turn over the "A" card, which is the

case of p , and the “7” card, which is the case of *not* q , to find out if they were possibly counterexamples.

The result of the subjects’ performance in the Wason’s task was alarming. Less than 10% of the people tested opted the combination of the cards p and *not* q , which could falsify the conditional rule (Wason, 1968). Most people chose the card combination of p and q , apparently seeking for confirmatory evidence rather than disconfirmation evidence. Similar findings were consistently replicated with various educational levels such as undergraduates, high school graduates, high school and middle school students, using the similar versions of selection tasks (Cheng, et al., 1986; Griggs & Cox, 1982; Jackson & Griggs, 1988; Lawson, 1990; Stenning & Lambalgen, 2004).

Both educational and cognitive science studies showed that students faced a common cognitive challenge in their reasoning with logical implications in both the arbitrary and mathematical context. Both type of tasks required them to justify logical implications, either arbitrary or mathematically meaningful, based on considerations of counterexamples. They were naturally inclined towards empirical verification of the implication based on empirical examples rather than towards refutation based on counterexamples. This posed a challenge to the teaching of logic in helping students overcome these difficulties. Next I will turn to the studies of the effectiveness of conventional logic instructions in students’ reasoning of logical implications.

Effects of Instructions on Students’ Reasoning of Logical Implications

Studies have shown that conventional instruction in logic emphasizing truth tables and construction of abstract proof have not influenced students’ reasoning of logical implications too positively. In a study by Cheng, et al. (1986), 53 students’ performance in their reasoning of

logical implications before and after a 40-hour introductory logic course were compared and found to have no difference. Considering that selection tasks were used to assess students' performance, the result might be due to the difficulty of the task. However, fewer students made errors for a particular version of selection tasks, known as "permission tasks," consistently before and after the class, indicating that task difficulty may not be the root of the cause here.

In this type of "permission tasks", real life contexts are introduced and the implications are posed as rules permitting an actor to take some particular action when some preconditions are fulfilled. One classical example of this type of task is the drinking age problem (Griggs & Cox, 1982; Lawson, 1990). The task describes a real life context in which a police officer is upholding a drinking law, "If a person is drinking beer, the person must be over 19 years old," and wants to check for violations made the customers of a restaurant. Students are asked to choose, on behalf the fictitious officer, which customers to check based on the provided descriptions of their age or drinks. Among the four cards "Drinking beer", "Drinking soda", "16 years old" and "22 years old", significantly more students choose the cards "Drinking beer" and "16 years old" which instantiates the case P and the case not Q, suggesting that contexts which cue a search for counterexamples helped students' reasoning with logical implications.

Another similar study that used less difficult reasoning tasks also reported the effects of conventional logic training in students' reasoning of logical implications (Leighton, 2006). 49 students went through 12-week training in symbolic logic involving quantifiers and implications. Before and after the training, they were then asked to derive, by selection and construction, the valid conclusions from a conditional statement like "If A then B" and a premise involving the antecedent or consequent. They experienced little obstacles in the easier tasks, scoring an average of 7.84 out of 8 points during the pretest and an average of 7.76 in the post test.

However, their performance in difficult tasks improved modestly from an average of 2.74 to 3.06. The effect of conventional logic training has on improving students' conditional reasoning is again shown to be of little effect.

Platt & Griggs (1993) have found that explication of the implication and provision of explicit instructions for seeking violations of the implication could enhance the subject's performance. With clarifications about the meaning of the implication and the relevance of the cards to the rule, subjects tend to choose cards which may constitute possible counterexamples. In addition, subjects were asked to provide reasons for their card choices and to seek violations explicitly in order to direct their attentions to the counterexamples to the logical implication. As a result, over 80% of the subjects chose the logically correct responses.

Effects of Logic Training on Students' Mathematical Proving

Studies of the benefits of logic training to students' abilities in mathematical proving are moderately encouraging. Mueller (1975) conducted an experimental study of the effects of teaching logic on 146 high school students' ability to write geometry proofs. These students are divided into six classes sorted into 2 different conditions. One condition had the logic unit taught before the geometry content. The other had the logic training inserted in between the basic and advanced geometry content. These classes were taught by four teachers. After 14 to 16 weeks of instructions in logic and geometry contents, students showed they had acquired some logical knowledge quite successfully and were able to interpret generic axioms to infer the validity of proposed model. However, the teaching of logic units was found to have little impact on students' ability to construct geometry proofs – only two classes taught by the same teacher performed better ($p=.05$). Also, it remained unclear whether students did acquire logical reasoning ability and how this might have impacted (or not) students' ability to construct proofs.

Similar studies of the minimal benefits of logic training to mathematical proving had also been reported previously (see Mueller, 1975 for a review of these studies).

Yet the necessity and possible benefits of logic training to proving have been advocated by several mathematics educators. Epp (2003) suggested that instructions in logical reasoning are required for students to acquire the reasoning principles underlying proof methods such as proof by contradiction and proof by contraposition. Stylianides & Stylianides (2008) suggested that instructions of categorical and conditional reasoning principles through the selection task activity may be a productive approach in equipping students with the necessary skills to engage in deductive mode of proving. Durand-Guerrier (2003) suggested that instructions of logic based on Tarski's semantic approach can benefit students in their understanding of mathematical implications and mathematical proving at large. A small number of studies also seemed to support this educational stance, though the benefits seem to be limited to students with high math abilities (Mueller, 1975; J. L. Platt, 1967).

In sum, despite many educators' support, a strong empirical case for the efficacy of logic training has not yet been developed, though anecdotal evidence of success was reported (Epp, 2003; Selden & Selden, 2003). One possible explanation, that can be derived when considering the mismatch between the truth table approach and the required search for counterexample associated with logical implications, is the lack of emphasis on counterexamples in logic training.

Students' Difficulty with Mathematical Proving

Similar issues seemed to persist in students' mathematical proving. When asked to provide proofs to justify a mathematical assertion, most of the students engaged in empirical

verification, i.e., construct a number of concrete examples that verify the assertion instead of property-based or general deductive arguments (Balacheff, 1988; Harel & Sowder, 1998; Healy & Hoyles, 2000; Simon & Blume, 1996; van Dormolen, 1977).

College and pre-college student held different conceptions of proofs but they mostly lacked understanding of the logical deductive character of proofs, especially at the pre-college level (Balacheff, 1988; Harel & Sowder, 1998; Simon & Blume, 1996). A lot of the students provided arguments based on external knowledge authority's approval (e.g., the teacher in class), empirical verifications of selected examples (empirical proofs) (Harel & Sowder, 1998). Pre-service teachers also struggled with deductive reasoning. They prefer to work with empirical examples while making sense of mathematical explanations (Simon & Blume, 1996). They seem not to be cognizant of possible counterexamples – the cases which satisfy the antecedent but not the consequent, which might falsify the mathematical statement or how to ascertain that such counterexamples cannot arise.

In addition, Healy & Hoyles (2000) also found that students held two different conceptions of algebra proofs: one which was convincing to themselves and the other which would get teacher's approval in the form of high test marks. Concerning with the use of empirical verification with examples, the students found them useful in convincing themselves about the truth of the statements. This finding concurred with the majority of the first-year college students' who considered empirical proofs as convincing to themselves but invalid to the public (Segal, 1999). Fischbein (1982) and Healy & Hoyles (2000) also reported that students who appear to understand correctly the deductive proof of a mathematical statement still needed to construct examples instantiating the statement to be convinced.

These studies suggested that students' ability in mathematical proving generally faced the cognitive challenge of moving from empirical verification to property based or generic arguments. They relied on empirical examples to understanding and prove mathematical statements but seldom attend to the need to eliminate possible counterexamples. At the conceptual level, students' difficulty with logical implications and mathematical proving converges towards a common source: students lacked cognizance of counterexamples and its roles in both the reasoning of logical implications and mathematical proving.

Aim of this study: In search for an effective logic training

Taken together, the above review indicates that there are two parts to the question of whether logic training can improve students' reasoning of logical implications and mathematical proving. The first concerns the students' ability in reasoning of logical implications: Conditional reasoning of implication is challenging for students but the existing approach of emphasizing truth table and abstract rules of logical inferences showed little promise in helping students to transit from empirical verification to logical reasoning. The second concerns the cognitive gap between students' ability reasoning of logical implications and mathematical proving: enhancing students' ability in logical implications does not seem to enhance students' proving, yet training students' logical reasoning of mathematical implications were still an educational concern for many mathematics educators (Durand-Guerrier, 2003; Healy & Hoyles, 2000; Hoyles & Küchemann, 2003). Students' proving adheres to an empirical-based scheme and the move to a deductive-based scheme seems to require more than traditional training in logic. At face value, one can claim that improving students' logical reasoning abilities and mathematical proving

abilities are altogether two different issues. Yet both issues can be traced back to a fundamental cognitive issue related to students' mental processing of counterexamples.

This study adopts the educational stance that students' reasoning in logical implications is central to students' ability in mathematical proving, as do many mathematics educators, and work on a hypothesis that productive logic training should emphasize students' active process of finding possible counterexample. Two empirical bases lend support to this research stance. Recent cognitive science studies of subjects' performance in the modified selection tasks had found that subjects' familiarity with available counterexamples and the formulation of the tasks facilitate students' reasoning of logical implication (Cheng, et al., 1986; Griggs & Cox, 1982; R. Platt & Griggs, 1993; Stenning & Lambalgen, 2004). Recent educational reviews of students' understanding of logical implications and proof also concur with the cognitive science findings about the importance of these two factors (Durand-Guerrier, 2003; A. Stylianides & Stylianides, 2009a; G. Stylianides & Stylianides, 2008; Zazkis & Chernoff, 2008).

The driving questions of this study are: (1) What and how can logic training emphasizing counterexamples enhance students' logical reasoning of mathematical implications, in comparison to conventional training approaches? (2) To what extent does logic training impact students' ability in mathematical proving? (3) To what extent does students' ability in logical reasoning of implications impact their ability in mathematical proving? In the following chapter, I will elaborate on the theoretical underpinnings of this study, in addition to the theoretical considerations of logical implications and proving ability.

CHAPTER 2 THEORETICAL FRAMEWORK

In this chapter, I will first present the conceptualization of the study's problem and then the hypotheses of the study and associated research questions. To frame the inquiry, I will first discuss logical implications, followed by mathematical proving. Following that, I will explain the factors affecting reasoning of logical implications, and finally the research questions of this inquiry. The conceptualization of the study's framework is guided by the overarching theme of active mental processing of examples and counterexample which matter to students' reasoning and proving of mathematical implications.

Mathematical Implication, its Logical Variants and Counterexamples

In the previous chapter, I have introduced the notion of logical implication as a *Generalized Conditional*. The statement of a logical implication takes the form of "If [statement P] then [statement Q]" where statement P is also known as the *antecedent* of the implication and statement Q the *consequent*. For a mathematical implication, sets of mathematical objects are quantified by the antecedent and consequent. Hence the logical criterion for the mathematical implication to be true is the non-existence of mathematical counterexample, i.e., a mathematical object satisfying the antecedent P but *not* the consequent Q; otherwise, the implication is falsified. Based on this notion, logical reasoning of implications is the reasoning of the statement in accordance with the criteria of logical truth and falsity. In particular, the statement "If P then Q" is concluded as false when a mathematical object satisfies the antecedent P and not the consequent Q and as true when such a counterexample to the implication does not exist. Other logical variants commonly found in logical reasoning of mathematical implications are defined. These are described as follows.

Converse, Contrapositive and Negation of Implication

For an implication of the form "If [statement P] then [statement Q]", logical variants related to the implication can be constructed by altering the order and the statements P and Q. When the order of statement P and statement Q is reversed and becomes "If [statement Q] then [statement P]," the logical variant, which is still an implication, is called the converse of the implication. A mathematical implication is not logically equivalent to its converse since their counterexamples are logically different. Counterexamples to the implication satisfy the statement P but not statement Q but counterexamples to the converse satisfy statement Q but not statement P.

Another logical variant form "If [statement not Q] then [statement not P]" is said to be the contrapositive of the implication "If [statement P] then [statement Q]." In the contrapositive, the negated consequent of the implication becomes the antecedent and vice versa. A mathematical implication is logically equivalent to its contrapositive since their counterexamples are logically identical, that is, both satisfy the statement P but not statement Q. Note that a contrapositive of an implication is itself an implication.

A negation of an implication is a statement that asserts one (or more) instance which constitutes a counterexample to the implication. For an implication of the form "If [statement P] then [statement Q]," its negation is a statement asserting the existence of one or more counterexamples which satisfy statement P but not statement Q. Using the implication "If the sum of two whole numbers is even, then their product is odd" as an illustration, a counterexample would be two numbers whose sum is even but their product is even. The

negation for the implication would thus be “There is a pair of whole numbers whose sum is even and product is even.”

Conceptualization of Mathematical Proving Ability

Mathematical proving ability has been regarded as proof writing ability in the past (Deer, 1969; Mueller, 1975; J. L. Platt, 1967; Sharon, 1989). However, the focus of the inquiries then was about the ability to present geometric proofs in the two-column format. Recently, the notion of mathematical proving has been revised and expanded beyond to capture a range of abilities related to mathematical proving, which include Proof Construction, Proof Validations and Knowledge of Proof Methods (Alcock & Weber, 2005; Moore, 1994; Selden & Selden, 2003; A. Stylianides & Stylianides, 2009a; Weber & Alcock, 2004).

Some contemporary researchers had made a distinction between the conviction (individual cognitive level) and validity (social aspects) when they look at students' argumentation process (Healy & Hoyles, 2000; Segal, 1999). The conceptualization of mathematical proving in this study, however, leaned heavily towards the context of individual cognition than the classroom context of social interactions. The study is set in an individual mathematical proving environment and social aspects are assumed to be of minimal influence. As such, social aspects of how students regard mathematical proving and validity are excluded from the scope of this study and mathematical proving is considered at individual cognitive level.

Generalized Conditional and Mathematical Proof and Proving

Built upon the *Generalized Conditional* notion of logical implication, mathematical proving is conceptualized as a task of searching for the domain in which the examples and counterexamples of mathematical objects satisfying a mathematical relationship, which is described as an implication. It involves a mathematical process of determining whether mathematical implications are logically true, i.e., the antecedent P will lead to the consequent Q by mathematical laws and no mathematical counterexamples can be found. As students attempted to prove or falsify mathematical implications, deductive inferences are involved and are carried out in a logically valid way. Every inference students make during the proving processes needs to be able to account for the possibility of counterexamples. Any counterexample that may be admitted during the process will render their proving process as invalid or illogical.

At the same time, students may consider examples that verify the mathematical implications. In the process of mathematical proving, student made and organized their inferences in a written form which constituted mathematical proofs for determining the truth of the implications. In addition, students quantified mathematical objects using “all,” “some” or “none” in the implication. This allowed modifications to the proposed statement so that a maximally specified set of mathematical objects satisfy the proposed conditional relationship, should there exist one.

Take for instance, the mathematical statement “If a number is less than 1, then the square of the number is less than itself.” The set of mathematical objects in the antecedent implicitly refers to the set of all real numbers and excludes imaginary numbers. The case of the number -1 will falsify the implication since the square of -1 is 1 and does not satisfy the consequent “the

square of the number is less than 1”. However, it does hold for a set of numbers which one can mathematically determine to be greater than -1.

Definition of Mathematical Proofs

While some researchers have adopted a wider definition of proofs as arguments that “remove one’s own doubt” to include arguments based on empirical verification and crucial experiments (Balacheff, 1988; Harel & Sowder, 1998), this study limits the proof definition to those that regard deductive arguments as the only “valid modes of argumentation” (A. J. Stylianides, 2007) and instead uses the word “empirical proofs” to refer to proofs based on inductive arguments and empirical verifications. Variety and flexibility for the presentation of the proof still remains since the criterion for valid mode of arguments does not restrict the nature of the representations used to develop the proof. In other words, the symbolic and formal type of proofs produced by mathematicians or college mathematics seniors is not the only acceptable genre in this study. How the variety within and in between deductive proofs and empirical proofs are distinguished are achieved through a coding scheme developed more elaborately based on Balacheff’s (1988) classification of proof schemes or van Dormolen’s (1977) three levels of proof characterizations.

Mathematical proofs are products of Mathematical proving. However, I use the term “Mathematical Proving” in this study, with a wider meaning than its usual connotations to proof productions (Harel & Sowder, 1998; Weber, 2001), to include other proof-related abilities which I will define in the next few sections, namely, Proof Validation and Knowledge of Proof Methods, or in short, Proof Knowledge. These two other abilities had been proposed as

important for students' proof and proving (Alcock & Weber, 2005; Epp, 2003; Selden & Selden, 2003)

Proof Construction

Proof Construction refers to the students' ability to construct deductive arguments that connect the given mathematical premises to the conjectured mathematical conclusions (A. Stylianides, 2007). It is similar to proof writing ability with regards to the theoretical interests in the students' ability to produce mathematically coherent proofs in written forms. However, Proof Construction places the analytic emphases on how students use their mathematical knowledge to connect the arguments in a proof, on top of the logical validity of the proof (Weber & Alcock, 2004). Also, it does not limit the proof format only to the two-column format. Types of proofs include formal or informal representations such as mathematical notations or diagrams, arguments presented in a narrative form, etc. Hence Proof Construction involved the interpretation of mathematical objects relevant to the implication, the representations of mathematical objects, and the use of these mathematical representations to make logical and deductive connections of mathematical statements.

Proof Validation

Proof Validation refers to the students' ability to evaluate a presented proof for its validity as a mathematical proof (Alcock & Weber, 2005; Selden & Selden, 2003). In Proof Validation, the students have to evaluate for use of invalid mathematical properties and/or fallacious logical principles in the proof. The cognitive demand of construction and validation of inferences affected students' performance differently (Leighton, 2006). Students' ability in

validation of proof was also shown to be relatively independent of their ability in construction of proof (A. Stylianides & Stylianides, 2009a; Weber, 2010).

One crucial indication of students' ability is their evaluation of Empirical Proof (Balacheff, 1988; Harel & Sowder, 1998). In this type of (invalid) proofs, one or more instances were generated and checked if they satisfied the implications. If none of the instances were falsifying the implication, the implication was concluded as mathematically true, which is logically unsound since not all instances were proven to satisfy the implication. However, students often made the logical error of regarding it as a valid proof because of the pattern of verification presented. Students' recognition of the invalidity of the empirical proof is thus a positive indication of their ability to validate proofs.

One common indirect proof method taught in transition-to-proof classes are Proof-by-contradiction. Otherwise known as the method of *reductio ad absurdum*, the proof started by first assuming the existence of a counterexample satisfying the antecedent but defying the consequent, and then proceeds to conclude that such a counterexample is mathematically and logically impossible, thereby establishing the original implication. On the apparent surface, the proof seemed irrelevant and was likely to throw students off because it assumed the negation of the implication. Students' ability to disregard the apparent mismatch but regard the proof as valid is thus a strong indication of their ability in evaluating proofs (Antonini & Mariotti, 2008).

Knowledge of Proof Methods

Another aspect of proving ability is the students' ability to recognize the logical non-equivalence between the proofs for the implication and its converse as well as the equivalence between the proofs for the implication and its contrapositive (Epp, 2003; Moore, 1994; A.

Stylianides, Stylianides, & Philippou, 2004). The contrapositive statement “If not Q then not P” is logically equivalent to the implication “If P then Q” because both are falsified by the identical counterexamples satisfying “P and not Q.” The argument starts by assuming the consequent of the original implication is not true and proceeds to conclude the antecedent is not true. Both methods have been reportedly difficult for students to understand the logical principles underlying its equivalence to the direct proof method (Antonini & Mariotti, 2008; Epp, 2003; Goetting, 1995; A. Stylianides, et al., 2004).

All three above aspects are related to the notion of mathematical proving as the search for examples or counterexamples of mathematical objects related to the implication. Proof Construction is the ability to construct a systematic and logical search for examples and counterexamples to the implication. Proof Validation is the ability to evaluate the logical coherence of the mathematical search being carried out. Knowledge of Proof Methods (Proof Knowledge) is the ability to recognize other logically equivalent and non-equivalent alternatives of the mathematical search.

Leveraging the use of Counterexamples in Reasoning and Proving of Mathematical Implications

Under the notion of generalized conditional, reasoning of logical implications inevitably requires students’ cognitive efforts in searching for and/or constructing mathematical examples and counterexamples to determine the truth of the implication. I will present cognitive science and educational research work which help to conceptualize the study’s proposal to leverage the use of counterexamples in improving students’ reasoning of logical implications and ability in mathematical proving.

Constrained Example Generation

Educational studies have also stressed the importance of counterexample in helping students' learning mathematical proving (A. Stylianides & Stylianides, 2009a; Zazkis & Chernoff, 2008). Counterexamples are instrumental in eliciting cognitive conflict in students to help them realize that empirical proofs are insufficient to establish conjectured mathematical implications. However, teaching experiments suggested these counterexamples must be within the potential cognitive reach of the students (A. Stylianides & Stylianides, 2009a).

One way of obtaining counterexamples useful for students' reasoning and proving is through the approach of constrained-example generation (CEG) (Rissland, 1991), where students are asked to generate mathematical instances that incorporate specific features but simultaneously exclude other features. In this study, generating counterexamples to a logical implication is considered to be a CEG process of generating instance which incorporates features specified by the antecedent but excludes the features specified by the consequent.

Counterexamples generated by students suggest that these counterexamples are within students' cognitive reach and mathematical expertise, and are likely to facilitate both their reasoning of logical implications and ability in mathematical proving. In the event that the logical implication is mathematically true and absent of counterexamples, students may be able to explain why it is impossible to have counterexamples, the success of which reflects students' understanding of the principle of *reductio ad absurdum* and also their knowledge of the mathematics domain involved. Whichever is the case, with all other factors being equal, the cognitive task of generating examples presumably suffices to increase students' cognizance of possible counterexamples and facilitate their reasoning of logical implications and ability in

mathematical proving, in particular, a transition from empirical-based proving to deductive-based proving.

Enhancing CEG through task formulations

Cognitively oriented studies have also suggested that the implication statement, its content and subjects' knowledge of the reasoning task guided their reasoning processes (Inglis & Simpson, 2006; Johnson-Laird & Byrne, 1991; Stenning & Lambalgen, 2004). As such, the formulation of the reasoning tasks influence students' generation of possible counterexamples, which in turn facilitated students' reasoning of logical implications (Cheng, et al., 1986; R. Platt & Griggs, 1993; Stenning & Lambalgen, 2004). In particular, formulations of logical implications as checking violations of rules in permission and obligation situations had been found to be quite successful in evoking students' reasoning schemas for logical reasoning. (Cheng & Holyoak, 1985; Cheng, et al., 1986; Stenning & Lambalgen, 2004).

Cheng, et al. (1985) found that students' reasoning of logical implications can be enhanced greatly by the evocation of certain schemas through the formulation of tasks. These schemas are understood to be clusters of abstract rules for situations involving permission and obligations. In permission situations, taking a particular action requires certain preconditions to be fulfilled. In obligation situations, the occurrence of certain conditions incurs the necessity of taking some follow-up actions (Cheng, et al., 1986). Violations of rules in permission situations are instantiated by cases in which an actor takes an action without the preconditions being fulfilled. Violations of rules in obligation situations are instantiated by cases in which an actor fails to take up necessary follow-up actions when the conditions do occur. These situations heightened the subjects' cognizance of possible violations of the permission or obligation rule

and guided the subjects to choose the two correct cards. For example, a membership rule “If one has been a member for at least five years, then one must have voted in the past elections” upheld in an obligation situation would heighten students’ cognizance to cases of possible violations in which either someone had been a member for five years or had not voted in the past. Guided by these schematic rules, the subjects tend to make correct choices of cards to turn over that corresponds to the general conditional requirement (Cheng, et al., 1986; G. Stylianides & Stylianides, 2008).

In sum, formulations of the reasoning tasks that facilitate interpretation of the implications and the context enhanced subjects’ cognizance of possible violations or counterexamples. Some math educational researchers had also advocated for an instructional application of these findings to help students improve their mathematical proving (Epp, 2003; G. Stylianides & Stylianides, 2008).

The Research Questions

In a nutshell, this study hypothesizes that logic training emphasizing generation of counterexamples can bring beneficial effects to students’ reasoning of logical implications as well as students’ ability in mathematical proving (Cheng, et al., 1986; R. Platt & Griggs, 1993; Stenning & Lambalgen, 2004; G. Stylianides & Stylianides, 2008). Furthermore, the formulation of the reasoning tasks can further enhance their logical reasoning by evoking their reasoning schemas of permission and obligations rules or logical interpretation. A corollary to the main hypothesis is also within the scope of this study’s interests: The effect of logic training using formulations that evoke reasoning schemas of permission and obligations should benefit students’ reasoning and proving more than the other which evoke logical interpretation. In

addition to the above hypotheses, an exploratory inquiry of how students modify a falsifiable mathematical implication to a mathematically true implication using the self-generated counterexamples may provide additional insights about the role of counterexamples in students' reasoning of logical implications as well as mathematical proving. In the event that the results of the study did not support the hypothesis, the analysis of the interview data might account for other possible factors.

Driven by the abovementioned hypotheses, this study is inquiring the following research questions:

- 1) Compared to the conventional approach, how does logic training emphasizing generation of counterexamples affect students' reasoning with logical implications across different formulations?
- 2) Compared to the conventional approach, how does logic training emphasizing generation of counterexamples affect students' validation of proofs across different formulations?
- 3) Compared to the conventional approach, how does logic training emphasizing generation of counterexamples affect students' construction of proofs across different formulations?
- 4) Compared to the conventional approach, how does logic training emphasizing generation of counterexamples affect students' Knowledge of Proof Methods across different formulations?
- 5) To what extent does students' reasoning of logical implications correlate with their ability in mathematical proving?

- 6) How do students modify a falsifiable mathematical implication to a mathematically true implication based on their self-generated examples and counterexamples?

Research question (1) explores the main hypothesis about the benefits of incorporating counterexamples generation into logic training on students' reasoning of logical implications. Research questions (2) to (4) queries the main hypothesis about the benefits of incorporating counterexamples generation into logic training on students' ability in mathematical proving, which is further distinguished into the three distinct aspects, construction, validation and Knowledge of Proof Methods, of the ability. Corollaries to the main hypothesis are also addressed by these four research questions. Research question (5) examines the correlation between students' logical reasoning and mathematical proving, which is a rekindled longstanding issue (Deer, 1969; Mueller, 1975; Platt, 1969; Hoyles and Kuchemann, 2003; Durand-Guerrier, 2003, Epp, 2003; Inglis, 2008). Research question (6) explores the strengths and limitations of self-generated counterexamples in students' reasoning of logical implications and supplements the answers to research questions (1) through qualitative inquiry.

CHAPTER 3 METHOD

In this chapter I will first describe the design of the study guided by my research hypothesis. Following that, I will describe the components of my data collection processes including the subjects involved, the design, the test instruments, the training materials and the procedure. Next I will elaborate on the process of coding students' data during the assessments, including the coding schemes, and video and written data of the post-study interview. Finally, I will outline the process of analyzing the coded data.

Subjects

The subjects of this study came from three Singapore school sites. The national rankings of these three schools in year 2010, based on their minimal required entry scores for enrolment were between the upper and lower quartiles. Students participating in the study were Secondary 3 students (equivalent to ninth graders) taking the Singapore-Cambridge General Certificate of Education (Ordinary Level) Mathematics (Syllabus D) as their core mathematics subject. Mathematics (Syllabus D) at Secondary 3 level introduced basic contents in algebra, trigonometry, arithmetic, rate and proportion, and graphs. In addition, the students also took another mathematics subject called Additional Mathematics, which placed heavy emphasis on algebraic thinking and computations, quadratic and trigonometric functions and graphs, and basic calculus at this level.

A total of 60 students from the three sites participated in the study. 13 students came from the first site, 39 from the second and 8 from the third. They were recruited through in-class invitations and flyer distributions. I gave a short presentation and a brief question-and-answer session to clarify the purpose of my study and to address students' concerns about the expected

time and efforts, and the level of challenge. They were assured that the study was not related to any academic assessments. To meet the targeted sample size for my study, I approached the third school for students' participation. Their participations were voluntary and were briefed about the irrelevance between their academic performance and the performance in this study. Book vouchers were given to the participants upon completion.

Design

This study used a pretest–intervention–posttest design over a contact period of four days, incorporating three training conditions: Control (C), Permission/Obligation (PO) and Wason (W). Table 1 below shows an overview of the study's design. The first column stated the time and activities for the participants in each condition.

Table 1: Overview of the design

Procedure/ Condition	Control (C)	Treatment (generate counterexamples)	
		Permission and Obligation (PO)	Wason (W)
Day 1 : Pretest	Pre-test		
Day 2: Training I	Training using Logical truth tables with application to Proof Construction.	Training using selection tasks involving permission and obligation situations.	Training using Wason selection tasks
Day 3: Training II	Proof construction and proof evaluation	Similar selection tasks, generation of violations of the rule and proof evaluation	Similar selection tasks, generation of counterexamples to the implication and proof evaluation
Day 4: Post test/ Post - study interview	Post-test and Semi-structured interview of randomly selected students		

The study inquired about the impact of logic training on students' logical reasoning and proving of mathematical implications. Each student was randomly assigned and distributed equally to each of the three (one control and two treatments) conditions (Control, Permission and Obligation, and Wason) with a counterbalanced design of the test sets over the test trials. Two sets of test instruments were used for the pretest and the post test. Based on the two-way design of three conditions (Control and two treatment) by two test sets (Set 1 and Set 2), 10 students were randomly assigned to each cell of the 3 by 2 block resulting in equal-sized sample cells across conditions and test sets. Each student attempted one set of test during the pretest and the other isomorphic set of test during the posttest. In the event that any student withdrew their participation unexpectedly during the study, students were randomly selected from the available pool to fill the vacancies.

During the two days of training, they were trained for logical reasoning of implications based on different approaches that subsequently led to either Proof Construction, generation of violations or counterexamples to the rule. The training used in the Control condition was a simplified version of using logic truth tables to illustrate and apply the process of logical reasoning of implications to mathematical Proof Constructions using worked-out examples. Learning mathematical proving from worked-out examples of mathematical proofs had been shown to be effective (Hilbert, Renkl, Kessler, & Reiss, 2008). The worked-out examples served as a means of cognitive modeling when the solution did not merely present an algorithmic process of proving but also explained the underlying thought processes of making mathematical inferences (Collins, Brown, & Newman, 1989; Schoenfeld, 1985). The training used in the two treatment conditions (PO and W) were adapted from the approaches used by Lawson (1990) and Cheng et al. (1985; 1986) to elicit students' logical reasoning based on counterexamples and rule

violations through solving the selection tasks. Similarly, worked-out examples illustrating the reasoning processes based on counterexamples and rule violations were provided.

In the Control (C) condition, logical truth tables were introduced to highlight the logical relation between the antecedent and consequent of an implication. Subjects then applied their understanding to construct chains of deductive inferences that derived the consequent from the antecedent to prove mathematical implications. In addition, subjects also worked on proof evaluation task(s) on the second day of training. Students' learning outcomes in this condition served as a baseline for comparing the effects of the other two treatment conditions.

In the Permission and Obligation (PO) condition, students worked on selection tasks formulated using contexts involving permission and obligations on the first day of logic training. On the next training day, students worked on similar tasks and proof evaluation tasks, each with an additional request of generating violations to the rules introduced.

In the Wason (W) condition, students worked on Wason's version of the selection tasks on the first day of training. Similar to the treatment in PO condition, students worked on similar tasks and proof evaluation tasks, each with an additional request of generating counterexamples to the implications introduced, on the next training day.

After the post test, four students were randomly selected from each condition and interviewed about how they attempted the proof-related items in the posttest and how they proved or disproved the mathematical implications, and modified falsified mathematical implications. This exploratory inquiry provided additional insights about the role of examples and counterexamples in students' logical reasoning and mathematical proving of implications.

Materials

Pre-test and Post-test instruments

The two sets of test instruments comprised four Wason's selection task items, two items of Deductive-proof Construction for proving mathematically true implications, two items of Proof-by-counterexample for falsifying mathematically false implications, two Proof Validation items and two Knowledge of Proof Method items (Table 2). Isomorphic items were constructed and matched according to the nature of each item and its mathematical topic.

Table 2: Composition of Test instruments

No.	Test Set 1	Test Set 2	Purpose	Content of Implication
1 to 4	Wason Selection tasks	Wason Selection tasks	Logical Reasoning of implication	Non-mathematical
5	Deductive-proof Construction	Deductive-proof Construction	Proof Construction	Elementary Number Theory
6	Proof by counterexample Construction	Proof by counterexample Construction		Elementary Number Theory
7	Deductive-proof Construction	Proof by counterexample Construction		Quadratics
8	Proof by counterexample Construction	Deductive-proof Construction		Quadratics
9	Invalidation of Empirical Proof	Invalidation of Empirical Proof	Proof Validation	Elementary Number Theory
10	Logical non-equivalence of converse	Logical non-equivalence of converse		
11	Logical equivalence of contrapositive	Logical equivalence of contrapositive	Proof Knowledge	
12	Validation of Proof-by-contradiction	Validation of Proof-by-contradiction		

The first four selection task items (labeled as Questions 1 to 4 in the test instrument) assessed subjects' logical reasoning of implications. As described earlier in Chapter 1 (pp. 23-24), the subjects were presented with four cards and a conditional statement in the form of "If there is P on one side of a card, then there is Q on the other," where P and Q were propositional statements. The four cards instantiated P , *not* P , Q and *not* Q of the implication. They were asked to indicate which cards they must turn over to look for evidence that supported or rejected the implication logically, which amounted to the correct combination of the cards that instantiated P and *not* Q (refer to Appendix A and Appendix B for details of the test sets). These tasks were established and widely used for assessing logical reasoning of implications (e.g., Cheng, et al., 1986; Jackson & Griggs, 1988; Stenning & Lambalgen, 2004).

The next four items were the Proof Construction items that assessed subjects' ability to construct mathematical proofs for mathematically true and false implications. Given a mathematical context and a related implication, the subjects were asked to decide whether it was true or false. They were further asked to justify their conclusions using the most convincing argument. For mathematically true implications, students were expected to construct deductive proofs (Deductive-proof Construction). For mathematically false implications, they were expected to construct counterexamples (Proof-by-counterexample Construction). All four items were phrased similarly, as illustrated by Figure 1 below. Each item described a mathematical situation and posed a mathematical implication. The students were prompted for their conclusion and justifications with a logically neutral tone (see Appendix A and Appendix B for details of the test sets)

6. A prime number is a whole number that has exactly two factors, 1 and the number itself (Note that 1 is not a prime number since it has only one factor). Two positive numbers, which may or may not be prime numbers, are added together. Decide whether the following rule is true or false:

If two prime numbers are added together, then the sum is an even number.

Justify why your conclusion must be true or false using the most convincing argument.

Figure 1: A Proof Construction item in the test set

Of these four items, two involved elementary number theory content whereas the other two involved quadratics content. Each content group consisted of one mathematically true and one mathematically false implication, and was numbered as shown in Table 2 above. In the other isomorphic test set (Test set 2), the order of the quadratics items was reversed to minimize any bias due to ordering. While all items of Test set 1 were isomorphic to the same-numbered item of Test set 2, item 7 of Test set 1 was isomorphic to item 8 of Test set 2 and likewise for item 8 of Test set 1.

The last four items assessed students' Proof Validation and Knowledge of Proof Methods (or Proof Knowledge). In the Proof Validation items, students were presented a mathematical implication and a proof, which was either an Empirical Proof or a Proof-by-contradiction, and asked to justify their validity. In the Empirical Proof item, three instances of numbers were shown to be verifying a mathematically false implication, e.g., "If n is an even number, then $n^2 + 7n + 7$ is a composite number." Students were asked to decide if the implication could be concluded as true based on the three verification instances. In the Proof-by-contradiction item, the negation was assumed to be true and a mathematical contradiction was drawn to conclude that the negation is mathematically impossible. For example, students were presented with the implication "Let x and n be two real numbers. If $x > 0$ and $n > 0$, then $\frac{x}{n} + \frac{n}{x} \geq 2$ " and its Proof-

by-contradiction which assumed the negation “*There exist a pair of numbers a and b such that $a > 0$, $b > 0$ and $\frac{a}{b} + \frac{b}{a} < 2$* ” that led to a mathematical contradiction of $(a - b)^2 < 0$. The original implication was then concluded to be true. Students were asked to justify whether and why the conclusion was valid.

In the Proof Knowledge items, students were asked to determine if a given mathematical implication was logically equivalent to its converse or contrapositive. In the former, students were presented with both an implication and its converse spoken by fictitious characters and asked to decide if both statements were expressing the same mathematical idea (e.g. “Gabriel says, ‘*If the product of two whole numbers is odd, then their sum is even.*’/ Dewey says, ‘*If the sum of two whole numbers is even, then their product is odd*’”). In the latter, students were asked to decide whether the truth of the contrapositive was the same as the implication, as proposed by a fictitious character, and why (e.g., “Henry says that the truth of [‘*Let N be an integer. If N^2 is odd, then N is odd*’] is the same as the truth of this statement: “*Let N be an integer. If N is an even integer, then N^2 is an even number.*”). The latter implication is a contrapositive (“If *not* Q then *not* P ”) since “not-odd” is “even.” In both sets of the test instruments, the Validation of Proof-by-contradiction was placed as the last item due to its bulk content (see Appendix A and Appendix B for details).

The content validity of the implications in all eight proof-related items was carefully considered. To avoid students’ under-performance in the assessments due to lack of mathematical knowledge, information were gathered from the textbooks and conversations with their teachers to verify that these mathematical content were taught before the study. In all cases, the topics and terms of Algebra, Elementary Number Theory and Quadratics used in the test sets

were within students' reach since they were either introduced or revisited in the current year's lessons in every school site before the time of conducting the study.

Training materials

A total of 18 training items were given in the training materials. Eight items of the materials were used for the first day of training, and the next ten items were used for the second day of training. Training in all three conditions took the form of self-paced problem solving with written solutions being provided for students' learning. The first 16 items (eight for the first day and the next eight for the next day) were practice problems on logical reasoning and its application to proving of implications, either using selection tasks in the two treatment conditions (PO and W), or problems on logical truth tables and mathematical Proof Construction in the control condition. The last two items were proof evaluation problems, which were introduced towards the end of training on the second day. Table 3 below shows an overview of the design of the training materials across three conditions. Details of these 18 practice problems and implications used in each of the three conditions are provided as follows.

Table 3: Design of training materials

Day	Content/ truth of implication	Number and Types of Practice Problems		
		Control (C)	PO treatment	W treatment
1	Non-mathematical; false	Two problems of logic truth table	Eight selection tasks (real-world context); Two were different; Six were modified to permissive/ obligatory rules	Eight selection tasks (Wason's version) ; Same implications as Control condition;

Table 3 (cont'd)

	Mathematical; false	Two Proof-by-counterexample Constructions		
	Mathematical; true	Four Deductive-proof Constructions		
2	Non-mathematical; false	Two problems of logic truth table	Eight selection tasks (real-world context) with generation of rule violations; Two were different, Six were modified to permissive/obligatory rules	Eight selection tasks (Wason's version) with generation of counterexamples; Same implications as Control condition
	Mathematical; false	Two Proof-by-counterexample Constructions		
	Mathematical; true	Four Deductive-proof Constructions	Two proof evaluations with generation of rule violations	Two proof evaluations with generation of counterexamples
		Two proof evaluations		

Training Materials used in the Control condition

The materials used in the control condition mirrored the typical logic training approach involving logic truth tables and mathematical Proof Constructions. For each day of training, the materials started with reading materials and worked-out examples of the logic truth table of an implication “If P then Q ” and its rules of logical inferences, which was adapted from the instructional materials used by Cheng, et al.’s (1986) study (see Appendix D for the actual content). Following that, two practice problems of applying the logical truth table were posed using specific non-mathematical implications.

The next six practice problems in each day’s training consisted of two Proof-by-counterexample Constructions and four Deductive-proof Constructions. Each problem was formulated in an isomorphic manner to the Proof Construction items in the test instruments; it started with a concise description of the mathematical situation before prompting for justification

for an implication statement. A worked solution accompanied with written explanations was provided on the following page for students' learning. The worked solution demonstrated the logical construction of a deductive proof or Proof-by-counterexample (see Figure 2).

Practice No.13

A set of five positive whole numbers are randomly chosen and their average is calculated. Decide whether the following implication statement is true or false:
If the five whole numbers are consecutive in order, then their average is a whole number.

Justify why your conclusion must be true or false using the most convincing argument.

Solution for Practice No.13

Let's call the first whole number as n , then any five consecutive numbers are namely, n , $n + 1$, $n + 2$, $n + 3$ and $n + 4$. By applying algebraic rules, the average of these numbers is calculated as:

$$\frac{n + (n + 1) + (n + 2) + (n + 3) + (n + 4)}{5}$$

$$= \frac{5n + 10}{5}$$

$$= \frac{5(n + 2)}{5}$$

$$= n + 2$$

Figure 2: A Proof Construction practice problem and its solution (Control Condition)

The last two proof evaluation practice problems, which only appeared on the second training day, were adapted from the proof evaluation tasks for studying students' algebra proof schemes (Healy & Hoyles, 2000). Each problem presented a mathematical implication and "proofs" provided by fictional characters, and asked students to choose one that best justified the statement. Explanations were provided on the following page to explain why some of the proofs were logically valid (see Appendix D to Appendix F for more examples of explanations).

Training Materials used in the PO condition

The materials used in the PO condition were adapted from the instructional material used by Cheng et al. (1985; 1986) in training students' reasoning of implications. Eight selection tasks involving two non-mathematical and six mathematical implications, which were formulated in permissive or obligatory contexts, were used for the first day of training. Each problem began with a description of a real-world situation, in which an implication "*If P then Q*" was formulated as an obligatory or permissive rule to be fulfilled, before asking students to check the four card choices for possible rule violations. Figure 3 below shows an illustration of formulating the practice problem No.13 in the control condition as a selection task with an obligatory rule (part (i) was not given on the first day). A worked solution accompanied with written explanations was provided on the following page to explain why the cards that instantiated the antecedent (P) and the negation of the consequent (not Q) might constitute a violation to the rule (see Appendix F for examples of explanations of the PO condition).

Practice No.13

You are helping your friend in checking some flash cards she made for investigating the average of whole numbers. She wrote any five positive whole numbers on one side of the flash cards and calculate their average on the other side. You want to make sure that her cards follow the mathematical rule, *"If the five whole numbers are consecutive in order, then my friend must get a whole number for the average."*

i) An instance violates the above rule when a set of five numbers are consecutive and their average is not a whole number. Can you think of such an instance of violation?

___ Yes. ___ No.

If "Yes," what might the five consecutive numbers be?

If "No," based on what you know about the average of numbers and consecutive numbers in general, why is it impossible to find an instance that violates the above rule?

[---Blank Space for students' responses---]

ii) Which of the four card(s) below would you need to turn over to check if your friend's work has violated the rule? Turn only those which you need to check. Tick the card(s) you want to turn.

33, 34, 35, 36, 37	67, 20, 42, 54, 36	The average is 45.2	The average is 25
(a)	(b)	(c)	(d)

Figure 3: Practice problem involving obligation situation

The second day of training focused on applying the idea of rule violation to generation of instance of violations. Eight similar selection tasks (two non-mathematical and six mathematical implications again) were presented except that, in each practice problem, students were to generate an instance of rule violation, if possible, or explain why violations were impossible. Of these eight selection tasks, instances of violations to the permissive or obligatory rules in the first four tasks (two non-mathematical and two mathematical) could be constructed, whereas violations to the permissive or obligatory rules in the next four tasks were mathematically impossible to construct. In addition to the solutions and explanations concerning the correct card

choices, the written explanations provided on the next page also demonstrated how an instance of violation was possible or impossible to construct based on the concurrent constraints of fulfilling the antecedent (P) but not the consequent (not Q). Establishing the mathematical impossibility of this constraint generation of an instance led to the conclusion of the rule being complied (see Figure 4).

Solution for Practice No.13 part (i)

To find an instance that violates the rule, you want to look for five consecutive positive whole numbers whose average is not a whole number. Let's call the first whole number as n , the five consecutive numbers are namely, n , $n + 1$, $n + 2$, $n + 3$ and $n + 4$. Since the average is not a whole number after the sum of these five numbers is divided by 5, it means that the sum of n , $n + 1$, $n + 2$, $n + 3$ and $n + 4$ is not a multiple of 5. That is, $n + (n + 1) + (n + 2) + (n + 3) + (n + 4) = 5n + 10$ is not a multiple of 5. So we have an instance of violation when we find a number n such that $5n + 10$ is not a multiple of 5.

But $5n + 10 = 5(n + 2)$, which actually means 5 multiplies a whole number (n is a whole number so $n + 2$ is also a whole number). It's impossible to pick a number n without letting $5n + 10$ become a multiple of 5. Thus we cannot find five consecutive whole numbers whose average is not a whole number. Because we have no instance that *violates* the rule, the rule is *complied*.

Figure 4: Solution for Why Violations were impossible

The last two proof evaluation practice problems were similar to the Control condition except that, with respect to the PO condition, student were prompted to generate an instance of violation or explain why such an instance was impossible, before being asked to choose one “proof” that best justified why violation was impossible. Worked solutions provided also explained the mathematical impossibility of generating an instance of violation and which “proof(s)” to choose (see Appendix E for examples of explanations of the W condition).

Training Materials used in the W condition

The materials used in the W condition were configured similarly to the PO condition – eight selection tasks (two non-mathematical and six mathematical implications) on the first day of training, eight selection tasks (two non-mathematical and six mathematical implications again) with prompts for generating instances or explanations and two proof evaluation practice problems on the next day. However, these selection tasks were formulated according to the Wason’s selection tasks. The implications appeared in these problems were identical to those used in the Control condition (e.g., “*If the five whole numbers are consecutive in order, then their average is a whole number*” appeared in both conditions). Each practice problem, either non-mathematically or mathematically related, began with a description of a situation and an implication, before asking students to choose the cards that helped decide the truth of the implication. Figure 5 below shows an illustration of formulating the practice problem No.13 in the control condition as a Wason selection task (part (i) was not given on the first day). Worked solutions for the correct choices of cards were provided on the following page (see Appendix E for details of solutions).

Practice No.13

Below is shown a set of four cards, of which you can see only the exposed face but not the hidden back. For each card, there is a set of five positive whole numbers written on one of its sides and their average written on the other.

Also below there is a rule which applies only to the four cards.

Rule: *If the five whole numbers are consecutive in order, then their average is a whole number.*

- i) A counterexample makes the rule false when a set of five numbers are consecutive and their average is not a whole number. Can you think of such a counterexample?

☐ Yes. ☐ No.

If “Yes,” what might the five consecutive numbers be?

If “No,” based on what you know about the average of numbers and consecutive numbers in general, why is it impossible to find a counterexample to the rule?

[---Next Page---]

- ii) Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don’t turn unnecessary cards. Tick the card(s) you want to turn.

33, 34, 35, 36, 37	67, 20, 42, 54, 36	Their average is 45.2	Their average is 25
(a)	(b)	(c)	(d)

Figure 5: Practice Problem in Wason Condition

The second day of training focused on applying the idea of counterexamples to generation of counterexamples, instead of rule violations. Students were to generate counterexamples that falsified the implication, or explain why the counterexamples were impossible. Of these eight selection tasks, counterexamples to the implications in the first four tasks (two non-mathematical and two mathematical) could be constructed, whereas counterexamples to the implications in the other four tasks were mathematically impossible to construct. Similar to the PO condition on the second day, the solutions also included demonstrations of how a counterexample satisfying the antecedent (P) but not the consequent

(not Q) was possible or impossible to construct. Establishing the mathematical impossibility of generating a counterexample led to the conclusion of the implication being true (see Figure 6).

Solution for Practice No.13 part (i)

To find a counterexample, you want to look for five consecutive positive whole numbers whose average is not a whole number. Let's call the first whole number as n , the five consecutive numbers are namely, n , $n + 1$, $n + 2$, $n + 3$ and $n + 4$. Since the average is not a whole number after dividing the sum of these five numbers by 5, it means that the sum of n , $n + 1$, $n + 2$, $n + 3$ and $n + 4$ is not a multiple of 5. That is, $n + (n + 1) + (n + 2) + (n + 3) + (n + 4) = 5n + 10$ is not a multiple of 5. So we have a counterexample that makes the rule false when we find a number n such that $5n + 10$ is not a multiple of 5.

But $5n + 10 = 5(n + 2)$, which actually means 5 multiplies a whole number (n is a whole number so $n + 2$ is also a whole number). It's impossible to pick a number n without having $5n + 10$ being a multiple of 5. Thus we cannot find five consecutive whole numbers whose average is not a whole number. Because we have no counterexample that makes the rule *false*, the rule is *true*.

Figure 6: Solution for Why Counterexamples were impossible

The last two proof evaluation practice problems were similar to the other two conditions except that, in consistency with the W condition, student were prompted to generate a counterexample or explain why such a counterexample was impossible, before being asked to choose one “proof” that best justified why counterexample was impossible. Worked solutions provided also explained the mathematical impossibility of generating a counterexample and which “proof(s)” to choose (see the last two items Appendix E for more details).

In sum, the training materials across the three conditions bore some similarities in their design and contents guided by the theoretical framework of the study. The implications of the Proof Construction or selection task practices and the proof evaluation practice problems were alike; statements were identical in both the Control and Wason condition but formulated differently in the PO condition. The non-mathematical implications were identical across both the Control and Wason conditions but different for the PO condition.

The practice problems in both treatment conditions (PO and W) shared common features in terms of the use of selection tasks for the first 16 problems, the initial focus on the logical reasoning and the subsequent focus on generating violations or counterexamples. In contrary, the practice problems in the Control condition consisted of few logical reasoning tasks involving logical truth tables but focused more on demonstration of Proof Construction on tasks which were similar to the Proof Construction items in the test instruments. The last two proof evaluation practice problems were also absent of the emphasis on violations or counterexamples in comparison to the proof evaluation problems in the two treatment conditions.

Students' reflection of learning

All students were asked to write down one to three things that they had learnt at the end of each training day. This reflection task helped students organize their training activities into meaningful learning experiences. Expressing their learning in words also allowed the students to recall what they had learnt previously. Their written summaries served as additional research artifacts of students' learning in each day.

Materials for Post Study Interview

A new Proof Construction task was posed during the semi-structured interview. The task was a generalized version of a practice problem which concerned whether "*If the five whole numbers are consecutive in order, then their average is a whole number*" is true. The antecedent was generalized to "*If the set of whole numbers are consecutive in order, then their average is a whole number,*" where the set of whole numbers had now included at least three members. This implication is false for cases where the set of whole numbers has even number of members, e.g.,

two , four or six consecutive numbers. However, for cases where the set of whole numbers has odd number of members, e.g., three, five or seven consecutive numbers, this implication is true because the sum of these numbers can be shown algebraically to be a specific multiple of the number of members in the set, e.g., a multiple of 3 for three consecutive numbers, a multiple of 5 for five numbers and so on. The specific multiple is in fact the middle number of the set.

Follow-up questions were posed about students' conclusion. For students who concluded the implication as true, they were asked whether counterexamples were possible and why to see if they could produce a proof. They would be prompted to consider other sets of consecutive whole numbers if they had been focusing on sets of three consecutive numbers. For students who falsified the implications, they were asked to elaborate how they determined so. Questions about how students would modify the implication into a true statement and whether they could prove it were asked towards the end of the interview.

Procedure

All data collection took place in the classrooms of the school sites over four to six weeks at each site. Each student participated in the study for four contact days over a two-week span. The pretest and the first day of training took place in the same week followed by the second day of training and posttest in the next. The pretest and the first training were at most one day apart; likewise for the second training and posttest. The lapse between the two training days was at most a week apart. The sessions were held after school to minimize disruptions to the students' class lessons. Students were assigned to different timeslots to maintain a group size of less than 15 at any time. This was to allow the researcher to attend to students' queries promptly and manage students' activities effectively during test and training sessions.

During the administration of pretest and post test, students were seated individually and were given about 60 to 90 minutes to attempt all the twelve items in the test instruments. During the pretest, ten out of 20 students in each training condition were assigned to attempt one test set while the other ten were assigned the other test set 2. During the posttest, all students were assigned to attempt the other set. In situations where students were not sure of their solutions while attempting the tests, they were reassured that the purpose of the test was to understand their thinking processes rather than being correct. They were also cooperative in observing silence throughout the test duration.

On the first day of training, students were given the designated training materials, as described in details previously. They were instructed to read the materials before attempting the practice problems. The students worked on their materials individually in their own pace, attempting each problem at a time sequentially. Once they had completed a problem, they would refer to the solution provided on the next page. If the students had any questions with regards to the contents of the problem or the worked solutions, the researcher would paraphrase the sentences to aid their understanding. In situations where students were unclear about certain mathematical terms, the researcher would provide the mathematical definitions to assist them. My general observation as the researcher was that the atmosphere in the room was light-mood and conducive during these self-learning sessions. Occasionally, students expressed concerns about the correctness of their responses but were reassured that the study was more interested in their thinking and reasoning. At the end of the training, the students were asked to write a summary of their learning on the last page of the materials. Depending on the training conditions assigned to the students, the training session lasted from 45 minutes to 70 minutes.

On the second day of training, students were given their materials from the last training to recall what they had learnt previously, after a few days' lapse. The old materials were then collected from the students after about ten minutes, in exchange of the designated training materials for the second training. The procedure was carried out in the same manner as described above. Depending on the training conditions assigned to the students, the training session took about 70 to 85 minutes. To reduce fatigue, students were prompted to write down their best responses and proceed to the worked solution if they made little progress after spending too much time on a particular problem. Throughout the training, the students could accomplish individual self-paced activities satisfactorily.

Four students were then randomly selected from each training condition for post-study interviews. The interview session was carried out on a one-to-one basis that lasted about nine to almost 23 minutes. All interview sessions were semi-structured leading to how they would modify a falsified statement to a true statement and justify that it is so (see Appendix G for post study interview task and sample interview questions). Students were presented the new Proof Construction task and were left to work on their own for some time until they reached a conclusion. Further inquiries about how students justified and how examples and counterexamples helped support their conclusions were also made before they were eventually asked to modify the provided implication to a mathematically true implication and to prove it.

Data Scoring and Coding

As there were multiple types of items, the data were scored and coded according to the schemes, which are elaborated as follows.

Scoring of selection task items

The students' responses in the selection task items were coded based on the logic index scheme (R. Platt & Griggs, 1993; Pollard & Evans, 1987). For each card choice, 1 point was awarded if a logically correct choice was made; otherwise, 1 point is subtracted. Students' scores in each item thus ranged from -4 to +4 with two-point intervals. The median score of 0 indicated that the student made two logically correct and two incorrect decisions, which reflected the case of majority (more than 90%) choosing the cards P and Q in the past studies. The subjects' total raw scores over four selection tasks hence ranged from -16 to +16 with two-point intervals, constituting more than 15 possible levels within.

Coding schemes for Students' responses to proof items

Logical reasoning and mathematical proving of mathematical implications involved making productive deductive inferences about possible examples of and counterexamples to the implication (Durand-Guerrier, 2003; Stenning & Monaghan, 2004). Coding schemes were constructed to identify how productive students were in making chains of deductive inferences about the examples and counterexamples of the mathematical implication. This might range from almost no consideration of relevant examples, to consideration of isolated examples and counterexamples, and to consideration of all possible examples and counterexamples using deductive inferences. These coding schemes thus assumed a hierarchical ordering of the extent of students' competence that concurs with contemporary views with respect to different aspects of mathematical proving (Balacheff, 1988; Coe & Ruthven, 1994; Durand-Guerrier, 2008; A. Stylianides & Stylianides, 2009b).

Overall, six different coding schemes were developed for analyzing students' Proof Constructions, Proof Validations and Proof Knowledge: (1) Deductive-proof Construction; (2) Proof-by-counterexample Construction; (3) Invalidation of Empirical Proof; (4) Validation of Proof-by-contradiction; (5) Consideration of the logical equivalence of an implication and its Converse; and (6) Consideration of the logical equivalence of an implication and its Contrapositive. Different schemes were mandated by the distinct criteria among various types of proof –related items for scoring students' responses.

Coding for Deductive-proof Construction

Students' deductive- Proof Construction were distinguished into seven levels (labeled level 0 to 6 in an ascending order). In general, the levels can be described as a progression through four phases: attempts that failed to relate the antecedent with the consequent (level 0), proofs that were based on examples or logical errors (levels 1 to 2), proofs that were based on incomplete deductive inferences (levels 3 to 4), i.e., “missing a step” (Lin, 2005), and proofs that were based on coherent deductive inferences (levels 5 to 6). Table 4 below showed the seven levels in the coding scheme. Note that students' responses that met multiple descriptors of different levels would suffice for the highest level code and students' responses that met any descriptor of a level with multiple descriptors would suffice for that level code.

Table 4: Descriptors for coding students' Deductive-proof Constructions

Level	Description of students' proofs
0	Irrelevant or show minimal engagement, i.e., the antecedent and consequent are not related
1	Generates an incorrect counterexample or example to conclude the statement is false or true incorrectly.

Table 4 (cont'd)

	• Generates one or multiple examples to verify and conclude that the implication is true (even if the statement is false) • Engaged in some erroneous form of logical reasoning such as using “if not P then not Q” to falsify the implication “if P then Q” • Derive a property which is not related to the conclusion
2	Generate one or more examples to verify the implication but also: • provides a rationale for the choices of examples by considering examples belonging to different cases of the antecedent • shows evident use of at least one extreme instance • uses mathematical properties inferred from generated examples to make conclusions
3	• Deduce relevant mathematical properties for proving the implication but missing one or two key inferences to deduce the implication. • Deduce the implication to be true for some cases of the antecedent but leave some others out
4	• Generate logical deductions to justify conclusions but one or two inferences may be interpreted as inductive due to insufficient substantiation • Generate logical deductions to justify conclusions but contain minor reasoning errors that may be interpreted as writing error from the context • Inferences made are not organized into a chain of logical inferences
5	Generate logically coherent and mathematically valid proofs
6	Generate logically coherent and mathematically valid proofs with inferences derived through use of mathematical symbols and notations

Students’ responses exhibited little or no deductive inferences at both levels 1 and 2. The distinction between them was based on whether students considered examples of the implications in isolation or in connected groups and, in turn, made productive inferences. Students might consider a mathematical example as a “crucial experiment” for testing the implication or as a representative of various subsets of the mathematical objects in question (Balacheff, 1988; Chazan, 1993). Some students might relate the antecedent to the consequent via inference of irrelevant property or erroneous logical reasoning (empirical-based inferences or invalid form of logical reasoning), which only qualified their responses as level 1.

Students were making productive but incomplete deductive inferences from the antecedent at both levels 3 and 4. The distinction between them was based on whether key inferences were omitted. This usually occurred when one case of the antecedent or one or two deductive inferences were skipped in students' proofs, e.g., deduced the implication for the case of even numbers but not the odd numbers, or skipped the inference of the product of odd numbers is an odd number (Lin, 2005).

At level 4, the chain of inferences connecting the antecedent and the consequent was almost complete. Distinction between levels 4 and 5 was based on whether one or two inferences might be interpreted as non-deductive or had minor errors. For instance, "3 times an odd number x will give an odd number" was not substantiated by "product of two odd numbers is an odd number" and might be interpreted as an inductive inference. Minor errors were either due to writing errors or lack of logical organization in sequencing the chain of inferences. Students' Proof Construction at levels 5 and 6 are logically valid and coherent, with the latter showing concise and clearer proof through the use of mathematical symbols and notations that approaches a formal proof (Boero, 1999; Miyazaki, 2000).

Coding Proof-by-counterexamples Construction for mathematically false implications

A similar scheme was also used for students' Proof-by-counterexamples Constructions for mathematically incorrect implications, except that there were only six ordered levels (levels 0 to 5, in ascending order). Table 5 below showed the six levels of the coding scheme in hierarchical order.

Table 5: Descriptors for coding students' Proof-by-counterexample Constructions

Level	Description of students' proofs
0	Irrelevant or show minimal engagement, i.e., the antecedent and consequent are not related.
1	- Generated an incorrect counterexample or example to conclude the statement is false or true incorrectly. - Generated one or multiple examples to verify and conclude that the implication is true (even if the statement is false) - Erroneous logical reasoning. Engaged in some invalid form of logical reasoning such as using "if not P then not Q" to falsify the implication "if P then Q" - Derive a property which is not related to the conclusion
2	Generate one or more examples to verify the implication but also: - provides a rationale for the choices of examples by considering different cases; - shows evident use of at least one extreme instance; or - uses mathematical properties inferred from generated examples to make conclusions
3	Deduced inferences based on incomplete cases of antecedent from misconceptions. Implication is proven true but would otherwise be falsified if not for the misconception
4	Falsify the implication by constructing one or few specific counterexamples
5	Falsify the implication and describe a general set of counterexamples, identifying the property of the set that falsifies the implication

Apart from level 0, the hierarchical levels posited a general distinction between proofs which considered the implications using examples (coded as levels 1 to 2) and proofs which considered the implication by deductive inferences (levels 3 to 5), with levels 4 and 5 indicating logically valid Proof-by-counterexamples had been constructed by students.

Similar to the previous coding scheme for Deductive-proof Construction, levels 1 and 2 marked little or no deductive inferences due to the use of example-based or logically invalid inferences and their distinction depended on how students conceived the examples of the

implication and made productive inferences. This is akin to incorrect empirical proof attempts to prove a false implication (Ko & Knuth, 2009).

Students' proofs showed productive use of deductive inferences but limited consideration of counterexamples at level 3. For whatever examples or cases of implications they had identified, they were able to make deductive inferences to prove the implication. However, students did not consider a subset of counterexamples due to their misconceptions (e.g., assumed that all prime numbers are odd numbers and did not consider that 2 is the only even prime number), which would otherwise lead to the falsification of the implication.

At levels 4 and 5, students' proofs had successfully constructed one counterexample to falsify the implication. Distinction between levels 4 and 5 was based on students' consideration of counterexamples as isolated instances or as a generic counterexample representing a group of counterexamples sharing a common property (Hoyles & Küchemann, 2002).

Coding scheme for Invalidation of empirical proof

The coding of students' response to this question consisted of three levels (level 0 to 2, in ascending order) which indicated the extent to which students considered possible counterexamples to the mathematical implication. At level 0, students did not exhibit any consideration of possible counterexamples, in their reasons for rejecting or accepting the empirical proof. Some students who generated their own "proofs" for the implication were also coded as level 0. At level 1, students questioned the truth of the implication and concluded that the empirical proof was invalid. They might have constructed a counterexample to the implication, reasoned that more instances needed to be tested for possible counterexamples, or that an established mathematical formula was required to prove the implication. However, these

responses were targeted more towards the truth of the implication than the invalidity of the empirical proof, i.e., such a proof did not verify that all instances were not counterexamples. At level 2, students rejected empirical proof based on the logical ground that the process of empirical proving itself did not verify that all instances were not counterexamples (A. Stylianides & Stylianides, 2009b). Table 6 below shows the three levels of the coding scheme.

Table 6: Descriptors for coding students' Invalidation of Empirical Proof

Level	Description of students' responses
0	Did not reject the proof or reject without relevant reasons
1	Reject the proof as invalid: • by falsifying the implication using a constructed counterexample • by specifying more instances or an extreme value needs to be tested. • due to absence of mathematical formula
2	Reject the proof as invalid because this type of proof is invalid in general terms; all instances needed to be verified but were not or that counterexamples are still possible.

Coding scheme for considering the logical equivalence between implication and converse

The coding of students' response to this question consisted of four levels (level 0 to 3, in ascending order). At level 0, students interpreted converse as the reverse of an implication and hence logically equivalent. Some students who "proved" both the implication and the converse to be "true" were also coded as level 0, indicating little consideration of counterexample. At level 1, students concluded that the implication and its converse were not logically equivalent based on their falsification of the converse, or that the antecedent of both the implication and its converse were referring to different sets of mathematical objects. At level 2, students showed considerations of counterexamples in deciding whether the pair of implications was the same. They constructed a deductive proof for the implication and a counterexample for the converse

(Hoyles & Küchemann, 2002). At level 3, students showed thorough considerations of counterexamples targeted towards a generic implication and its converse. They inferred that the counterexamples for falsifying a pair of implications were not identical and hence not logically equivalent. Table 7 below shows the coding scheme at each level.

Table 7: Coding scheme for the implication and its converse

Level	Description of students' responses
0	Did not reject, reject without logical reason or with logically incorrect reasons, e.g., if p then q is the same as if q then p
1	Conclude that the pair is different because: • one is true and the other is false, shown by counterexample, or • different mathematical objects were referred by different antecedents
2	Concluded that the pair is different by providing deductive proofs for the implication and counterexample for the false statement.
3	Concluded that the pair is different because the counterexamples for the implication and its converse are different in general

Coding scheme for considering the logical equivalence between implication and contrapositive

The coding of students' response to this question consisted of three levels (level 0 to 2, in ascending order) that generally reflected the extent to which students recognize the logical equivalence through comparing examples and counterexamples of the implication and its contrapositive. At level 0, students showed little consideration of examples and counterexamples and rejected the logical equivalence due to irrelevant or mathematically false reasons. Students who "falsified" one of the implications incorrectly using invalid counterexamples were also coded as level 0.

At level 1, students concluded that the implication and its contrapositive were logically equivalent based on the truth of the implications. They concluded both implications were

mathematically true, which was an inadequate explanation of why one implication being true would imply the other to be true, and vice versa. In other words, they no longer relied on the appearances of the sentence to judge logical equivalence (Antonini & Mariotti, 2008). However, they had yet to recognize logical equivalence based on validity of mathematical propositions (Durand-Guerrier, 2008).

At level 2, students identified the basis for logical equivalence being that the counterexamples to both implications was identical, which was the only logically sound explanation in comparison to other levels. Both implications are either true or false concurrently, depending on whether the shared counterexamples exist. Table 8 below shows each level of the coding scheme.

Table 8: Coding scheme for the implication and its contrapositive

Level	Description of students' responses
0	• Irrelevant, mathematically false reasons • Use of invalid counterexample to falsify an implication
1	Concluded that both are logically equivalent because both statements are true and thus both implications have same truth value; proofs are constructed
2	Concluded that both are logically equivalent and explain the equivalence based on identical counterexamples that both implications have

Coding scheme for the validation of Proof-by-contradiction

The coding of students' response to this task consisted of four levels (level 0 to 3, in ascending order) that generally reflected the extent to which students recognized the validity of Proof-by-contradiction. At level 0, students failed to accept Proof-by-contradiction. Some responded with irrelevant remarks or 'substitute' proofs which were mathematically invalid, e.g., empirical verification or proving its converse. Some regarded the proof as "faulty", either

because the negation of the implication was being “proved” or the particular statement that carried the mathematical contradiction was “wrong.” (Antonini & Mariotti, 2008; Thompson, 1996) Others accepted the proof as valid because the mathematical steps in the proof are correct and disregarded the contradiction statement.

At level 1, students accepted the validity of Proof-by-contradiction because they evaluated and agreed with the conclusion that the particular contradiction statement in the proof was mathematically impossible. However, such responses were short of explaining why the implication would then be true.

At level 2, students based their acceptance of the proof on the basis that contradiction led to the falsification of the negation and thus the implication was true. Students' explications about the logical consequence of deriving contradiction from a negation were distinct from level 1 because they indicated a logical relationship between the contradiction statement and the original implication (Antonini & Mariotti, 2008). However, they did not further substantiate why the falsification of the negation led to the implication being true.

At level 3, students accepted the proof based on their recognition that the contradiction in the proof for the negation of the implication had logically eliminated all possible counterexamples and consequently, the original implication was true. Table 9 below shows the coding scheme at each level.

Table 9: Coding scheme for validating the Proof-by-contradiction item

Level	Description
0	- Irrelevant response or response without mathematical reasons, including affective remarks - Failed to accept Proof-by-contradiction based on logical consideration of counterexamples

Table 9 (cont'd)

	Accept or rejects the proof due to empirical verification or invalid counterexamples, the opposite of the implication is assumed or some subsequent algebraic steps being correct or incorrect.
1	Accepts the proof as valid because students evaluate and agree that the particular statement, which contradicts some established mathematical properties, is mathematically impossible.
2	Accepts the proof as valid because the derived mathematical contradiction showed that the logically opposite statement is mathematically incorrect and thus the implication is true
3	Accepts the proof as valid because counterexamples to the implication are impossible due to the contradiction.

Coding of Post Study Interview Data

The interview session was intended to find out how students modify an implication after falsifying it and how they went about proving it. The audio-video data were transcribed.

Episodes of each proving attempt were identified. The student's attempts were coded for the conclusion and the level of the proof constructed. For coding the interviewee's mathematical proving performance in each attempt, the coding scheme of Deductive-proof Construction and Proof-by-counterexample Constructions were used (see Tables 4 and 5 for both coding schemes).

In addition, the implications that students had considered, whether students had been prompted to consider different sets of whole numbers and how they modified the given implication were also identified. Instances related to examples and counterexamples in these episodes were then coded. A coding scheme of the modifications made, examples and counterexamples used were then developed from the interview data.

Analysis

The data analyses served to address the main hypothesis driving the study – logic training emphasizing generation of counterexamples can bring beneficial effects to students’ reasoning of logical implications as well as students’ ability in mathematical proving. Furthermore, formulations that evoke reasoning schemas of permission and obligations should benefit students’ reasoning and proving more than the other. In addition to the above hypotheses, analysis of post study interview also served to inquire the role of examples and counterexamples in students’ reasoning and proving of mathematical implications; in particular, students’ modification of a false mathematical implication to a mathematically true implication. Before presenting the data analysis plan, the preliminary data analysis is discussed.

Preliminary Analyses

Preliminary analyses served to surface any peculiarity in the collected data that might cause concern or ambiguity in interpreting the main findings in the next chapter. The scope of the preliminary analyses encompassed students’ written data from the test instruments, video and written data from the interview and written data from students’ reflection of training. Three aspects were targeted: (1) Reliability of the coding schemes, applied to all written and video data; (2) internal consistency of the test instruments; and (3) Identification of systematic biases prior to the training sessions, if any, in students’ pretest assessment across conditions and test sets. To this end, inter-rater agreement on 20% of the students’ data (12 students) from test items no. 5 to 12, training reflections and post study interview, Cronbach’s alpha of the pretest and post test scores and two-way ANOVA (3 conditions by 2 test sets) of the pretest scores were used.

Interrater agreement ratings

Table 10 shows the inter-rater agreement of the coding schemes for all the 12 items based on 20% of the sample, i.e., 24 out of 120 responses, including pretest and post test, for each item. Each response to a particular item was also coded by a second coder, who was either a mathematics teacher or a mathematics major graduate. The inter-rater agreement ratings for all the proof –related items were satisfactory though some items’ ratings are slightly lower than 80% (see Table 10 below). However, pooled ratings of the items using the same coding scheme, i.e., Deductive-proof Construction items and the Proof-by-counterexample Construction items, turned out to be 85.4%.

Table 10: Inter-rater agreement of coding and Reliability coefficients

Test items	Interrater agreement	Cronbach’s alpha	
		Pretest	Posttest
Logical reasoning Items 1 to 4	100% for all items	0.12	0.866
Deductive-proof Construction Item 5 Item 7 (test set 1)/ Item 8 (test set 2)	83.3% 87.5%	0.523	0.597
Proof-by-counterexample Construction Item 6 Item 8 (test set 1)/ Item 7 (test set 2)	91.7% 79.2%	0.420	0.360
Proof Validation Empirical proof (Item 9) Proof-by- contradiction(Item 12)	83.3% 79.2%	Not applicable	
Proof Knowledge Logical non-equivalence of converse (Item 10) Logical equivalence of contrapositive (Item 11)	79.2% 100%	Not applicable	

Reliability coefficients

Table 10 also shows the internal consistencies (Cronbach's alpha) of the test items for logical reasoning and Proof Constructions. The internal consistencies of the pretest selection tasks for the logical reasoning were unusually lower ($\alpha = 0.12$) than 0.8. The items were highly similar in how the task and the prompts were phrased and posed but the students' card choices were not consistent and seemed random at times. 23 out of 60 students had chosen a card combination in at least one or two tasks which were different than their other responses.

The Cronbach's alpha of the Deductive-proof Construction items in pretest and posttest were at least 0.5. Pretest-posttest spearman correlation analysis of these two items yielded 0.47 and 0.17, indicating small to medium but positive correlation between students' pretest and posttest performance in Deductive-proof Constructions in general. The Cronbach's alpha of the Proof-by-counterexample Construction items in pretest and posttest were lower than 0.5. The source of low consistencies may be originated from the use of two different mathematics topics, elementary number theory and quadratics, in these two items. Pretest-posttest spearman correlation analysis of these two items yielded 0.28 and 0.19, indicating small but positive correlation between students' pretest and posttest performance in Proof-by-counterexample Constructions in general.

Principal Component Analysis

Considering the low internal consistencies and the similar task formulations between the two Deductive-proof Construction items, a principal component score was derived for each student's score for the Deductive-proof Construction (KMO: 0.5). Likewise, a principal

component score was also derived for each student's score for the proof-by counterexample construction (KMO: 0.5). Table 11 below shows the result of the principal component analyses for each type of Proof Construction. Each principal component accounted for at least 62% of the total variance and the factor loading was at least 0.79. Principal component scores of Deductive-proof and Proof-by-counterexample for each subject were generated from the original raw scores using the statistical software SAS 9.2. The adjusted scores are reported in the next chapter.

Table 11: Principal component analysis

	Factor loading	Explained Variance
Deductive-proof Construction items	0.83760	0.7016
Proof-by-counterexample Construction items	0.79306	0.6290

Biases in Pretest results

Preliminary analyses of pretest results were carried out for detecting any systematic bias due to the assignments by three conditions and two test sets prior to intervention. Two-way ANOVA were carried out on students' total raw scores of selection tasks, their principal component scores of Deductive-proof Construction and Proof-by-counterexample Construction and their ranked scores of each of the Proof Validation and Proof Knowledge items. The Type I error and the power of ANOVA would not be affected by the use of Rank Transform method on the ordinal scores of the items in the last group (Agresti, 2010; Akritas, 1990; Conover & Iman, 1981).

Overall, no pre-existing bias attributable to conditions or test sets are found in student pretest scores prior to training intervention (Conditions: $0.19 < F(2, 54) < 1.94$, $0.1531 < p <$

0.8287; Test sets: less than $0.001 < F(1, 54) < 3.68$, $0.0603 < p < 0.9458$; Interaction effect: $0.06 < F(2, 54) < 2.15$; $0.1266 < p < 0.9382$). Differences observed in students' mean pretest scores in each cell were not statistically significant across two test sets and three conditions. The assignments of conditions and test sets did not contribute any significant pre-existing bias to the students' pretest scores in all items.

Overview of the analyses plan

Recall that the research questions of this study compared the benefits of logic training emphasizing generation of counterexamples to the conventional approach (Control condition) in enhancing the following student's performance: (1) Reasoning with logical implications, (2) Proof Constructions, (3) Proof Validations, and (4) Knowledge of proof methods.

In addition, the study also investigated the following research questions: (5) To what extent does students' reasoning and various abilities of mathematical proving of logical implications correlate with each other? (6) How do students modify a falsifiable mathematical implication to a mathematically true implication based on their self-generated examples and counterexamples?

In the research questions 2 to 4, Proof Constructions consisted of Deductive-proof and Proof-by-counterexamples constructions, Proof Validations consisted of Invalidation of Empirical Proofs and Validation of Proof-by-contradiction, and Knowledge of Proof Methods concerned the Logical non-Equivalence of implication and its converse and Logical Equivalence of implication and its contrapositive. Analyses from the study are carried out in the order of addressing the above research questions. Improved performances in the test items are taken as evidence of enhanced students' reasoning and proving abilities.

To inform the effects of different training conditions on students' logical reasoning and various aspects of mathematical proving ability as spelt out in research questions (1) to (4), two-way repeated measures ANOVA (3 conditions by 2 assessments) on total scores of logical reasoning items, principal component scores of Proof Construction items and ranked scores of other proof-related items from students' pretest and posttest data were carried out. The choice of ANOVA was made depending on whether the scores were interval-based or ordinal-based. Additionally, students' written reflection data were also analyzed to understand the impact of each training condition on their logical reasoning. Table 12 below presented an overview of the various statistical analyses applied to the students' pretest and post test scores and the research question being addressed.

Table 12: Overview of the analyses of students' test responses

Focus of analysis	Outcome variables	Statistical analysis	Addressing
Logical Reasoning	Sum of scores of Items 1 to 4 (Pretest and posttest)	Two-way repeated measure ANOVA on students' total scores	Question (1)
Mathematical proving			
Deductive-proof Construction	Collation of Deductive-proof items scores (Pretest and posttest)	Principal Component Analysis; Two-way repeated measure ANOVA on Principal Component Scores	Question (2)
Proof-by-counterexample Construction	Collation of Proof-by-Counterexample items scores (Pretest and posttest)		
Proof Validation – Empirical Proof	Rank Transform of Item scores No. 9 (Pretest and posttest)	Two-way repeated measure ANOVA on ranked scores	Question (3)
Proof Validation – Proof-by-Contradiction	Rank Transform of Item scores No. 12 (Pretest and posttest)		
Knowledge of Proof Method – Converse	Rank Sum of Item scores No. 10 (Pretest and posttest)		Question (4)

Table 12 (cont'd)

Knowledge of Proof Method – Contrapositive	Rank Sum of Item scores No. 11 (Pretest and posttest)		
Correlation between students’ gain scores in logical reasoning and proving			
Logical Reasoning	Total gain score of Items 1 to 4 (Posttest- Pretest)	Spearman Correlation Matrix	Question (5)
Deductive Proof	Gain in Principal Component score (Posttest – Pretest)		
Proof-by-counterexample			
Validation of Empirical Proof	Gain in raw scores of Item 9, 12, 10, 11 (Posttest minus Pretest)		
Invalidation of Proof-by-Contradiction			
Logical non-equivalence of Converse			
Logical equivalence of Contrapositive			
Post study interview			
Modification of false implications and Proof Construction	Post study interview video and audio data	Qualitative analyses	Question (6)

A Spearman correlational matrix comprising of the seven gain scores (students' reasoning and various proof-related scores) were computed to investigate the association between students' logical reasoning and various aspects of mathematical proving. Students' gain scores in these seven components of the pretest and post test were calculated from their raw scores and then transformed into their ranked scores for computation of the correlational matrix.

A theoretical model of Proof Construction was also developed based on the coded interview data to account for the role of examples and counterexamples in Proof Construction. In the event that the treatments which emphasized generation of counterexamples did not benefit

students' mathematical proving as proposed, the analysis of the interview data might account for other possible factors.

CHAPTER 4 RESULTS

In this chapter, I will present the main findings pertaining to students' logical reasoning, Proof Construction, Proof Validation and Proof Knowledge, using data from students' pretest and posttest responses to address the first five research questions set out in Chapter 2 (the sixth question is addressed in the next chapter). The first four questions compared the training effects of the logic training approach emphasizing counterexamples (Condition W) or rule violations (Condition PO) with that of the conventional approach (Condition C, the Control condition) emphasizing logical truth tables on students' logical reasoning and mathematical proving abilities, in the following order: (1) Logical reasoning of implications, (2) Proof Construction consisting of Deductive-proof Construction and Proof-by-counterexample, (3) Proof Validation consisting of Invalidation of Empirical Proof and Validation of Proof-by-contradiction, and (4) Proof Knowledge consisting of logical non-equivalence of converse and logical equivalence of contrapositive. The fifth research question examined the correlation among students' logical reasoning and mathematical proving abilities.

In the next four sections, students' responses to the total of 12 pretest and posttest items are presented in the order corresponding to the research questions mentioned above. Each section will report the students' performance in the test items related to that ability in a two-way table (three conditions by two tests), the outcomes of a repeated-measure two way ANOVA on the students' scores, the relative frequency distribution of students' scores and exhibit samples of students' works that might be worthy of mention. A significance-level of 0.05 was used for all ANOVA. Partial η^2 is used as an effect size measure which qualified values of about 0.01 as weak effect, values of about 0.06 as medium effect, and values of about 0.14 or bigger as large effect (Cohen, 1988, pp. 285-287). Following that, the next section will address the fifth research

question by presenting the correlation analysis of students' gain scores related to the various abilities using a matrix.

According to our main hypothesis, students' scores should show significant differences between all three conditions and across pretest and posttest, with students' improved scores in the Conditions W and PO shown to be significantly higher than Condition C in a post-hoc multiple comparison test. Moreover, similar to the findings of Cheng, et al. (1986), I expected that students' logical reasoning scores in the Condition PO will be better than that of Condition W. According to the conjecture that the logic training emphasizing rule violations is the most effective for enhancing students' mathematical proving, students' proving-related scores in Condition PO should also show greater improvement (G. Stylianides & Stylianides, 2008).

As the readers will soon realize, the findings of a study seldom match what one sets out to find, which can be said of this study too. At the end of this chapter, I will summarize these findings and discuss a few issues related to the interpretation of results briefly. These discussions support the inquiry and report of interview findings in the next chapter. Jointly, the findings in these two chapters will lead to the conclusions of the study in the final chapter.

Effect of training on Students' logical reasoning

Recall that the research question (1) concerns the effect of logic training on students' reasoning of the implications. Table 13 below shows the descriptive statistics of students' total raw scores in the logical reasoning component, i.e., the Wason's selection tasks in items 1 to 4, before and after the training. Overall, the scores ranged from -16 to +16 with step of two-point intervals.

Table 13: Students' performance in logical reasoning by condition

<u>Condition</u>	<u>Pretest</u>		<u>Posttest</u>	
	<u>Mean (Min, Max)</u>	<u>SD</u>	<u>Mean (Min, Max)</u>	<u>SD</u>
<u>C (N=20)</u>	1.9 (0, 10)	3.21	2.0 (-8, 16)	6.12
<u>PO (N=20)</u>	0.7 (-2, 8)	2.08	5.6 (0, 16)	7.50
<u>W (N=20)</u>	0.3 (-2, 8)	2.77	12.6 (4, 16)	5.39

The mean score of the control group (Condition C) group increased slightly while the mean score of the treatment groups increased more (Condition PO and Condition W) after the logic training. On average, the students in the Condition C performed slightly better, while the students in the Conditions PO and W performed much better in their logical reasoning of implications. The standard deviation of the posttest scores also increased for all groups. Students' posttest scores varied wider than their pretest scores, with some scoring the highest possible points.

A repeated measure two-way ANOVA on students' scores was carried out to determine whether students in Conditions PO and W performed significantly better in logical reasoning. Analysis showed significant between-subject effects of condition ($F(2, 57) = 8.27, p = .0007, \eta^2 = 0.33$) and within-subject effects of repeated assessments ($F(1, 57) = 45.10, p < 0.0001, \eta^2 = 0.52$) with interaction effects ($F(2, 57) = 17.08, p < 0.0001$).

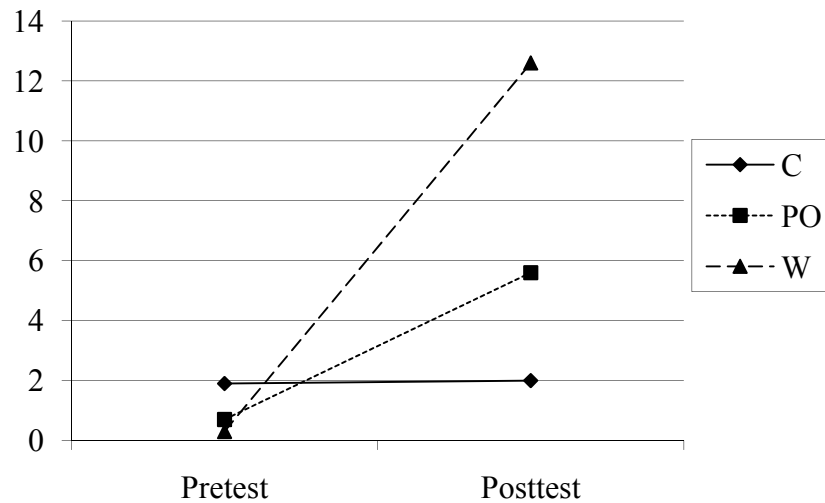


Figure 7: Pretest and Posttest mean scores by conditions

Figure 7 above indicated that the treatment which emphasized the generation of counterexamples (condition W) was significantly more effective than the approach of emphasizing logical truth table (condition C) and the treatment of emphasizing rule violations (condition PO). Scheffe's multiple comparison tests affirmed that the mean total score of students logical reasoning in condition W was significantly different from those of condition C and PO (minimum significant difference = 5.08, $\alpha = 0.05$, mean (W) = 12.6, mean (PO) = 5.6, mean (C) = 2.0). This explained the main effects due to condition and assessment, as well as their interaction effects mentioned earlier. While the logic training in Condition PO did not make significantly larger impact than the training in Condition C, the training in Condition W created a significant impact when compared to the other two. Logic training that emphasized generation of counterexamples benefited students' logical reasoning of implications significantly and better than the training which emphasized violations of rule and which involved logic truth tables.

Effects of training on Students' Proof Construction

Recall that the research question (2) concerns the effect of logic training on students' Proof Construction, which comprised Deductive-proof Construction and Proof-by-counterexamples Construction. Table 14 and Table 15 show the summaries of students' total raw scores and principal component scores for these two types of Proof Construction by condition. The raw scores of Deductive-proof and Proof-by-counterexample Construction were assigned based on the extent to which deductive inferences were productively used. The principal component score for each student's Deductive-proof and Proof-by-counterexample Construction was derived from the students' two-item raw scores related to each construction by means of orthogonal transformation. The derived principal component scores for each type of Proof Construction thus reflected students' performance, which were uncorrelated with other factors, in their use of deductive inferences.

Table 14: Students' total raw scores in Proof Construction by condition

<u>Condition</u>	<u>Pretest</u> <u>Mean (min, max)</u>	<u>SD</u>	<u>Posttest</u> <u>Mean (min, max)</u>	<u>SD</u>
<u>C (N=20)</u>				
Deductive-proof	3.55 (1, 10)	2.438	4.25 (1, 10)	2.954
Proof-by-counterexample	5.05 (0, 10)	2.946	5.70 (1, 10)	2.227
<u>PO (N=20)</u>				
Deductive-proof	3.75 (0, 8)	2.124	4.05 (1, 8)	2.089
Proof-by-counterexample	5.45 (2, 10)	2.305	6.00 (2, 9)	1.919
<u>W (N=20)</u>				
Deductive-proof	3.1 (0, 5)	1.586	4.15 (1, 9)	2.277
Proof-by-counterexample	4.25 (0, 8)	2.314	5.50 (1, 9)	2.282

Overall, students' total raw scores ranged from 0 to 10 for Deductive-proof Construction and Proof-by-counterexample Construction in both pretest and posttest. Students' principal

component scores for Deductive-proof Construction ranged from -1.67 to 2.75 in pretest and -1.24 to 2.75 in posttest. Students' principal component scores for Proof-by-counterexample ranged from -2.41 to 1.89 in the pretest and -1.92 to 1.89 in the posttest. The minimum and maximum score by each condition and assessment is provided in Table 14 and Table 15.

Table 15: Students' principal component scores in Proof Construction by condition

<u>Condition</u>	<u>Pretest</u>		<u>Posttest</u>	
	<u>Mean (min, max)</u>	<u>SD</u>	<u>Mean (min, max)</u>	<u>SD</u>
<u>C (N=20)</u>				
Deductive-proof	-0.11 (-1.23, 2.75)	1.08	0.20 (-1.21, 2.75)	1.30
Proof-by-counterexample	-0.12 (-2.41, 1.89)	1.25	0.18 (-1.93, 1.89)	0.91
<u>PO (N=20)</u>				
Deductive-proof	-0.02 (-1.67, 1.89)	0.94	0.10 (-1.24, 1.84)	0.91
Proof-by-counterexample	0.05 (-1.54, 1.89)	0.95	0.30 (-1.55, 1.50)	0.79
<u>W (N=20)</u>				
Deductive-proof	-0.32 (-1.67, 0.52)	0.69	0.15 (-1.24, 2.32)	1.01
Proof-by-counterexample	-0.46 (-2.41, 1.12)	1.03	0.06 (-1.92, 1.50)	0.96

For both Deductive-proof and Proof-by counterexample constructions, the mean total scores and the mean principal component scores increased across all conditions after the logic training. Hence students exhibited more productive use of deductive inferences in constructing proofs for proving or falsifying mathematical implications. However, closer look at the students' scores were required to conclude whether the improvement was systematic and whether they were better in constructing mathematically valid proofs.

Deductive-proof Construction

A two-way repeated measures ANOVA (3 conditions by two tests) of students' principal component scores for Deductive-proof Construction was carried out to determine whether students in Conditions PO and W performed significantly better. Analysis showed no significant

between-subject effects of conditions ($F(2, 57) = 0.15, p = 0.8601, \text{partial } \eta^2 = 0.0053$) but significant within-subject effects of repeated assessments ($F(1, 57) = 4.67, p = .0349, \text{partial } \eta^2 = 0.076$). Interaction between condition and assessment were not significant ($F(2, 57) = 0.49, p = .6157, \text{partial } \eta^2 = 0.017$). Training had a medium positive effect on students' construction of deductive proofs in all three Conditions. Students showed improved use of deductive inferences in constructing deductive proof after the respective logic training but their improvements were not significantly better or worse than each other across the Conditions.

Deductive-proof Construction of Item related to Elementary Number Theory

Figure 8 showed the distribution of students' raw scores using a stacked bar graph. The item that students attempted posed an implication related to elementary number theory. Recall that the coding scheme classified irrelevant attempts as level 0, empirical or logically erroneous proofs as levels 1 and 2, incomplete deductive proofs as levels 3 and 4, and coherent deductive proofs as levels 5 and 6. Since each condition had 20 students per assessment, 5% of each column would represent 1 student.

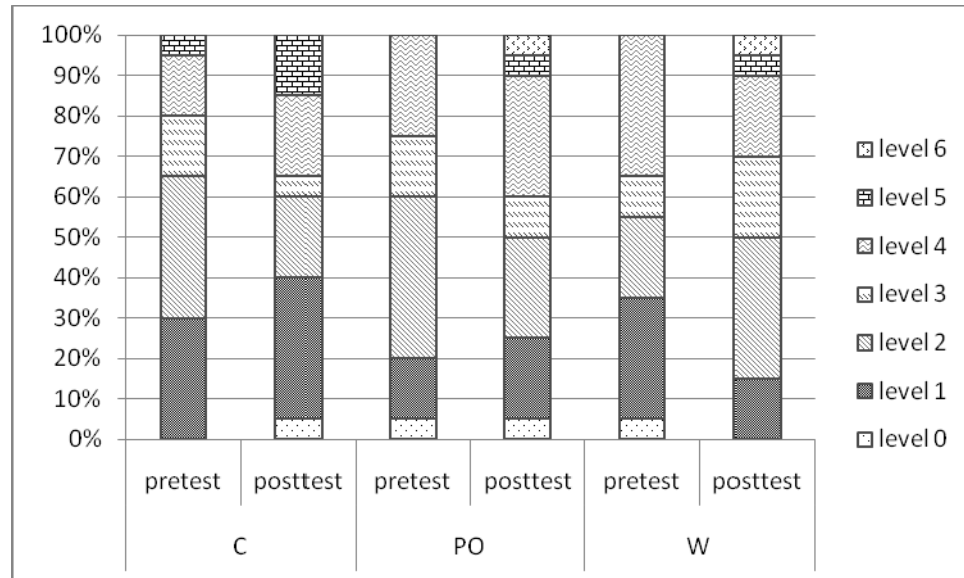


Figure 8: Students' Deductive-proof Construction (Elementary number theory)

Referring to the bar legend, the bar columns representing levels 5 and 6 increased their proportions in all three Conditions after the logic training, indicating more coherent proofs (level 5 and 6) were constructed across all conditions. The total number of students scoring levels 5 and 6 increased from 1 (5% in pretest of Condition C) to 7 (15% + 10% + 10% in posttest of each condition). The bar columns representing levels 3 to 6 also increased from 24 (35% + 40% + 45% in pretest of each condition) to 28 (40% + 50% + 50% in pretest of each condition).

This figure showed students' proofs shifted towards higher level of Deductive-proof Construction in this number theory task with moderately more students used deductive-inferences instead of examples and logically erroneous inferences. Note that least 50% of the students across all conditions had yet to make productive use of deductive inferences in constructing proofs (level 3 and above), which concurred with other similar studies (Healy & Hoyles, 2000; Hoyles & Küchemann, 2003; Knuth, et al., 2009).

Deductive-proof Construction of Item related to Quadratics

Figure 9 below is a stacked bar graph showing the distribution of students' raw scores for the other item that posed an implication related to Quadratics. Recall that the implication posed in the pretest item was different from its counterpart in the posttest due to the isomorphic design of the assessment instrument.

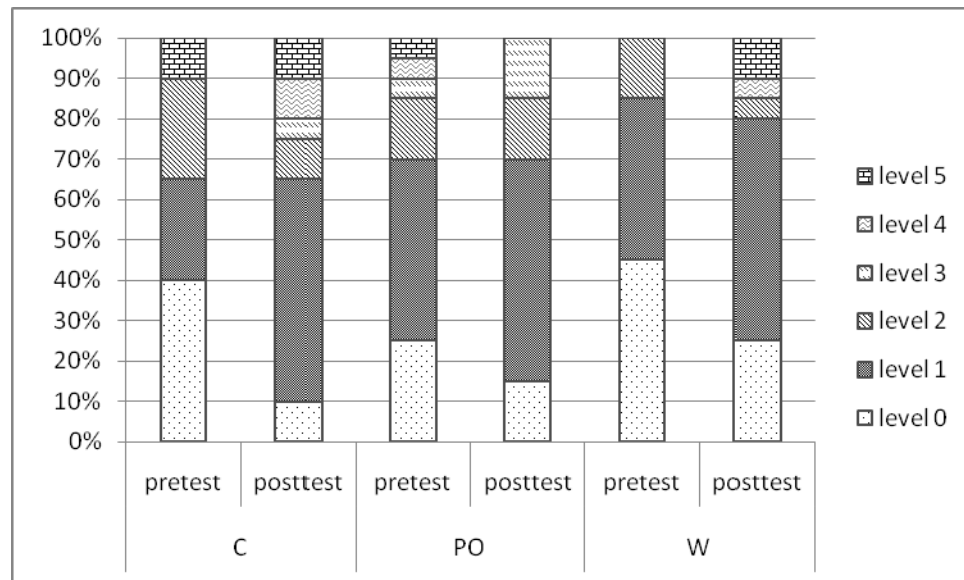


Figure 9: Students' Deductive-proof Construction (Quadratics)

Referring to the bar legend, the bar columns representing levels 0 to 2 summed up to be at least 85% during pretest and at least 75% during posttest in all conditions. The bar columns representing level 0 ranged between 25% to 45% during pretest and decreased to between 10% to 25% after training with a difference as large as 30% in Condition C. The bar columns representing levels 3 and above increased from 10% to 25% in the Condition C and 0% to 20% in Condition W after training. Surprisingly, the columns remained the same at 15 % in Condition PO with the bar columns representing levels 4 and 5 decreased from 10% to 0% after training.

The bar columns representing levels 4 and 5 were 10% in the other two Conditions during posttest.

Majority of the students might find this item challenging. Despite the training received, at least 75% of the students in all conditions still relied on examples, with some general inferences generated at best, but had yet to make any productive use of deductive inferences to prove the implication. Students in Conditions C and W exhibited some progress in using deductive inferences. Only four students (10% of Condition C + 10 % of Condition W) were able to construct coherent proof after training. Overall, students' improved use of deductive inferences was moderate in some conditions or little in other condition. The prevalent use of Empirical Proofs concurred with similar studies about students' proof behavior when the task becomes more challenging (Coe & Ruthven, 1994; Healy & Hoyles, 2000; Knuth, et al., 2009; Recio & Godino, 2001).

Students' Indirect approach in Deductive-proof Construction

Some students in Condition W or PO adopted an indirect approach of Proof Construction that resembled the approach provided in their training materials. Instead of inferring how the antecedent leads to the consequent, students considered whether a counterexample to the implication could possibly be constructed. Figure 10 below shows such an instance of Alex's (pseudonym) work in deciding the truth of the implication "[For graphs of $y=ax^2+bx+c$] If a is positive and c is negative, then the x -intercepts of the graphs are one positive number and one negative number."

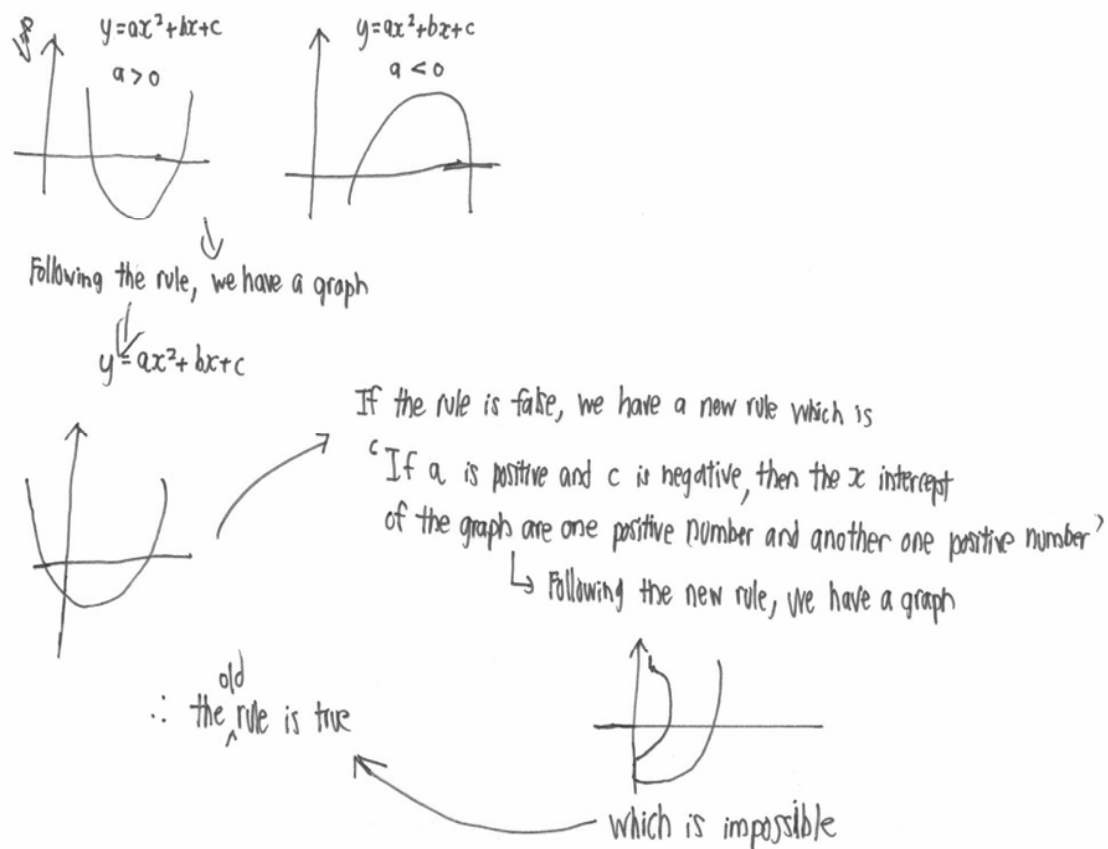


Figure 10: Sample of Indirect Proof approach in Deductive-proof Construction (This figure was hand drawn by: Alex)

The student considered “If a is positive and c is negative, then the x -intercepts of the graphs are one positive number and one positive number,” a “rule” which he inferred from considering the original implication as “false.” In accordance with this new “rule,” he drew a graph which ended up as a non-quadratic graph. Thus the student concluded that it was impossible to produce any such instances. His proof was scored at level 3 because the other case of x -intercepts being both negative, which is also another possible inference if the original implication was false, were not ruled out. The number of students adopting indirect approach was, however, rare and limited to the item involving Quadratics.

Proof-by-counterexample Construction

Next we turn to students' performance in Proof-by-counterexample Construction. A two-way repeated measures ANOVA (3 conditions by two assessments) of students' principal component scores for Proof-by-counterexample Construction was carried out to determine whether students in Conditions PO and W performed significantly better. Analysis showed no significant between-subject effects of conditions ($F(2, 57) = 1.52, p = 0.2275, \text{partial } \eta^2 = 0.051$) and within-subject effects of repeated assessments ($F(1, 57) = 3.71, p = .0590, \text{partial } \eta^2 = 0.061$). Interaction between condition and assessment were also not significant ($F(2, 57) = 0.21, p = .8077, \text{partial } \eta^2 = 0.0075$). No training condition enhanced students' construction of Proof-by-counterexample significantly, though the difference between students' pretest and posttest performance approached significance.

Proof-by-counterexample Construction of Item related to Elementary Number Theory

Figure 11 below is a stacked bar graph showing the distribution of students' raw scores for the item that posed an implication related to Elementary Number Theory. Recall that the coding scheme classified irrelevant attempts as level 0, empirical or logically erroneous proofs as levels 1 and 2, deductive proofs with misconceptions as levels 3, and successful falsification by counterexamples as levels 4 and 5. Likewise, 5% of each column would represent 1 student.

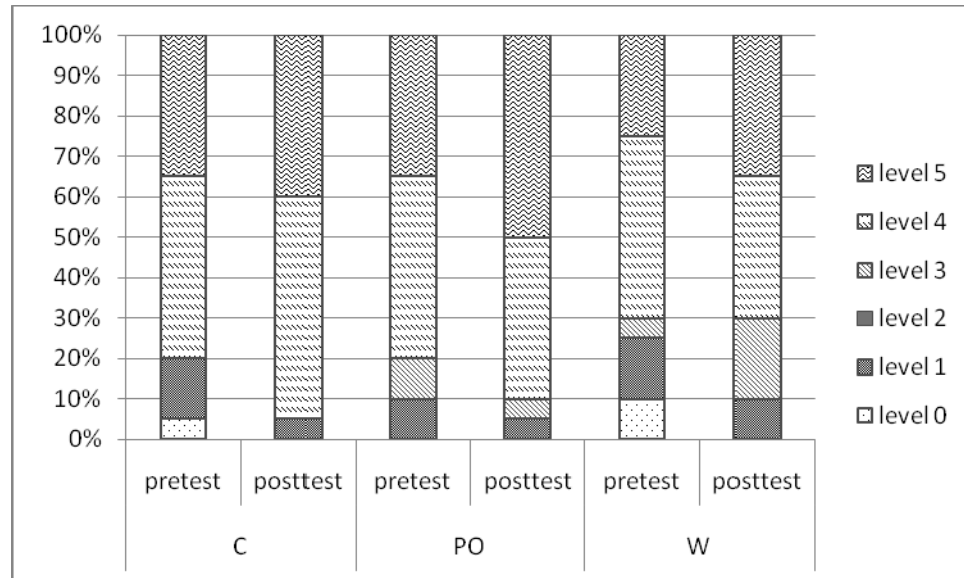


Figure 11: Students' Proof-by-counterexample Construction (Elementary Number Theory)

Referring to the bar legend, the bar columns representing levels 4 and 5 in all conditions were at least 70% during pretest. While that proportion remained constant in Condition W, its counterparts increased from 80% to 95% in Condition C and from 80% to 90% in Condition PO.

Majority of the students (at least 70%) across all conditions might find this item easy. They might be well-versed with the implication and thus were capable of constructing specific or a set of counterexamples (levels 4 and 5) before training. After training, more students were able to identify a set of counterexamples (level 5) across all conditions in the isomorphic posttest item. The high success rate of constructing counterexamples in both the pretest and posttest indicated a possible ceiling effect.

Proof-by-counterexample Construction of Item related to Quadratics

Figure 12 below is a stacked bar graph showing the distribution of students' raw score for the other item which posed an implication related to Quadratics. Likewise, the implication posed

in the pretest item was different from its counterpart in the posttest due to the isomorphic design of the assessment instrument.

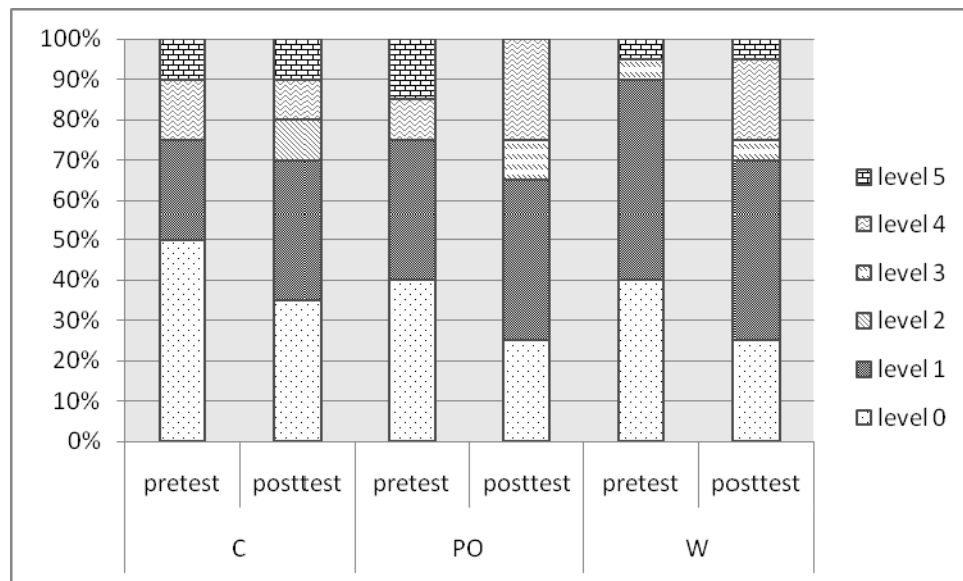


Figure 12: Students' Proof-by-counterexample Construction (Quadratics)

Referring to the bar legend, the bar columns representing levels 0 to 2 in all conditions were at least 75% during pretest. The bar columns representing level 0 in all conditions ranged from 40% to 50% while the bar columns representing level 1 in all conditions ranged between 25% to 50%. After training, the bars representing levels 0 and 1 in all conditions still remained around 65% and 70% during post test. The bar columns representing levels 4 and 5 decreased from 25% to 20% in Condition C but increased from 5% to 25% in Condition W. The bar column representing level 5 in Condition PO also decreased from 15% to 0%.

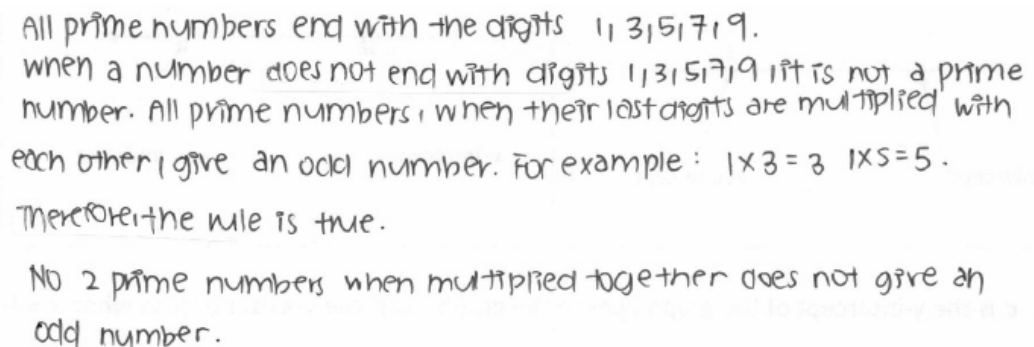
Majority of the students across all conditions might find this item challenging. Overall, 25% or less of the students was able to construct the counterexamples during the pretest. They were only able to construct an example or make logically erroneous inferences to conclude the implication incorrectly as true. After training, at least 65% of the students still responded

similarly to a different but isomorphic item. The combined effects of the task demand and the content might be too great for most students and constituted possibly a floor effect, which might have undermined the overall training effects.

Furthermore, the findings of students' performance in this item contrasted sharply with that of the above item. A ceiling effect was observed for the implication involving elementary number theory while a floor effect was observed in the implication involving Quadratic. Both effects were likely to render the test items insensitive to any between-subject effects. This extreme contrast between the two items might have also contributed to the low reliability coefficient of these two items, as reported in the previous chapter.

Students' False Deductive Proofs with Inadequate Mathematical Considerations

Some students proved the implication using a chain of deductive inferences falsely because of inadequate consideration of all mathematical objects. There seemed to be a brief lapse of logical reasoning in their Proof Construction. Figure 13 below shows such an instance of Brady's (pseudonym) posttest work in deciding whether the implication "*If two prime numbers are multiplied together, then the product is an odd number*" was true.



All prime numbers end with the digits 1, 3, 5, 7, 9.
When a number does not end with digits 1, 3, 5, 7, 9 it is not a prime number. All prime numbers, when their last digits are multiplied with each other, give an odd number. For example: $1 \times 3 = 3$ $1 \times 5 = 5$.
Therefore the rule is true.
No 2 prime numbers when multiplied together does not give an odd number.

**Figure 13: Students' Proof-by-counterexample Construction with inadequate consideration
(This figure was hand drawn by: Brady)**

The student started considering the set of prime numbers based on the end-digits of the numbers and asserted that all prime numbers were odd numbers with end-digits 1, 3, 5, 7 and 9. Then she deduced that when two prime numbers with odd end-digits were multiplied together, the resulting end-digit was also an odd number. From here she concluded that the implication was true. The proof was scored at level 3 due to a lapse of logical reasoning in asserting her end-digit model of prime numbers. However, her subsequent inferences and conclusion were logically valid.

In summary of the training effects on students' Proof Construction, the study found that the logic training emphasizing counterexamples or rule violation benefited students' Deductive-proof Construction equally as the logic training emphasizing truth tables. Students used deductive inferences more productively in constructing deductive proofs. However, all three logic training approaches did not benefit students' Proof-by-counterexample Construction. Existential evidence of students' use of indirect proof approach and inadequate considerations of mathematical objects of a false implication were also found.

Effects of training on Students' Proof Validation

Recall that the research question (3) concerns the effect of logic training on students' Proof Validation, which comprised Invalidation of Empirical Proof and Validation of Proof-by-contradiction. Table 16 shows the summaries of students' raw scores for these two types of Proof Validations in each condition. Each type of Proof Validation was assessed with a test item. The raw scores of each test item were assigned based on the extent to which students considered

possible counterexamples when determining the validity of a given proof. These scores ranged from 0 to 2 for Invalidation of Empirical Proof and from 0 to 3 for Proof-by-contradiction.

Table 16: Students' performance in Proof Validation by condition

<u>Condition</u>	<u>Pretest</u>		<u>Posttest</u>	
	<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
<u>C (N=20)</u>				
Empirical Proof	0.80	0.523	0.80	0.767
Proof-by-Contradiction	0.30	0.732	0.50	0.760
<u>PO (N=20)</u>				
Empirical Proof	0.65	0.489	0.85	0.670
Proof-by-Contradiction	0.15	0.489	0.40	0.598
<u>W (N=20)</u>				
Empirical Proof	0.55	0.510	0.95	0.826
Proof-by-Contradiction	0.15	0.489	0.90	1.252

The mean scores of the students' Invalidation of Empirical Proof varied differently across all conditions after the logic training. They remained the same in Condition C but showed increase in the other two Conditions. In comparison, the mean scores of the students' Validation of Proof-by-contradiction increased across all conditions.

Rank Transform method and a two-way repeated measure ANOVA were each applied to students' raw scores in their Invalidation of Empirical Proofs (Item 9) and Validation of Proof-by-contradiction (Item 12) (Agresti, 2010; Akritas, 1990; Conover & Iman, 1981). As the coding schemes were hierarchical, the raw scores were ordinal measures in nature. Since the levels of the scores were too few, the ordinal scores were converted into rank sum scores before applying the ANOVA.

Invalidation of Empirical Proof

A two-way repeated measures ANOVA (3 conditions by two tests) of the students' ranked scores for Invalidation of Empirical Proof was carried out to determine whether students in Conditions PO and W performed significantly better. Analysis showed no significant between-subject effects of conditions ($F(2, 57) = 0.08, p = 0.9229, \text{partial } \eta^2 = 0.0028$) and within-subject effects of repeated assessments ($F(1, 57) = 2.25, p = .1388, \text{partial } \eta^2 = 0.038$). Interaction between condition and assessment were also not significant ($F(2, 57) = 1.13, p = .3299, \text{partial } \eta^2 = 0.038$). No training condition enhanced students' Invalidation of Empirical Proof significantly across conditions and assessments.

Figure 14 below is a stacked bar graph showing the distribution of students' raw score for the item of Invalidation of Empirical Proof. Recall that the item concluded an implication was true based on three examples. The coding scheme classified students' attempts which did not considered possible counterexamples as level 0, attempts which questioned the truth of the implication as level 1, and attempts which recognized that the given proof was invalid due to a fundamental lack of logical considerations for possible counterexamples as level 2. The implication posed in the pretest item was different from its counterpart in the posttest but the structures of the proof were isomorphic.

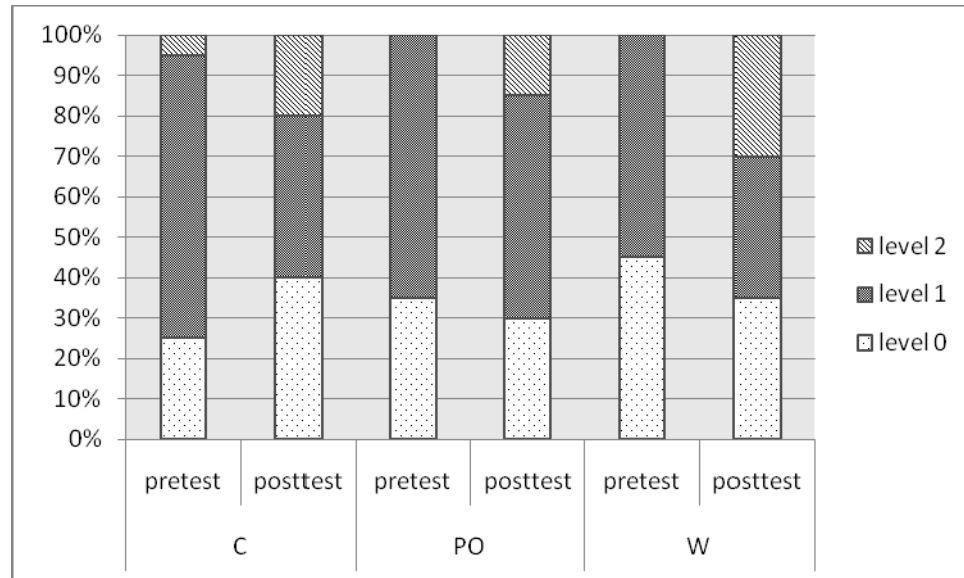


Figure 14: Students' Invalidation of Empirical Proof

Referring to the bar legend, the bar columns representing level 1 made up the main bulk of the distribution. Across all conditions, these bar columns decreased after the training. The bar columns representing level 2 increased from 5% to 20% in Condition C, 0% to 15% in Condition PO and 0% to 30% in Condition W. The bar columns representing level 0 were between 25% and 45% across all conditions during pretest and remained about the same (30% to 40%). In addition, only the bar columns representing level 0 in Condition C increased after training, while its counterparts decreased in Conditions PO and W.

Majority of the students across all conditions questioned the truth of the implication of the proof but not the validity of the proof itself. Some students understood the logical inadequacy to consider possible counterexamples as the invalidity of the given proof. However, a substantial proportion of students' responses did not reject the proof or reject it without relevant reasons during pretest and after training. Such proportion even increased after training in Condition C.

Overall, logic training in all three conditions benefited some students in understanding why Empirical Proofs were invalid based on logical ground. However, such benefits was limited

and mixed across Conditions, which might have explained why the training effects bore no significant improvements.

Validation of Proof-by-contradiction

A two-way repeated measures ANOVA (3 conditions by two tests) of the students' ranked scores for validation of Proof-by-contradiction was carried out to determine whether students in Conditions PO and W performed significantly better. Analysis showed no significant between-subject effects of conditions ($F(2, 57) = 0.28, p = 0.7604, \text{partial } \eta^2 = 0.0096$) but significant within-subject effects of repeated assessments ($F(1, 57) = 8.34, p = .0055, \text{partial } \eta^2 = 0.128$). Interaction between condition and assessment were not significant ($F(2, 57) = 0.46, p = .6325, \text{partial } \eta^2 = 0.016$). Training had medium positive effect on students' Validation of Proof-by-contradiction in all three Conditions. Students showed improved consideration about counterexamples in validating Proof-by-contradiction after logic training. However, their improvements were not significantly better or worse than each other across the Conditions.

Figure 15 below is a stacked bar graph showing the distribution of students' raw scores for the item of Proof-by-contradiction. Recall that the item presented a Proof-by-contradiction to prove an implication. The coding scheme classified attempts which failed to accept the proof-by-contradiction approach as level 0, attempts which accepted the proof by agreement with the contradiction statement as level 1, attempts which asserted that the contradiction falsified the negation of the implication as level 2, and attempts which asserted that the contradiction eliminated all possible counterexamples logically as level 3. The implication posed in the pretest

item was different from its counterpart in the posttest though both concerned Elementary Number Theory and the structures of the proof were isomorphic.

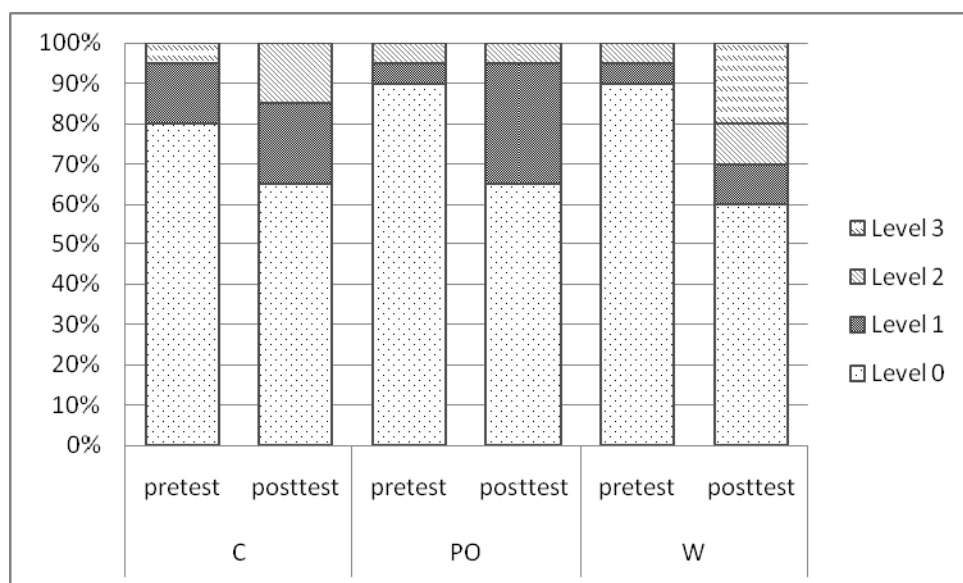


Figure 15: Students' Validation of Proof-by-contradiction

Referring to the bar legend, the bar columns representing level 0 were at least 80% across all condition during pretest. After training, the bar columns decreased by 15 % to 30% across all conditions but were still at least 60%. The bar columns representing level 1 increased across all condition with a moderate 5% increase in Conditions C and W but a substantial 25% increase in Condition PO. Bar columns representing level 2 increased moderately from 5% to 10% in Condition W and more substantially from 0% to 15% in Condition C. However, the latter increase in Condition C was compensated by a 5% decrease of the bar column representing level 3. Note that only the bar column representing level 3 increased substantially from 0% to 20% in Condition W.

Majority of the students across all conditions might find this item challenging. After training, 60 % or less of the students still failed to accept the proof and might constitute a floor

effect. They thought that the negation of the implication was “incorrectly” assumed or the contradiction did not prove the implication, both of which were reported cognitive difficulties (Antonini & Mariotti, 2008). Students’ improved performance in validating the Proof-by-contradiction seemed to be composed of mixed characters across conditions. In Condition W, more students recognized the validity because of the elimination of counterexamples due to the mathematical contradiction. In Condition C, more students recognized the validity by the falsification of the negation. In Condition PO, more students recognized the validity by evaluating and agreeing with the contradiction statement. Overall, students improved performance seemed to be more substantially due to their evaluation and agreement with the contradiction statement.

Students’ Understanding of Proof-by-Contradiction by Counterexample Elimination

Some students were able to accept a Proof-by-contradiction because the proof justified the implication through eliminating possible counterexamples. They seemed to have little cognitive issues in understanding why the proof worked. Figure 16 below shows an instance of students’ work in determining whether the Proof-by-Contradiction provided by a fictitious character named Gabriel was valid. In the task, the implication “*Let x and n be two real*

numbers. If $x > 0$ and $n > 0$, then $\frac{x}{n} + \frac{n}{x} \geq 2$ ” was used and Gabriel began the proof by

assuming a counterexample of a pair of positive numbers, a and b , exists, i.e., $\frac{a}{b} + \frac{b}{a} < 2$. The

algebraic operations of this inequality later led to a contradiction of the supposition and thus proved the implication (see Appendix A and Appendix B for the full details of the task). Students were asked if they agreed with Gabriel’s way of making his conclusion and why.

: 'Gabriel is trying to find an instance that makes the rule false
if the instance can be solved but his result shows that the values of
a and b cannot be solved. Therefore, he agrees that the rule is true.
Thus, I agree with gabriel's statement as his supposition is false
and he cannot find any instances to prove that the rule is false.
The statement is thus, true.

Figure 16: Student's justification for the Validity of Proof-by-contradiction (This figure was hand drawn by: Carl)

Carl (pseudonym) regarded the proof as an attempt to find a counterexample if the instance can be solved. She further understood the contradiction as a result of no solution for such an instance. She supported Gabriel's assertions on two bases: (1) his supposition was falsified by the contradiction and as a result, (2) no instance could be found to prove that the rule is false. Hence she made the logical conclusion that the implication is true. Her responses showed that she had associated the purpose of Proof-by-contradiction with the attempt of finding counterexample to the implication and contradiction indicated that the counterexample could not be "solved" or exist.

In summary of the training effects on students' Proof Validation, the study found that the logic training emphasizing counterexamples or rule violation benefited students' Validation of Proof-by-contradiction equally as the logic training emphasizing truth tables. Students accepted the Proof-by-contradiction with varying understanding about the purpose of reasoning to a contradiction. All three logic training approaches did not benefit students' Invalidation of Empirical Proof, though a number of students rejected Empirical proof due to its logical deficiency to address possible counterexamples. Existential evidence of students' ability to

understand Proof-by-contradiction as a logical failure to find a counterexample was also found. Next we looked at the training effects on students' Proof Knowledge.

Effects of training on Students' Proof Knowledge

Recall that the research question (4) concerns the effect of logic training on students' Proof Knowledge, which comprised knowledge about the logical non-equivalence between a mathematical implication and its converse as well as the logical equivalence to its contrapositive. Table 17 shows the summaries of students' raw scores for these two types of Proof Knowledge in each condition. Each type of Proof Knowledge was assessed with a test item. The raw scores of each test item were assigned based on the extent to which students applied deductive inferences when considering the logical equivalence or non-equivalence of statements to an implication. These scores ranged from 0 to 3 for students' consideration of converse and from 0 to 2 for students' consideration of contrapositive.

Table 17: Students' Consideration of Logical Equivalence

<u>Condition</u>	<u>Pretest</u>		<u>Posttest</u>	
	<u>Mean</u>	<u>SD</u>	<u>Mean</u>	<u>SD</u>
<u>C (N=20)</u>				
Converse	0.35	0.670	0.60	0.753
Contrapositive	0.75	0.444	0.90	0.307
<u>PO (N=20)</u>				
Converse	0.40	0.680	0.80	1.056
Contrapositive	0.90	0.308	0.95	0.223
<u>W (N=20)</u>				
Converse	0.50	0.827	0.60	0.883
Contrapositive	0.65	0.489	0.95	0.394

The mean scores of the students' consideration of the logical non-equivalence between an implication and its converse and the logical equivalence to its contrapositive increased across all

conditions after training. Similar to the analysis of students' scores for Proof Validation, the ordinal scores in their consideration of the logical non-equivalence of converse (Item 10) and the logical equivalence of contrapositive (Item 11) were converted into rank sum scores before the ANOVA was carried out.

Logical non-equivalence between an implication and its converse

A two-way repeated measures ANOVA (3 conditions by two tests) of the students' ranked scores for Logical non-equivalence of converse was carried out to determine whether students in Conditions PO and W performed significantly better. Analysis showed no significant between-subject effects of conditions ($F(2, 57) = 0.06$, $p = 0.9404$, partial $\eta^2 = 0.0022$) but significant within-subject effects of repeated assessments ($F(1, 57) = 4.04$, $p = .0492$, partial $\eta^2 = 0.066$). Interaction between condition and assessment were not significant ($F(2, 57) = 0.36$, $p = .7012$, partial $\eta^2 = 0.012$). Training had a medium positive effect on students' Proof Knowledge about converse in all three Conditions. Students showed improved application of deductive inferences when considering the logical non-equivalence between an implication and its converse. However, their improvements were not significantly better or worse than each other across the Conditions.

Figure 17 below is a stacked bar graph showing the distribution of students' raw scores for the item of logical non-equivalence of converse. Recall that the item presented a mathematical implication and its converse for students to determine whether both implications were logically equivalent. The coding scheme classified responses which considered both implications equivalent as level 0, responses which focused on the falsification of converse or

the comparison of antecedents as level 1, responses which constructed deductive proofs as level 2, and responses which compared the equivalence of counterexamples to both implications as level 3. Note that students' responses at level 2 or 3 indicated that an adequate knowledge of the logical non-equivalence between an implication and its converse. The implication posed in the pretest item was different from its counterpart in the posttest but the same mathematical content and sentential structure were used.

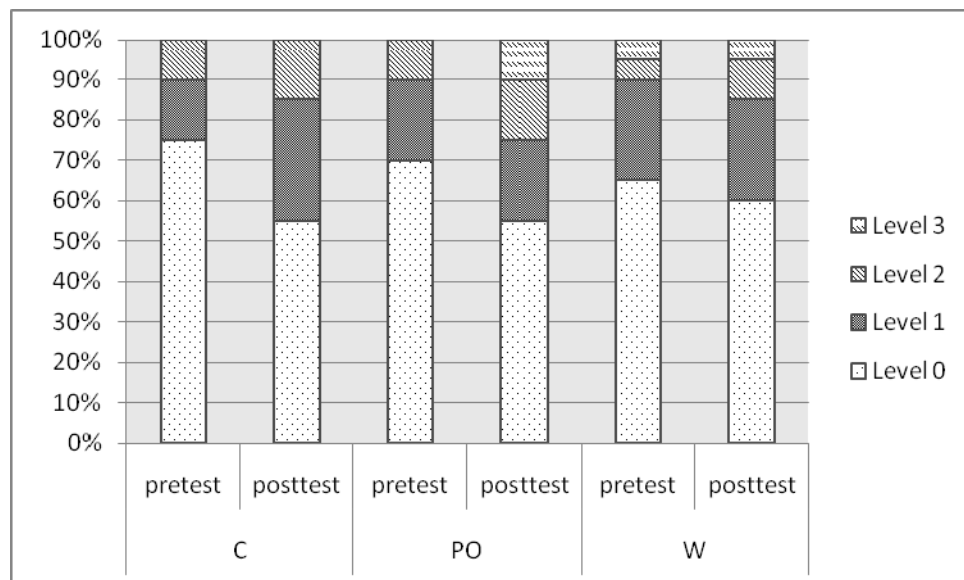


Figure 17: Students' Consideration of Logical non-equivalence of Converse

Referring to the bar legend, the bar columns representing level 0 were at least 65% across all conditions during pretest. After training, the bar columns still remained at least 55%. The bar columns representing level 1 only increased from 15% to 30% in Condition C, while the bar column representing level 2 increased 5% across all Conditions. The bar columns representing level 3 were found in Conditions PO and W but only the bar column increased from 0% to 10 % in Condition PO.

Despite the logic training received, more than half of the students still considered the converse being logically equivalent to the implication, which concurred with the well-documented phenomenon of students' mathematical proving (Healy & Hoyles, 2000; Hoyles & Küchemann, 2003; Knuth, et al., 2009). The high proportions also indicated that students might find the item challenging. Students' improved performance beyond this level showed a mixed character in terms of their application of deductive inferences across conditions. At the very least, the benefits of logic training to students' knowledge about the logical non-equivalence of converse were significant across all conditions to the extent that more students applied deductive inferences to consider logical equivalence instead of irrelevant features of the implications.

Logical equivalence between implication and its contrapositive

A two-way repeated measures ANOVA (3 conditions by two tests) of the students' ranked scores for logical equivalence of contrapositive was carried out to determine whether students in Conditions PO and W performed significantly better. Analysis showed no significant between-subject effects of conditions ($F(2, 57) = 1.34, p = 0.2689, \text{partial } \eta^2 = 0.045$) but significant within-subject effects of repeated assessments ($F(1, 57) = 5.94, p = .0179, \text{partial } \eta^2 = 0.094$). Interaction between condition and assessment were not significant ($F(2, 57) = 1.09, p = .3428, \text{partial } \eta^2 = 0.037$). Training had a medium positive effect in students' Proof Knowledge about contrapositive in all three conditions. Students showed improved application of deductive inferences when considering the logical equivalence between an implication and its contrapositive. However, their improvements were not significantly better or worse than each other across the Conditions.

Figure 18 below is a stacked bar graph showing the distribution of students' raw scores for the item of logical equivalence of contrapositive. Recall that the item presented a mathematical implication and its contrapositive for students to determine whether both implications were logically equivalent. The coding scheme classified responses which rejected the logical equivalence as level 0, responses which accepted the logical equivalence by comparing the truth of both implications as level 1, and responses which accepted the logical equivalence by comparing the counterexamples to both implications as level 2. The implication posed in the pretest item was different from its counterpart in the posttest but the same mathematical content and sentential structure were used.

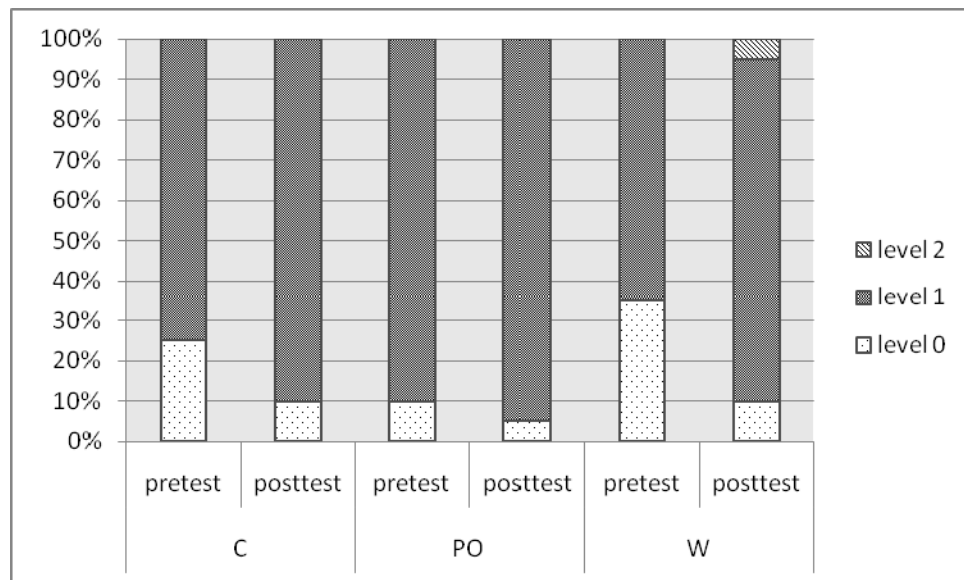


Figure 18: Students' Consideration of Logical equivalence of Contrapositive

Referring to the bar legend, the bar columns representing level 1 constituted the main bulk of the distribution. Across all conditions, the bar columns were at least 65% during pretest and increased to 90% or 95% after training. In complement, the bar columns representing level 0

in all conditions decreased to 10 % or less after training with a magnitude of 5% to 25%. The only bar column representing level 2 was 5% in the posttest of Condition W.

Majority of the students' still considered logical equivalence between an implication and its converse based on identical truth-values, despite the logic training received. However, fewer students rejected the logical equivalence based on mathematically irrelevant or incorrect reasons after training. Students who would consider the logical equivalence by comparing counterexamples of both implications after training were rare.

In summary of the training effects on students' Proof Knowledge, the study found that logic training emphasizing counterexamples or rule violation benefited students' consideration of logical equivalence involving converse and contrapositive significantly and equally as the logic training emphasizing truth tables. More students applied deductive inferences when comparing an implication and its converse or contrapositive. However, the extent of the improvement seemed to be only effective in shifting students' consideration away from irrelevant features or mathematically false reasoning.

Correlation between Logical Reasoning and Various aspects of Mathematical Proving

Recall that research question (5) inquired about the correlation between students' logical reasoning and various aspects of mathematical proving. Table 18 below shows the matrix of Spearman rank correlation coefficients comprising students' ranked gain scores of the logical reasoning and various components of mathematical proving. In each cell below, the number at the top is the correlation coefficient between the two components associated to the cell while the number in italics is the p-value of the coefficient being non-zero.

Table 18: Spearman correlational matrix of students' gain scores

Spearman Correlation Coefficients, <i>p</i> -value (N = 60)							
Ranked Gain Scores	Logical reasoning	Deductive-proof	Proof-by-counter-example	Invalidation of Empirical Proof	Validation of Proof-by-Contradiction	Converse	Contra-positive
Logical reasoning	1	.07122 .5887	.0166 .9	.1857 .1554	.0706 .5918	.0002 .9988	.00556 .9664
Deductive-Proof		1	-.21539 .0984	.25422 .05*	.09235 .4828	-.11595 .3777	.20671 .113
Proof-by-counter-example			1	-.03494 .791	.27373 .0343*	.15534 .236	-.1097 .404
Invalidation of Empirical Proof				1	.09007 .4937	-.16274 .2141	-.0567 .6667
Validation of Proof-by-Contradiction					1	-.1661 .2047	.09751 .4586
Converse						1	-.1242 .3444
Contra-positive							1

**p-value is significant at 0.05 level*

Among all pairs of correlation between the seven components, only two coefficients were significant. Both the correlation between Deductive-proof Construction and Invalidation of Empirical Proof (Spearman's $\rho = 0.254$, $p = 0.05$), and between Proof-by-counterexample Construction and Validation of Proof-by-contradiction (Spearman's $\rho = 0.274$, $p = 0.0343$) showed medium positive correlation. The correlation between students' gain in logical reasoning and the various aspects of mathematical proving were small and not significant (Spearman's $\rho < 0.19$ and $p > 0.15$). This suggested that students' improvements in logical reasoning in the selection tasks associated weakly with their improvements in various aspects of mathematical proving.

Summary of the Experimental Results

This chapter reported the experimental findings gathered from students' pretest and posttest data to address the hypothesis of the study, i.e., logic training emphasizing generation of counterexamples or rule violations enhanced students' reasoning of logical implications, as well as students' ability in mathematical proving, more significantly than the logic training emphasizing truth tables. Each approach constituted a training condition in the design of the experiment. Derived from the main hypothesis were the first four research questions pertaining to the comparison of the benefits of three logic training approaches on students' logical reasoning and various aspects of mathematical proving. These aspects were classified into Proof Construction, Proof Validation and Proof Knowledge, with finer distinction made within each aspect. The fifth research question concerned to the correlations between these abilities.

In each section of this chapter, training effects addressing each research question were reported in order for its statistical significance of the between-subject effects of training conditions, the within-subject effects of repeated assessments and the interaction between these two types of effects. Detailed characteristics of students' improved performance in the assessment items were also furnished to describe the extent of the training effects qualitatively.

Table 19 below summarized the outcomes of ANOVA and the conclusions concerning the effects of logic training on students' logical reasoning, Proof Construction, Proof Validation and Proof Knowledge. Each entry stated the statistical significance and the effect sizes for each of the aspects assessed. Effect sizes were reported only when the training effect was statistically significant ($p < .05$ or less).

Table 19: Summary of ANOVA outcomes and Conclusions

Effects of training	Condition	Assessments	Interaction
---------------------	-----------	-------------	-------------

Table 19 (cont'd)

Logical reasoning	$p < .01$; large effect; Condition W is better than Conditions C and PO	$p < .01$; large effect	$p < .01$; large effect
Deductive-proof Construction	$p > .05$	$p < .05$; moderate effect	$p > .05$
Proof-by-counterexample Construction	$p > .05$	$p > .05$	$p > .05$
Invalidation of Empirical Proof	$p > .05$	$p > .05$	$p > .05$
Validation of Proof-by- contradiction	$p > .05$	$p < .05$; moderate effect	$p > .05$
Logical non-equivalence of converse	$p > .05$	$p < .05$; moderate effect	$p > .05$
Logical non-equivalence of contrapositive	$p > .05$	$p < .05$; moderate effect	$p > .05$

Overall, the effects of these training approaches on students' logical reasoning exhibited significant differences but their effects on students' mathematical proving were statistically similar across conditions, repeated assessments, as well as interaction. Students undergoing the logic training emphasizing counterexample showed significant improved logical reasoning over the other two approaches. For each aspect of mathematical proving, no statistical significance across training conditions and interaction effects were reported in students' pretest and posttest performances. Students undergoing all three logic training approaches showed significant improvements over repeated assessments for Deductive-proof Construction, Validation of Proof-by-contradiction, logical non-equivalence of converse and logical equivalence of contrapositive and no significant improvements for other aspects of mathematical proving. In other words, all three approaches were equally effective in some aspects but ineffective in other aspects of mathematical proving. These outcomes were opposite to the predicted outcomes derived from the hypothesis of the study.

A closer look at the distribution of students' scores revealed more detailed characteristics of improvements in students' mathematical proving. In Deductive-proof Construction, more students exhibited productive use of deductive inferences though not necessarily lead to construction of coherent proofs. In Validation of Proof-by-contradiction, more students accepted the proof because they agreed with the contradiction statement, or because the negation of or the counterexamples to the implication were shown to be false. In determining the logical equivalence between an implication, its converse, and its contrapositive, a lot of students focused on proving whether an implication was logically equivalent to the other statement (converse or contrapositive) over the similarities or dissimilarities in the sentence structure.

Preliminary Discussion

The three logic training approaches had dissimilar effects on students' logical reasoning but similar effects on students' mathematical proving. The contrast of these outcomes required further inquiry. In particular, the phenomenon of similar effects on mathematical proving, significant or not, across all conditions stood opposite to the hypothesis of this study and required some conjectured explanations. This will be carried out in the next chapter when students' interview data were analyzed. However, the issue regarding concerned the practice effects of repeated assessments needed to be resolved before the analyses of students' interview data became meaningful for conjecturing plausible explanations for the similar training effects on students' mathematical proving across conditions.

Students' Improved Aspects of Mathematical Proving: Practice Effect or Training Effect

Practice effects occurred when students enhanced their abilities solely due to repeated attempts of a task. In this study, training approaches in all three conditions provided worked-out examples to serve as feedback for students' learning in a self-paced problem-solving context. The design in this study did not include a condition for controlling practice effects on Proof Construction since prior similar studies had established that worked-out examples of problem solutions and mathematical proofs facilitated learning effect for enhancing students' problems solving and Proof Construction (Atkinson, Derry, Renkl, & Wortham, 2000; Hilbert, et al., 2008; Zhu & Simon, 1987). Students' processing of the explanation and solution steps provided in the worked-out examples served as a means of cognitive modeling (Collins, et al., 1989; Schoenfeld, 1985).

To add another guard against practice effects, the pretest and posttest posed different mathematical implications across isomorphic items for Deductive-proof Construction and Proof-by-counterexample Construction. For instance, students who were asked to prove an implication involving quadratic graphs in one test would be asked to prove a different implication involving quadratic factoring. Also, the other tasks posed implications using different combinations of number operations across the two test sets.

Likewise, two worked-out examples supporting students' performance in Proof Validation were provided in the training materials across all conditions. The examples explained why deductive proofs were valid and empirical proofs were invalid in accordance to the logical bases emphasized by each training condition. Moreover, different mathematical implications were also being posed in the assessment item of Validation of Proof-by-Contradiction across the

two test sets. One concerned the addition and division of two numbers and the other concerned the average and square roots of numbers.

The absence of worked-out examples illustrating Proof Knowledge related to the converse and the contrapositive of an implication in the training materials seemed to lend suspect to practice effects on students' improved performance in the assessments. However, the practice effects were unlikely due to the posing of different implications across isomorphic test items. Furthermore, the students' responses to these tasks were unlikely improvements due to practice effects. Despite the logic training received, at least 55% of students' responses still considered the converse as logically equivalent to the implication (Figure 17, the bar columns representing level 0), and that at least 75% of students' responses were focusing on proving the contrapositive and the implication (Figure 18, the bar columns representing level 1), when considering their logical equivalence with an implication. For the former, practice effects were unlikely when more than half of students still failed to deduce the logical non-equivalence. Improved use of deductive inferences in the latter had already been established as a learning effect due to the worked-out examples of Proof Construction instead of practice effects.

In sum, I considered the likelihood of practice effect being observed in students improved performance in the mathematical proving tasks. Based on three considerations: (1) Prior studies of students' learning from worked-out examples in mathematical problem solving and mathematical proving, (2) the design of isomorphic test instruments, and (3) the students' responses in the Proof Knowledge tasks, practice effects was assessed to be unlikely to account for students' gains. Instead, their improved performances were attributed to the effects of the three training conditions. With this issue being resolved, we will now turn to the next chapter for more in-depth accounts of the experiment results based on students' interview data.

CHAPTER 5 POST-STUDY INTERVIEW RESULTS

This study hypothesizes that logic training emphasizing generation of counterexamples can bring beneficial effects to students' reasoning of logical implications as well as students' ability in mathematical proving, and pit this approach as a better alternative than the approach using truth tables. The experimental results reported in the previous chapter were mixed. On one hand, students' performances were significantly improved for Deductive-proof Construction, Validation of Proof-by-contradiction, logical non-equivalence of converse, and logical equivalence of contrapositive across conditions after training. On the other hand, no significant improvements were found for Proof-by-counterexample Construction and Invalidation of Empirical Proof. Evidently, the experimental outcomes do not support the hypothesis of this study. An interesting question thus arises: Why did students in all three Conditions exhibited similar improved performance (or unaffected performance) in these outcomes?

The purpose of this chapter is to re-examine the design of the study and the underlying theoretical framework of the proposed hypothesis, and generate some conjectured explanations. I would first explore the controlled and uncontrolled factors of the experimental design, as implied by the experimental results. Guided by the theoretical framework of the study, I would then propose plausible theoretical explanations. Using these theoretical explanations to structure the analysis of students' mathematical proving works in the post-study interview data, I would next proceed to report the findings that provided evidences for the conjectured explanations. Posing a mathematical proving task in an interview setting was a credible source of evidence as it captured the processes of students' Proof Constructions, including relevant characteristics which were otherwise unobserved in the experiment. Experimental results would be drawn upon to inform the inquiry process whenever they were useful.

Controlled and uncontrolled factors of the experimental design

Recall that the experimental study is a two-factor study, one of which is a between-subjects factor of three levels (three training conditions) and the other a within-subjects factor of two levels (two repeated assessments). The ANOVA of students' outcomes showed that no significant impact of the training conditions but significant impact of the repeated assessments was evident. Evidently, some factors, other than the different training approaches, were the causes and remained to be identified.

Since practice effect had been ruled out for Condition C (see Chapter 4, *Preliminary Discussion*), the impact could only be attributed to one or more factors controlled within Condition C. By the same argument, the observed impact in Conditions PO and W was attributed to one or more factors controlled by each Condition. These controlled factors might be unique to each Condition or shared by two or more Conditions. As differences between Conditions did not exert a differential impact, the impact of those factors uniquely controlled by each Condition must be comparable or such factors did not exist.

The above possibilities only accounted for the training outcomes in which significant impact was observed, i.e., students' improved performance in Deductive-proof Construction, Validation of Proof-by-contradiction, logical non-equivalence of converse, and logical equivalence of contrapositive. For the other two outcomes (Proof-by-counterexample Construction and Invalidation of Empirical Proof) in which no significant impact was observed, they were not affected by the controlled factors. However, some other uncontrolled factors of the experimental design might be at play.

The re-examination of the experimental study suggested that the underlying factors affecting students' various aspects of mathematical proving could be some controlled factors located within and also other uncontrolled factors were located outside of the training conditions of the experimental study. Next we identify these factors based on the theoretical constructs of logical reasoning of implications and Proof Construction set out in the study.

Theoretical constructs: Logical reasoning and Mathematical Proving

This study conceptualized logical reasoning of implications as deriving a conclusion about the implication statement “If P then Q ,” where P is the antecedent and Q is the consequent, according to the logical criteria of truth and falsity. The implication is false when a counterexample to the implication exists, i.e., a mathematical object is found to satisfy the antecedent P and not the consequent Q ; otherwise, the statement is concluded true. Mathematical proving of an implication was conceptualized as the search for examples or counterexamples of mathematical objects related to the implication (Durand-Guerrier, 2003). The search for examples and counterexamples implied that the process of mathematical proving involved the interpretation of mathematical objects specified in the implication, the representation of these objects and the use of these mathematical representations in a deductive manner.

These three theoretical causes circumscribed the sources of controlled and uncontrolled factors that affected students' various aspects of mathematical proving in the experimental study. The training materials used in all conditions was probably the prime source of the controlled factors since students' improved performances were primarily driven by the designed proving tasks and the worked-out examples in the materials. The theoretical cause most relevant to the

design of training materials would plausibly be hypothesized as the controlled factor while the theoretical cause least relevant would plausibly be hypothesized as the uncontrolled factor.

All worked-out examples demonstrated the mathematical proofs using the same set of mathematical representations across conditions. It would then seem that the use of mathematical representations was a controlled factor underlying all training conditions in common. Though students had prior formal classroom exposure to the mathematical contents used in the training materials, their mathematical competence in interpreting the mathematical objects specified in the tasks and the implications would likely be an uncontrolled factor. In relation to that, students' choice of mathematical representations was also likely uncontrolled since the training materials did not instruct students on how to evaluate mathematical representations for the proving processes. Based on the list of causes generated, there seemed to be no candidate for controlled factor that was unique or common to only two of three training conditions.

Together, the theoretical inquiry and re-examination carried out thus far had hypothesized plausibly that students' interpretation of implication, their choice and use of the representation were the factors underlying the experimental outcomes of students' mathematical proving. However, the credibility of these conjectured factors remained dubious without any empirical substantiation. The next step is thus to verify the plausibility and provide further insights into these conjectured factors through an empirical inquiry, which motivated the subsequent sections on the analysis and findings of the students' post-study interview data in this chapter.

Apart from explaining why students exhibited similar improved or unaffected performance in Proof Construction, Proof Validation and Proof Knowledge across Conditions, the analysis of the students' interview data also needed to address the sixth research question: How do students modify a falsifiable mathematical implication to a mathematically true

implication based on their self-generated examples and counterexamples? Given the experimental outcomes, I expected students' processes observed in the data would look similar across conditions. However, the findings might provide more insights into students' logical reasoning and mathematical proving of implications.

A caveat is warranted at this point. The proposed theoretical causes were, of course, non-exhaustive as the theoretical framework of this study might have excluded other plausible factors. The limitations of theoretical issues will be addressed in the last chapter. Nonetheless, the inquiry had laid out a framework for structuring the subsequent analyses and findings.

Analysis of Interview Data

In this section, a brief background about the logical reasoning and mathematical proving abilities of the selected interviewees is provided at the start. Before going to the findings of the interview, the interview and the data coding are briefly described. Coding schemes and samples of coded transcripts showing the assignments of codes are presented to illustrate how the findings from the interview were generated.

Background of the interviewees

We will first look at an overview of the 12 interviewees' mathematical proving work. 12 students (4 boys and 8 girls; labeled with IDs S1 to S12) participating in the interviews came from all three school sites and across three training conditions: 4 from each site and 4 from each training condition. Their raw scores for each assessed component in the posttest, in terms of logical reasoning, Proof Construction, Proof Validation and Proof Knowledge, are shown in Table 20. The legend of the abbreviations in the column headings were provided at the bottom of

the table. The ranges of the scores for each component are shown by the parenthesized numerical range found in the first row. Two sub-columns, “NT” (Number Theory) and “Quad” (Quadratics), were created within the columns of DP (Deductive-proof) and PC (Proof-by-counterexample). The “NT” sub-column recorded students’ scores for tasks involving Elementary Number Theory while the “Quad” sub-column recorded their scores for tasks involving Quadratics.

Table 20: Post test performance of Interviewees

ID	Cond	LR (-16 to 16)	DP (0 to 6)		PC (0 to 5)		I-EP (0 to 2)	V-PC (0 to 3)	LnE-C (0 to 3)	LE-C (0 to 2)
			[NT]	[Quad]	[NT]	[Quad]				
S1	W	16	2	5	3	1	2	0	1	0
S2	W	16	5	4	3	1	2	0	0	2
S3	W	12	1	1	4	4	2	2	0	1
S4	C	2	1	1	4	1	0	0	2	1
S5	PO	16	3	2	5	0	1	0	0	1
S6	C	0	3	1	5	2	1	0	1	1
S7	C	0	4	4	5	4	2	2	0	1
S8	W	16	6	1	4	4	2	3	0	1
S9	PO	16	5	3	3	4	2	1	3	1
S10	PO	0	4	1	5	3	2	1	1	1
S11	PO	0	2	3	4	1	1	0	3	1
S12	C	-8	4	1	5	1	0	0	2	1
Legend ID – Identification LR – Logical Reasoning NT – Number Theory PC – Proof-by-Counterexample V-PC – Validation of Proof-by-Contradiction LE-C – Logical Equivalence of Contrapositive Cond – Condition DP – Deductive-proof Quad – Quadratics; I-EP – Invalidation of Empirical Proof LnE-C – Logical non-Equivalence of Converse										

The logical reasoning and mathematical proving abilities of these interviewees were somewhat aligned with the experimental results. The logical reasoning scores of the interviewees from the Condition W (12 to 16) were generally higher than those from the Condition PO (0 to

16), which was in turn higher than those from the Condition C (-8 to 0). They performed better for the Proof Construction tasks (Deductive-proof and Proof-by-counterexample Construction) involving Elementary Number Theory in comparison to Quadratics. Their understanding of the invalidity of Empirical Proofs was not uniform across conditions – all four interviewees from condition W were clear about why Empirical Proofs were invalid (at level 2) while none from Condition C did. However, more than half of the interviewees did not accept the Proof-by-contradiction. Their performances in the Proof Validation tasks were reflective of the post test results in which high proportion of students (at least 60%) did not accept the valid proof.

The Proof Knowledge exhibited by the interviewees was also not uniform. Two out of four interviewees from condition PO were able to relate the logical non-equivalence of converse to the counterexamples (level 3), while none from condition W did. In comparison, the interviewees exhibited uniform knowledge of the logical equivalence of contrapositive (mostly at level 1). Overall, this selection of interviewees was representative of the experimental findings in terms of logical reasoning and most aspects of mathematical proving, except for their invalidation of Empirical Proofs and logical knowledge about converse.

The Interview Task and Data Coding

The video-recorded interview session was carried out by me on a one-to-one basis. Students were presented the following impromptu task: “A set of at least 3 whole numbers are randomly chosen and their average is calculated. Decide whether the following statement is true or false: / *If the set of numbers were consecutive in order, then their average is a whole number.* / Justify why your conclusion must be true or false using the most convincing argument.” Note that the given implication referred to sets of at least three numbers and is thus false whenever an

even number of consecutive numbers is considered, e.g., the average of “1, 2, 3, 4” is 2.5 and not a whole number. However, students might consider different sets of numbers pending their interpretations.

The student worked on the task for some time with paper and calculator provided. The interviewer then asked the student about their conclusions and justifications when they appeared to have reach a conclusion. If the student’s conclusion was “True” and had no intention of revision, the interviewer would prompt the student to consider other sets of numbers that would falsify the implication. After students had falsified the implication, they were then asked to modify the false implication to a mathematically true implication and to provide mathematical justifications for it.

Coding of Students’ Proving Attempts

The audio-video data were transcribed. The beginning of a proving attempt was marked by the student’s mathematical actions and its ending by the student’s response of his/her conclusion. The student’s attempts were coded for the conclusion and the level of the proof constructed. For coding the interviewee’s mathematical proving performance in each attempt, the coding scheme of Deductive-proof Construction was used when sets of odd number of mathematical objects were considered. The coding scheme of Proof-by-counterexample Construction was used when sets of even number of mathematical objects were considered (See Chapter 3, Table 4 and Table 5 for both coding schemes).

The coding of students’ proving attempt was guided by the theoretical inquiry mentioned earlier. In coding for students’ interpretation of mathematical objects, the set of mathematical objects they considered in their attempts were identified from the transcripts. In addition, the

number of mathematical objects considered was also coded. The types of mathematical representations used in each attempt were also identified. These were mainly numerical or algebraic representations as preliminarily observed from students' articulation or written work.

Table 21 below shows the coding scheme used.

Table 21: Coding scheme for students proving attempts

Category of data	Codes to be assigned
Prove attempt	Indicate the number of attempts thus far
Conclusion of proof	'True' 'False'
Level of Proof	If the conclusion is 'True,' code student's proof using the coding scheme for Deductive-proof Construction If the conclusion is 'False,' code student's proof using the coding scheme for Proof-by-counterexample Construction
Sets of mathematical objects considered	Indicate the number of objects considered
Representations	'Numerical' if numeric symbols were written or spoken 'Algebraic' if algebraic symbols were written or spoken
Use of representations	Indicate the specific symbols used

A sample of the coded transcript is shown in Table 22.

Table 22: Sample of coded transcript

	Transcribed conversation and actions	Codes assigned
S11	wait a while. [takes out a calculator.]	
I	so you understand the situation?	
S11	yes. [long pause, working on calculator a few times. three numbers added each time]	Proof attempt 1; Numerical representations; True
S11	yes, the statement will always work	
I	ok, so your conclusion is true	
S11	yeah.	
I	ok, um, why? It will always work, so how?	

Table 22 (cont'd)

S11	[laughs, starts writing down] because if, I can try to explain or I can just list out the again the 15 examples?	15 sets of three consecutive numbers
I	whichever way that you can support your reason, your decision here.	
S11	um, how to say? You always have 3 as a factor.	Level 2 (Proof attempt 1)
I	Ok	
S11	because if consecutive numbers right, so it's like [writes down, $x + (x + 1) + (x + 2)$] $x + x + 1 + x + 2$	Proof attempt 2; Algebraic representations; three terms considered – “ $x, x + 1, x + 2$ ” True Level 6 (Proof attempt 2)
I	Mmhmm	
S11	if this is the set of consecutive numbers	
I	Ok	
S11	Right	
I	yeah. Mmhmm	
S11	then you will always have, you will become $3x + 3$	
I	Ok	
S11	so this will always be divisible by 3	
I	divisible by 3. ok if it is divisible by 3, then the average would be	
S11	[writes down 'average = $(3x + 3)/3 = x + 1$ ']	
I	which is $x + 1$	
[Part of transcripts here were omitted for illustrative purposes]		
I	but the set of whole numbers here says,' at least three whole numbers'	Prompted to consider four numbers
S11	Yeah	
I	so at least three whole numbers refers to three whole numbers, but if it can also be other numbers of whole numbers	
S11	ok, I try [use calculator to check].	

These codes were then further consolidated and classified for analyses. The sample of coded transcripts will be referenced again in some subsequent sections to highlight the typical aspects of interviewee's mathematical proving attempts.

Coding of Students' Modification of Implication

As the implication relates the antecedent to the consequent, students' modifications to the implication were classified into three types: (A) modifications made to the mathematical objects

described by the antecedent, (C) modifications made to the mathematical objects described by the consequent, (M) modifications made to mathematical objects described by both antecedent and consequent. In the first type of modifications, students restricted the set of consecutive numbers being specified in the antecedent. In the second type of modifications, student extended the numerical property of the average as specified in the consequent. In the third type, student restricted the set of consecutive numbers as specified in the antecedent and altered the property of the average as specified in the consequent. Table 23 below shows a sample of modified implication of each type to illustrate the coding. The underlined words and phrases were students' modifications.

Table 23: Sample of Modified Implications

Student	Modified implications	Type	Modification
S4	If the set of whole numbers are consecutive in order <u>and are of an odd number, like 3 or 5 numbers</u> , then their average is a whole number.	A	Odd numbers
	If the set of whole numbers are consecutive in order <u>and of an even number, like 4 to 6 numbers</u> , then their average is <u>not</u> a whole number.	M	Even numbers; non whole numbers
S5	If the set of whole numbers are consecutive in order, then their average is a whole number <u>or a decimal number</u> .	C	Decimal numbers

As the modification task was open-ended, the number of coded modifications varied across students as shown in the above table.

Overview of students' proving behaviors

Table 24 below shows the frequencies of proving-related actions taken by the interviewees and the duration of the interviews. The number of proving attempts made in the

entire task, the number of modifications made to the implication and the number of attempts to prove the modified implications was counted from the coded data. Whether students were prompted to consider sets of more consecutive numbers was also noted.

Table 24: Overview of Students' proving of the impromptu task

Interview events	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
Number of attempts	2	4	2	3	1	3	4	2	5	4	3	5
Prompted for other sets	Yes	No	No	No	No	Yes	No	Yes	No	Yes	Yes	Yes
Number of Modifications	2	1	3	2	2	1	1	3	5	2	1	2
Number of attempts after modifications	3	2	3	3	4	3	3	4	3	3	3	3
Time spent (min:sec)	10:32	22:46	18:29	15:19	9:00	10:16	16:13	17:04	16:55	9:38	21:14	12:15

Overall, the time spent by each student varied substantially (from nine to almost 23 minutes) and about five to nine proving attempts were made. However, the duration of the interview did not necessarily correspond to the number of the attempts, e.g., compare S12 and S2. Most students attempted three or more proofs in proving the given implication except S1, S3 and S5. Almost all attempted three or more proofs after they had modified implications, except for S2.

Recall that the analysis of interview was concerned with explaining the experiment outcomes of students' mathematical proving based on the three theoretical factors as well as addressing the sixth research questions. In the subsequent sections of reporting the findings from the interview, I first turn to how students considered the sets of mathematical objects specified in the implication to address students' interpretation of implication, which was conjectured as an

uncontrolled factor. Following that, I will examine students' proving attempts, the representations used and the level of the constructed proof (Deductive-proof or Proof-by-counterexample) scored in each attempt to address students' representations of mathematical objects and their use of representations. Subsequently, I will examine the types of modifications student made to the implication in relation to their proving attempts. Finally, these findings were put together to infer conjectured explanations for the experimental results and to address the sixth research question.

Students' Considerations of Mathematical Objects

Students considered different sets of mathematical objects based on their interpretation of the given implication which referred to sets of at least three numbers. This description was less specific than what they had encountered in the training materials. Hence this task appeared more novel. However, some students had limited considerations of the number sets and needed to be prompted after some attempts. Table 25 below shows the set of mathematical objects, specifically the number of numbers, being considered by each student in each attempt. A distinction between whether numbers or algebraic terms were used was also noted. Attempts in bold print indicated the student needed to be prompted to consider the task description "set of at least three numbers" for the attempt.

Table 25: Students' considerations of mathematical objects for proving implication

ID	Attempt 1	Attempt 2	Attempt 3	Attempt 4	Attempt 5
S1	3 numbers – "1, 2, 3"; and "2, 3, 4"	4 numbers – "1, 2, 3, 4"			
S2	17 terms – " x , $x+1$, $x+2$, ..., $x+17$ "	4 terms – " x , $x+1$, $x+2$, $x+3$ "	5 numbers – "2, 3, 4, 5, 6"	4 numbers – "2, 3, 4, 5"	

Table 25 (cont'd)

S3	4 numbers – “7, 8, 9, 10”	6 numbers – “1, 2, 3, 4, 5, 6”			
S4	4 terms – “ $x, x + 1, x + 2, x + 3$ ”	3 terms – “ $x, x + 1, x + 2$ ”	5 terms – “ $x, x + 1, x + 2, x + 3, x + 4$ ”		
S5	4 numbers – “1, 2, 3, 4”				
S6	3 terms – “ $x, x + 1, x + 2$ ”	4 terms – “$x, x + 1, x + 2, x + 3$”	5 terms – “ $x, x + 1, x + 2, x + 3$ ”		
S7	Odd number of numbers	3 numbers – “1, 2, 3”	Odd number of numbers	4 numbers – “1, 2, 3, 4” and “2, 3, 4, 5”	
S8	3 numbers – “1, 2, 3”, “2, 3, 4”, “3, 4, 5”, “4, 5, 6”	4 numbers – “1, 2, 3, 4”			
S9	3 numbers – “1, 2, 3”, “2, 3, 4”, “4, 5, 6”	3 terms – “ $n, n + 1, n + 2$ ”	4 terms – “ $n, n + 1, n + 2, n + 3$ ”	4 terms – “ $n + 1, n + 2, n + 3, n + 4$ ”	4 numbers – “1, 2, 3, 4”
S10	3 numbers – “1, 2, 3” and “4, 5, 6”	3 terms – “ $n, n + 1, n + 2$ ”	4 terms – “$n, n + 1, n + 2, n + 3$”	4 numbers – “100, 101, 102, 103”	
S11	3 numbers – 15 sets	3 terms – “ $x, x + 1, x + 2$ ”	4 terms – “$x, x + 1, x + 2, x + 3$”		
S12	3 numbers – “1, 2, 3”	3 numbers – “1, 4, 6”	4 numbers – “1, 2, 3, 4”	5 numbers – “1, 2, 3, 4, 5”	6 numbers – “1, 2, 3, 4, 5, 6”

As illustrated in Table 25, the mathematical objects considered were mostly expressed as numbers or algebraic terms, either verbally or in writing. Seven students started considering three consecutive numbers or terms and extended their consideration to sets of more objects subsequently. Of these students, six students were prompted to consider what “set of at least three numbers” meant. Only upon prompted, they inferred that sets of four or more numbers were meant and made another proving attempt.

The other four students started considering sets of four or other mathematical objects instead. While students S7 and S9 choose other sets of numbers for unknown reasons, some

students made their decision because they wanted to avoid being biased, e.g., S2 chose 17 terms, which involved some complicated calculations for her, and S3 chose four numbers (see Table 26).

Table 26: Students’ reasons for choosing a set of numbers

ID	Reasons provided during interview (in verbatim)
I:	So I see you do a lot of things and press the calculator, what happens, what are the numbers you have considered?
S2:	just add, 1, 2, 3, 4, 5, 6, then add until some point that use calculator to divide.
I:	so you try four whole numbers.
S3:	Yeah, actually I try four and try five, I did not try three because sometimes in the mathematical way, you didn’t do something complicated, the number would be what you are thinking.
I:	So far what have you done with the calculator?
S8:	spamming with random numbers
I:	random numbers. How many of them?
S8:	for countless, all from 1-2-3, 2- 3- 4, 3-4-5, 4-5-6,
I:	Ok, so if you, so now based on what you have found out about this, would you say this [pointing to the given implication] is true or false again?
S6:	False. Because I didn't know it was "at least three" [<u>underline the three words in the question</u>].

S3 made a conscious decision to avoid the set of three numbers because of her concern that simple mathematical actions tend to confirm her prediction of mathematical outcomes. S2 made a “random” selection of numbers by making a spontaneous stop with keying of numbers and proceeded with a calculation of their average. Contrast with the “random” move made by S8, S2 randomized her choice about the size of the number set to be considered while S8 limited his choice to three consecutive numbers but “randomly” selected what those numbers were. S6 limited her choice to three consecutive numbers and did not notice more numbers were included until prompted. Her interpretation of the set of objects was representative of the six students.

Students' consideration of mathematical objects suggested that their considerations were often influenced by their prior mathematical knowledge and lacked a process of logical check. Half of the students' consideration of "at least three numbers" was fixated to only three numbers and lacked a cognizance of other sizes. Of the other six students, their considerations of numbers were varied and often grounded in some sophisticated heuristics of bias aversion other than logical reasoning. Overall, students' considerations of mathematical objects appeared to lack an effective process of logical reasoning to rectify the errors in their interpretations.

Students' Deductive-proof and Proof-by-counterexample Constructions

In all attempts (except the first and third attempt of S7, which shall be discussed later) made by students, interviewees either used numerical or algebraic representations and operations to prove or falsify the implication with the set of mathematical objects considered and concluded whether the implication was true or false. Table 27 below shows what type of representations and operations were used by each student in each mathematical proving attempt. In each cell, the last number is the level of proof being scored for the particular attempt and the word, "True" or "False," is the conclusion students made for that attempt. Specific numbers or terms considered by the students have been listed in Table 26.

Table 27: Types of representations, Conclusions and Level of Proof

ID	Attempt 1	Attempt 2	Attempt 3	Attempt 4	Attempt 5
S1	3 numbers; Numerical; True – 1	4 numbers; Numerical; False – 4			
S2	17 terms; Algebraic; True – 6	4 terms; Algebraic; False – 5	5 numbers; Numerical; True – 1	4 numbers; Numerical; False – 4	
S3	4 numbers;	6 numbers;			

Table 27 (cont'd)

	Numerical; False – 4	Numerical; False – 4			
S4	4 terms; Algebraic; False – 5	3 terms; Algebraic; True – 6	5 terms; Algebraic; True – 6		
S5	4 numbers; Numerical; False – 4				
S6	3 terms; Algebraic; True – 6	4 terms; Algebraic; False – 5	5 terms; Algebraic; True – 6		
S7	‘Odd’; ‘Center’; True – 3	3 numbers; Numerical; True – 1	Odd; ‘Balance’; True – 4	4 numbers; Numerical; False – 4	
S8	3 numbers; Numerical; True – 1	4 numbers; Numerical; False – 4			
S9	3 numbers; Numerical; True – 1	3 terms; Algebraic; True – 6	4 terms; Algebraic; True – 1	4 terms; Algebraic; False – 5	4 numbers; Numerical; False – 4
S10	3 numbers; Numerical; True – 2	3 terms; Algebraic; True – 6	4 terms; Algebraic; True – 1	4 numbers; Numerical; False – 4	
S11	3 numbers; Numerical; True – 2	3 terms; Algebraic; True – 6	4 terms; Algebraic; False – 5		
S12	3 numbers; Numerical; True – 2	3 numbers; Numerical; True – 1	4 numbers; Numerical; False – 4	5 numbers; Numerical; True – 2	6 numbers; Numerical; False – 4

Use of Numerical Representations in Deductive Proofs or Proof-by-counterexample

For students who adopted numerical representations and operations to prove the implication, their proofs were usually scored at level 1 or 2. The latter depended on whether students inferred additional explanations, such as observed properties, to justify why the implication was true. For example, S11’s first attempt was scored level 2 because he calculated the total of three consecutive numbers and divided by 3 for 15 sets of numbers and provided

additional explanations (See Table 22, proof attempt 1 by S11). For students who “proved” the implication false more than once, their final conclusion for the implication is false.

For students who adopted numerical representations and operations to falsify the implication, they usually selected a set of four or six small consecutive numbers, e.g., 1 to 4. They calculated the average of the number set with a calculator, by summing the numbers and dividing the total by the number of consecutive numbers in the set. Their proofs were scored at level 4, indicating that they had falsified the given implication with isolated instances of the average being a non-whole number.

Use of Algebraic Representations in Deductive Proofs or Proof-by-counterexample

For students who adopted algebraic representations and operations, their proofs were usually scored at level 6. The algebraic representations of any consecutive numbers in algebraic terms similar to “ $x, x + 1, x + 2 \dots$ ” supported the students in constructing a concise and clearer proof. Students applied subsequent algebraic additions and divisions to obtain the average of the terms in the form of $x + a$. For example, S11 represented three consecutive numbers in the form of “ $x, x + 1, x + 2$ ” and obtained the average $x + 1$. Since x represented a whole number, he deduced that the average is a whole number if the set of three numbers were consecutive (See Table 22, proof attempt 2 by S11).

For students who falsified the implication, they chose a set of 4 (or other even number) terms, n to $n + 3$, to represent any four (or other even number) consecutive numbers. They obtain an average of $n + 2.5$ by dividing the algebraic sum by 4 (or other even number). Their proofs were usually scored at level 5, indicating that they had falsified the implication based on a

generic set of counterexamples in which the average was not a whole number for any four (or other even number) consecutive numbers.

Most students falsified the implication when they considered an even number of consecutive numbers. However, S9 and S10 concluded the implication to be true (see the third attempt of S9 and S10) due to some mathematical errors made during the computation of the average. Their use of incorrect mathematical examples led to logically invalid conclusions of being true and thus their proofs were scored at level 1.

Students' choice of numerical and algebraic representations appeared to have an impact to the types of Deductive proof and Proof-by-counterexample constructed. However, how students used the representations and the operations to derive a logical conclusion from the results also mattered.

Alternative use of representations in Deductive Proofs

Among all interviewees, S7's attempt was atypical – she considered a generic set of odd number of mathematical objects and based her inferences on mathematical knowledge other than numerical or algebraic operations after she was briefed on the task. She held the premise that the average of a generic set of consecutive numbers lies exactly at the 'center' of the set based on her prior knowledge. If the set has an odd number of consecutive numbers, that 'center' would be the middle number of the set. Upon request for more explanations, she illustrated her idea of 'center' using the set "1, 2, 3" but was pressed for an explanation that applied to a generic set. S7 then introduced a 'balanced' conception in her third attempt (Excerpt 1).

Excerpt 1: Average is the ‘Center’ number of a ‘balanced’ set

- S7: Yeah, because basically you are just finding the, how to say, um, average, you just finding the, like everything then the *moderate number*, yeah, whereby it is something like *constant*. yeah, then, so the center one is basically the more fair, because both are of *equal* [underlines the space to the left of '2'], equal [underlines the space to the right of '2'] value. I mean they both have *equal number of integer* [circling '1' and '3'].
- I: Ok, so you take the *middle number* [points to '2'] and on the right hand side you have equal number of integers [point to the space to the right of '2', including '3'],
- S7: Yeah.
- I: On the left side integers [point to the space to the left of '2', including '1'],
- S7: Yeah *equal number*.
- I: Of integers on the left side.
- S7: Yeah. Because this one [points to '1, 2, 3'] only have 1 [points to '1'], this one only have 3 [points to '3'], so 1 and 3 *equal balance*.

S7's knowledge of the average being the ‘center’ number of a set having an odd number of consecutive numbers was grounded in a ‘balance’ conception, in which as many numbers were on the left side as on the right side of the middle number. The number of consecutive numbers was not crucial as long as there were an odd number of them. However, her justifications for the conception were not logically valid. Nevertheless, just as the numerical and algebraic representations and operations supported other students in their Proof Constructions, the ‘balance’ conception supported S7 in applying her knowledge of average to generic sets of consecutive numbers though it was unclear what the underlying mathematical operations were. Further analysis of students’ Deductive-proof Construction for modified implications would provide more insights about students’ representations and their use of representations.

Students’ Deductive-proof Construction for modified implications

In between the process of modifying implications, students constructed deductive proofs to justify the implication they made. These attempts in the data were identified and labeled in

order of appearance. Note that the number of proofs constructed by a student did not correspond exactly to the number of modifications. Although these attempts occurred after students' modification of implication in the interview transcripts, they were reported ahead in this section for a coherent analysis of students' representations and their use of representations.

Table 28 below groups the students' attempts according to their use of representations. Each entry described the representations students generated. The students were classified into four groups according to how the representations were used throughout these attempts: (1) Use of numerical representations only; (2) Use of algebraic representations only; (3) Mixed use of numerical and algebraic representations; and (4) Other uses. Students' attempts often included a modified implication in which an odd number of consecutive numbers was specified in the antecedent (except for S5, see Table 29 at later section for details). Such attempts were italicized in print.

Table 28: Students use of Mathematical Representations by groups

ID	Representations used
Group 1: Numerical representations only	
S1	1. Three numbers – “1, 2, 3”, “2, 3, 4”, “3, 4, 5”, “4, 5, 6” 2. Random sets of three numbers – “55, 56, 57”, “99, 100, 101” 3. <i>Multiple sets of three, four, and five consecutive numbers</i>
S3	1. Three consecutive numbers – “1, 2, 3”, “4, 5, 6”, “7, 8, 9” 2. Five numbers – “1, 2, 3, 4, 5” 3. <i>Averages of the sets – “1, 2, 3, 4, 5”, “1, 2, ..., 6, 7”, “1, 2, ..., 8, 9” and “1, 2, ..., 10, 11 “</i>
S12	1. <i>Each sum of three, five, seven consecutive numbers can lead to a whole number average – $1 + 2 + 3$, $1 + 2 + 3 + 4 + 5$, $1 + 2 + 3 + 4 + 5 + 6 + 7$</i> 2. <i>The sum of three, five, seven, nine and eleven numbers of $1 + 2 + 3 + \dots$, is a</i>
	3. <i>The sum of three, five, seven, nine numbers of $1 + 2 + 3 + \dots$, are multiples of the number of terms - 3 times 2, 5 times 3, 7 times 4. The second multiplicand increased by 1 as the odd number increased.</i>

Table 28 (cont'd)

Group 2: Algebraic representations only	
S4	1. Average of five consecutive terms, "$x, x + 1, \dots, x + 4$" 2. Average of six consecutive terms, "$x, x + 1, \dots, x + 5$" 3. Sum of consecutive numbers, x to $x + n$, is $[x + (x + n)] n$.
S6	1. Three, five, seven, nine, and eleven terms – "$x, x + 1, x + 2, \dots$," 2. Total of a set of consecutive terms $-ax + b$; b is a multiple of a 3. Nine terms – "$x, x + 1, \dots, x + 8$"
S10	1. Three terms – "$n, n + 1, n + 2$" 2. Sum of three, five and seven consecutive terms are $3n + 3, 5n + 10$, and $7n + 21$. 3. Total of an odd number of consecutive numbers is $an + b$; b is related to a.
Group 3: Mix of Numerical and Algebraic Representations	
S2	1. Total of three terms to Total of eight terms – $[x + x + 1 + x + 2 = 3x + 3]$ to $[x + \dots + x + 7 = 8x + 28]$ 2. Average of 3, 5, 7 consecutive terms increased by 1
S5	1. Four numbers – "$1, 2, 3, 4$" 2. Five numbers – "$1, 2, 3, 4, 5$" 3. Three terms – "$x, x + 1, x + 2$" 4. Sum of a set of consecutive numbers
S8	1. Three numbers on calculator 2. Three terms – "$x, x + 1, x + 2$"; Average is $x + 1$. 3. Five terms – "$x, x + 1, x + 2, x + 3, x + 4$"; Average is $x + 2$. 4. Seven terms – "$x, x + 1, \dots, x + 5, x + 6$"; Average is $x + 3$.
S9	1. Five terms – "$n, n + 1, \dots, n + 4$"; Average is $n + 2$. 2. Four numbers - "$1, 2, 3, 4$" 3. Random set of consecutive numbers – 111 to 115, 10 to 14, and 11 to 17; Average is the middle number when the set has odd numbers.
S11	1. Three and four terms – "$x, x + 1, x + 2$" and "$x, x + 1, x + 2, x + 3$"; 2. Average of 1 to 11 3. Average is "The sum of constants / constants = whole number". 4. Average is the algebraic formula "$(1 + 2 + 3 + 4 + 5 + \dots + n)/n = \text{whole number}$."
Group 4: Other uses	
S7	1. Repeated the 'balance' conception for an odd number of consecutive numbers. 2. Three numbers – "$1, 2, 3$" and the average is the 'center' of the set. 3. Sum and average of the last digit of consecutive numbers – "$71, 72, 73$" and "$74, 75, 76$"

Students in group 1, who used numerical representations and operations throughout, could only construct empirical proofs and inferred some relevant properties at most. For

example, S12 inferred a numerical relation between the two factors of the sum of consecutive numbers based on a consistent pattern among a few instances.

Students in group 2, who used algebraic representation throughout, could construct deductive proofs if the set of consecutive terms was of a specific odd number. The algebraic sum of the set of numbers was often reduced to the form of $ax + b$, where a is the number of terms and b is the sum of the constants. However, when the sets were extended to a generic odd number of terms, they could at best infer a mathematical pattern of b being a multiple of a from their observations of a collection of instances. S4 observed that the consecutive terms can be paired in a way that each pair summed up to be the same term as $(x + x + n)$ but he made an error in over-counting the number of pairs.

Similar to group 2, students in group 3 could construct deductive proofs if the set was made up by a specific odd number of consecutive terms. However, they could at best infer from a number of instances that the average is the middle number of the set or in an equivalent algebraic form.

S7 was the only student in group 4. As shown in Excerpt 5.1, she used a ‘balance’ conception to consider how the average of an odd number of consecutive numbers was the ‘center’ number of the set but could not prove clearly how her conception worked for a generic set. In her last attempt, she proposed how the end-digits of the numbers might determine the actual average, using two triplets of numbers, “71, 72, 73” and “74, 75, 76” (Table 28).

None of the students were able to construct a deductive proof of why the average is a whole number depends on whether an odd number of consecutive numbers was considered. Their proofs were all scored at level 4 or below. They met an impasse sooner or later as they tried to improve their proofs. Some were able to infer mathematical properties (e.g., the average

is the middle number of the set or an algebraic expression for the average of consecutive numbers) which, if proven, would constitute a deductive proof.

Students' representations and their uses in Proof Construction

Students' choice of numerical or algebraic representations seemed to be resistant to impasse in logical reasoning. In proving the given implication, some adhered to numerical representations and were limited in the proofs they produced. Others who began or later switched to algebraic representations were successful in producing deductive proofs for a specific odd number of consecutive numbers. In proving the modified implication for an odd number of consecutive numbers, students who used algebraic representations rarely represented the consecutive terms in a more general form of " $x, x + 1, \dots, x + n$ ". This aspect of students' mathematical proving was uncontrolled in experiment since students were not trained to evaluate the representations in the worked-out examples and thus contingent to the task situations and individual experiences. Rather, students were trained to construct a proof or evaluate whether a proof was valid by considering all possible examples and counterexamples expressed in a given type of representations.

Students' use of representations seemed to be generally effective in their Proof Constructions once they had chosen the appropriate representations. They were able to use numerical or algebraic representations to construct Proof-by-example and use algebraic representations to construct Deductive proof. Based on the numerical or algebraic results, they were able to deduce whether the consequent followed from the antecedent. This was likely a factor contributing to the training conditions since the training materials across conditions used

algebraic representations and operations frequently to deduce whether an implication was true or its counterexample or rule violation were mathematically impossible.

Students' modifications of the implication

Upon realizing the implication was false for sets of at least three numbers, students were asked to modify the implications to make it true. The sets of mathematical objects in each modified implication were identified by its content. Table 29 shows the type of modification of the implication made by each student and the modified set of mathematical objects (the detailed modifications made in each attempt is found in Appendix I).

Table 29: Modifications of Implication by Students

ID	Type	Modification
S1	A	Specify three numbers
	A	Specify odd numbers
S2	A	Specify odd numbers
S3	A	Specify three numbers
	A	Specify three or five numbers
	A	Specify odd numbers
S4	A	Specify odd numbers
	M	Specify even numbers; Average is a not a whole number
S5	C	Average is a whole number or a decimal number
	C	Average is a whole number if the sum has a factor decimal number with additional conditions
S6	A	Specify odd numbers
S7	A	Specify odd numbers
S8	A	Specify three numbers
	A	Specify three numbers
	A	Specify odd numbers
S9	A	Specify three numbers
	C	Average may be non-whole numbers
	A	Specify five numbers
	M	Specify even numbers; Average is the middle number
	M	Specify odd numbers; Average is the middle number

Table 29 (cont'd)

S10	A	Specify three numbers
	A	Specify odd numbers
S11	A	Specify odd numbers
S12	A	Specify odd numbers or multiples of 3
	A	Specify odd numbers

During the course of proving their modified implications, some students made as many as five attempts to modify implications. I had only made the request once but the students revised their modifications spontaneously. Apparently, students' responses of making more modifications were motivated by some other factors than the task request.

There were three types of modifications. For type A, modifications were made through insertion of words (e.g., "If the set of three whole numbers are all consecutive in order, then the average is a whole number"), or rephrasing of the antecedent (e.g., "If the set of numbers is an odd number of consecutive numbers, then the average is a whole number" [underlined words indicated the modifications made]). It was apparent that most of the students (N=11) made this type of modifications, i.e., restricting the set of consecutive numbers described by the antecedent, to either sets of three consecutive numbers, five consecutive numbers or an odd number of consecutive numbers, at least once. All students' modified implications, except S5's and S9's, eventually directed the restriction towards the set of odd number of consecutive numbers, which was the maximal set of consecutive numbers for the average to be a whole number.

For Type C, modifications were made through insertion of either words (e.g., "If the set of whole numbers are consecutive in order, then their average is a whole number or a decimal number"), or phrases (e.g., "If there is a set of n consecutive whole numbers, then their average would be a whole number if the sum is [has] a factor of n , or a decimal number if the sum is not [does not have] a factor of n " [underlined words indicated the modifications made]). S5 and S9

were the only two students that made this type of modifications by suggesting alternative outcomes for the average. However, S9 made another modification of Type A which reverted the average to being a whole number and restricted the set of consecutive numbers to five numbers subsequently. In comparison, S5 made another modification of Type C by distinguishing the conditions (namely, whether the sum is a factor of n or not) that led the average to be a whole number or a decimal number. In either case, modifications were made to the consequent to rephrase the outcome of the average again.

For Type M, modifications were made either through insertion of words (e.g., “If the set of whole numbers are consecutive in order and of an even number, like 4 to 6 numbers, then their average is not a whole number”), or replacement of phrases (e.g., “If there are odd number of consecutive whole numbers, then the average is the middle number” [underlined words indicated the modifications made]). These two modifications were last made by S4 and S9 respectively. S9’s final modification about average being the middle number of the set, if proven, would imply that the average is a whole number.

In sum, when students modified a false implication, they tended to modify the sets of mathematical objects described in the antecedent to satisfy the consequent, rather than to introduce another outcome to the consequent or alternative outcomes to both the antecedent and consequent. For students who modified the sets of consecutive numbers, specifying the sets to have an odd number of consecutive numbers was common to, if not eventually, their modifications. Note that this collection is the maximal collection of sets of consecutive numbers for the average to be a whole numbers.

Students considered different sets of mathematical objects in making their modifications to the implication. Next I will explore further about the relation between these two aspects.

Students' consideration of objects and modification of implications

Students considered a wide range of sets of consecutive numbers, either in numerical or algebraic terms, as they modified the given implication. As mentioned earlier, majority of them, except S5, had considered the sets having an odd number of consecutive numbers in one of their modifications either on their first attempt (S2, S4, S6, S7, S11 and S12) or had eventually come to it (S1, S3, S8, S9 and S10). Taking students' prior considerations of mathematical objects into account (Table 25), both groups of students had considered at least three different-sized sets of consecutive numbers by the time they modified the implication to an odd number of consecutive numbers. For example, S6 had considered sets of three, four and five consecutive terms prior to the modification task, and S8 had considered sets of three and four numbers while proving the given implication (Table 25), and three, five, and seven consecutive terms before making a similar modification.

The excerpt of the interview with S6 below (Excerpt 2) highlighted how working with three, four and five consecutive terms helped her generate a tentative modification to the antecedent. S6 interpreted the implication as intended for a set of three numbers but was prompted to think of sets of more than three consecutive numbers.

Excerpt 2: S6's Proof Constructions prior to modification

- I: So what you did is only for three whole numbers.
S6: Mmhmm.
I: But the question says "at least three." "At least three" meaning,
S6: Meaning, what do you mean?
I: "At least three" meaning this set can have four, five, six, seven, eight, nine whole numbers.
S6: Oh, then it goes on then. To calculate I get x , Ok, I do with four now. I just go with the formula, $x + 1$, I see whether I have something in mind but I don't know whether it will

Excerpt 2 (cont'd)

work.

S6: Oh, then it goes on then. To calculate I get x , Ok, I do with four now. I just go with the formula, $x + 1$, I see whether I have something in mind but I don't know whether it will work.

S6: Plus 3, so I will end up getting $4x$ plus 6

I: Mmhmm.

S6: Hmm, now I add it, it's different already because if, to get this average I had to get four. Likewise for five, I get $5x+10$. Ahh, this is another exception. The formula differs. So if I had to have four numbers, this is my formula [points to " $4x + 6$ "].

I: Mmhmm.

S6: And to get the average, I had to divide by four, which means cannot already. So I can't get whole number. But if I were to get $5x+10$, I divide by 5, I would get $x + 2$. so this one must satisfy the equation but the four numbers, they are wrong. So for six numbers it will be $x + 4$, $+ 5$, so $x + 7$. Ah, cannot satisfy also. From what I see is like, when there are even numbers, that means four is, the even numbers of [pointing to "the set of whole numbers"],

I: whole numbers.

S6: Ahh.

I: Ok.

S6: Yeah. Um, I can't satisfy the equation. But when I have a odd number of the set of numbers, I can satisfy the equation.

I: Ok, so if you, so now based on what you have found out about this, would you say this [pointing to the given implication] is true or false again?

S6: False. Because I didn't know it was "at least three" [underline the three words in the question].

As discussed earlier, S6's interpretation of the implication seemed to lack a cognizance for logical verification. When she realized that the set of numbers was not limited to three consecutive terms, S6 continued to construct proofs for four and five consecutive terms using algebraic representations. As she computed the sum, she began to deduce that the implication did not work for four consecutive terms but work again for five consecutive terms. She seemed to have considered six consecutive terms and also began to infer that set having an even number of consecutive terms did not satisfy the consequent. Later she also inferred that sets with an odd number of consecutive terms satisfied the consequent. Consideration of three, four, five (and perhaps six) consecutive terms helped her recognize an emerging mathematical pattern, i.e.,

three, five and seven consecutive terms yielded a whole-number average, which supported her modification of the implication to an odd number of consecutive numbers (Table 29).

S5's modifications seemed to be contradicting the above statement because she had explored sets of three, four and five consecutive numbers but was the only exception that did not make the above modification at all. However, the excerpt of my query with S5 (Excerpt 3), which occurred after she had falsified the implication with a single attempt (Table 27), indicated otherwise.

Excerpt 3: S5's first modification of the implication

- I: Ok, given that you have found some examples that show that this statement is false. Is there any way you change the statement to a true statement?
- S5: True ah?
- I: Yeah.
- S5: [Thinks for a while, points to the consequent statement and say] their average can be a whole number or decimal [laugh]
- I: Their average and be a whole number or decimal [laugh] ok. Why do you want to change it into this way?
- S5: Huh?
- I: Why did you want to change it into this way?
- S5: Because consecutive whole number,
- I: Mmhmm.
- S5: Like you add 1, 2, 3, 4, then you divide by 4. Because 10 might not be, like 4 might not be a factor of 10.
- I: Right,
- S5: So if your numbers add up together is not a factor of 4.
- I: Mmhmm,
- S5: Then you will have a decimal.
- I: I see.
- S5: Then if you add three numbers, it must be a factor of three.

Throughout the interview, S5 meant “multiple” when she said “factor.” While she was modifying the implication, she had only considered the set of three and four consecutive

numbers. Each set led to two opposite properties for the average. Her modification of the implication thus accounted for these two opposite properties observed.

At a macro level, all types of modifications were attempts to classify sets of numbers into supersets that determined whether the average of sets of consecutive numbers is a whole number or a decimal number. Upon identifying multiple sets of consecutive numbers that led to their averages being a whole number, students who recognized these sets as belonging to a single group revised their implications to refer to sets having an odd number of consecutive numbers. Similarly, students who modified their implications to specify the average being a decimal number classified the sets as “of an even number, like 4 to 6 numbers” (S9) or “if the sum is not [does not have] a factor of n ” (S5). What was obvious but worth highlighting was that no modification was made to the “consecutive” description. Students’ considerations of various sets of numbers yielded observations of mathematical properties which students recognized as common to various sets of mathematical objects. This constituted a mathematical pattern which guided students’ processes of modifications.

Summary of findings from the interview

In this chapter we looked at the students’ logical reasoning and Proof Construction in an impromptu task with an intention to identify conjectured explanations for the experimental results and to address the research question of how students modified a false mathematical implication to a true implication based on their examples and counterexamples. The group of students ($N=12$) selected for interview was representative of the logical reasoning and mathematical proving abilities of the study’s sample. They were presented with an implication concerning the average of a set of at least three whole numbers, namely, “If the set of numbers

are consecutive in order, then their average is a whole number.” The given implication is false but can be made true if modified. Students were asked to decide if the implication is true, and then modify the implication to be true. Students spent about nine to 23 minutes on the entire task.

Students’ interpretations of the implication were examined through their considerations of mathematical objects. Half of the students’ considered only three numbers though “at least three numbers” was mentioned and needed to be prompted about other numbers. Some grounded their considerations in some sophisticated heuristics other than logical reasoning. Overall, students’ consideration of mathematical objects suggested that their interpretation of implications lacked logical reasoning for detecting interpretation errors.

Students’ choice of numerical and algebraic representations appeared to have an impact to the types of Deductive proof and Proof-by-counterexample constructed. Their success were somewhat limited by their choices of representations. Algebraic representations and operations supported the students in constructing deductive proofs and generic counterexamples to some extent while numerical representations were limited to falsification of implications by isolated counterexamples. Yet their choice of representations might not change even when they encountered difficulties in Deductive-proof construction. Either the students adhered to numerical representations or algebraic representations but seldom considered more general algebraic representations or alternatives. The experimental study did not factor students’ ability to evaluate their choice of representations into the training materials.

Students used the representations and the operations to derive a logical conclusion from the results also mattered and seemed to interact with their choice of representations. In the situation that they chose an appropriate representation, they were able to construct Proof-by-example and Deductive proof, and were able to deduce whether the consequent followed from

the antecedent. This factor might have contributed to students' improved performance since the training materials across conditions used algebraic representations and operations frequently in drawing mathematical conclusions.

When students were asked to modify the given implication to a true implication, majority of the students restricted the set of mathematical objects described in the antecedent to satisfy the consequent. They modified to a reasonably generic set of mathematical objects after making sufficient exploration of various sets. One student who considered a limited variety of sets modified and further refined the consequent to match the antecedent. Overall, all types of modifications were guided by the mathematical patterns they recognized, as attempts to classify sets of mathematical objects into supersets related to the mathematical patterns. These mathematical patterns described the mathematical properties that students recognized as common to the sets of objects that gave rise to a general mathematical outcome.

Account of possible Contextual factors affecting Students' Performance

The interviews were conducted after school, most of which were on the same day as the posttest. Students might be compelled to complete the interview task as soon as possible to gain some early rest. Certain performance lapses, such as the considerations of only three numbers, or the fixation on the use of numerical representations, might be contingent to the motivational and situational factors.

Consider that the interviews were conducted in a one-to-one setting; the student might feel obliged to do the best they could, out of respect, face or other social factors at play in a Singapore school environment. That might have explained why students made some repeated attempts in proving or modifying the implications.

Nevertheless, the findings from the interview substantiated students' interpretation of statements, their choice and use of representations as the three controlled or uncontrolled factors conjectured for explaining the students' performance in the proving task. In addition, students' recognition of mathematical patterns was found to be also instrumental in students' reasoning of the implications. Given the empirical evidence from interview data, these factors were the most credible causes for explaining the experimental results and understanding students' logical reasoning and mathematical proving.

CHAPTER 6 DISCUSSION AND CONCLUSION

My study examined the role of logic training in students' logical reasoning and proving of mathematical implications, i.e., students inferred why an “If... then...” mathematics statement is true or false deductively. This study inquired how three logic training approaches, one conventional approach emphasizing truth tables and two experimental approaches emphasizing counterexamples and rule violations, might benefit students' logical reasoning and three aspects of mathematical proving. In particular, I examined students' logical reasoning and Proof Construction, Proof Validation and Knowledge of Proof Method (Proof Knowledge) and inquired the extent to which each approach of logic training was effective. Students' reasoning and proving were examined in the context of mathematical implications, which are statements of the form “If P then Q ,” where the statement P is the antecedent and the statement Q is the consequent. Additionally, the study sought to inquire the role of counterexample in enhancing students' logical reasoning and mathematical proving. These inquiries aimed to contribute towards understanding the instructional role of logic training and counterexamples in developing students' mathematical reasoning and proof in the classroom.

In the first part of this chapter, a summary of the study and its findings is presented before the conclusion. Subsequently, the significance of the findings will be discussed and their implications for education and research are drawn. Finally, the limitations of this study will be reviewed and future studies will be proposed.

Summary of the Study and its Findings

In this study, the approaches of emphasizing counterexamples (Condition W) and violations of permissive and obligatory rules (Condition PO) were compared with the approach

of emphasizing truth tables (Condition C, the control condition) in training students' logical reasoning and mathematical proving. The analyses were driven by the hypothesis of the study: logic training emphasizing counterexamples benefits students' reasoning of logical implications as well as their abilities in mathematical proving. Emphasizing counterexamples during logic training increased students' cognizance of counterexamples and thereby enhanced their abilities to determine the logical truth of an implication. Emphasizing rule violations evoked students' reasoning schemas through a real-life context and thereby enhanced their abilities to search for instances that would violate the logical truth of an implication.

To address this hypothesis, a pretest-intervention-posttest experimental design (3 conditions by 2 test trials) with a post-study interview was carried out. Students' written and interview data ($N = 60$) were collected from three Singapore school sites, each over a four-day contact period (including the pretest and posttest sessions). The test instruments assessed students' logical reasoning of implications and the three aspects of their mathematical proving, i.e. Proof Construction, Proof Validation and Proof Knowledge. Proof Construction was further distinguished into Deductive-proof Construction and Proof-by-counterexample Construction. Proof Validation was further distinguished into the Invalidation of Empirical Proof and the Validation of Proof-by-contradiction. Proof Knowledge was further distinguished into the logical non-equivalence of a converse, and the logical equivalence of a contrapositive. These finer aspects were essentially related to students' success in mathematical proving (Alcock & Weber, 2005; Epp, 2003; Healy & Hoyles, 2000; Selden & Selden, 2003).

Students' logical reasoning and mathematical proving was further examined during the post study interview. A representative group of 12 students worked on a new Proof Construction

task. They were asked to decide if the implication given in the task is true with justifications, and then modified the implication to become a mathematically true implication.

The four research questions derived from the main hypothesis inquired the effects of logic training emphasizing counterexamples and rule violations (Condition W and PO), in comparison to the conventional training approach (Control condition), in enhancing the following students' abilities:

Compared to the conventional approach,

- (1) how does logic training emphasizing generation of counterexamples affect students' reasoning with logical implications across different formulations?
- (2) how does logic training emphasizing generation of counterexamples affect students' validation of proofs across different formulations?
- (3) how does logic training emphasizing generation of counterexamples affect students' construction of proofs across different formulations?
- (4) how does logic training emphasizing generation of counterexamples affect students' Knowledge of Proof Methods across different formulations?

In addition, the following research question was also investigated:

- (5) To what extent does students' reasoning of logical implications correlate with their performances in mathematical proving?

These research questions were addressed by the findings generated from the students' pretest and posttest. Apart from explaining the findings from students' tests, the findings gathered from the students' interview data addressed the sixth research question:

- (6) How do students modify a falsifiable mathematical implication to a mathematically true implication based on their self-generated examples and counterexamples?

Experimental Findings from the Pretest and Posttest

The logic training emphasizing counterexamples (Condition W) was found to be significantly more effective in improving students' logical reasoning than the other two training conditions (Condition C and Condition PO). The students were better able to identify the counterexamples satisfying antecedent but not the consequent by which the implication might be falsified.

However, for students' mathematical proving, no significant differences were found for the effectiveness of the training approaches across all three Conditions. Logic training significantly enhanced students' Deductive-proof Construction, Validation of Proof-by-contradiction, and logical non-equivalence of converse and logical equivalence of contrapositive across all Conditions. Though not always successful in providing logically valid responses for these tasks, students in all three Conditions demonstrated more use of deductive inferences, as shown by the distribution of students' scores in each of these tasks. In contrast, there was no improvement in students' Proof-by-counterexample Construction and Invalidation of Empirical Proof in any Condition. Some evidence of students constructing deductive proofs using indirect approach or evaluating proofs based on reasoning of counterexamples was found.

Comparison between students' gain scores in logical reasoning and the various aspects of mathematical proving indicated relatively independent training effects between students' logical reasoning and the finer aspects of mathematical proving. However, to some significantly positive extent, students' constructions of a deductive proof were associated with their rejections of Empirical Proofs, and that their constructions of Proof-by-counterexample for falsifying an implication were associated with their acceptance of Proof-by-contradiction.

Put together, the hypothesis for the better effectiveness of emphasizing counterexamples in logic training was only substantiated for students' logical reasoning but not their mathematical proving; evidence for the better improvements due to the training Condition W was only found in students' reasoning scores. The hypothesis was not substantiated for students' mathematical proving; logic training emphasizing counterexamples seemed to be as effective (or ineffective) as the other approaches. Improvements in some finer aspects of students' mathematical proving were found, namely, Deductive-proof construction, Validation of Proof-by-construction, and both finer aspects of Proof Knowledge related to converse and contrapositive. Moreover, students' logical reasoning of implication was weakly associated with their mathematical proving, though some finer aspects of Proof Construction were moderately associated with Proof Validation.

Findings from the Post-study interview

Theoretical inquiry and empirical evidence from the interview data suggested students' interpretation of implication, their choice and use of representations, and their recognition of mathematical patterns as four most credible factors which were instrumental in students' reasoning and proving of the implications. Students' interpretation of implication refers to their inferences about the sets of mathematical objects meant by the implication. Students' choice of representation refers to the mathematical symbols chosen to represent the interpreted mathematical objects. Students' use of representations refers to how they perform mathematical operations on the representations to make deductive inferences from the result of operation. Students' recognition of mathematical pattern refers to the mathematical properties that students

recognized as common to the sets of objects and that generated a consistent mathematical outcome.

Students' consideration of mathematical objects during their proving attempts suggested that their interpretation of implications were often influenced by their mathematical knowledge. Their inferences of the mathematical objects were often inadequate and needed to be prompted for missing objects. Students' interpretation of implication seemed to lack a deductive process of checking whether the mathematical objects they had considered were logically matched to what were meant by the implication.

Students' choice of representations for the mathematical objects appeared to have an impact on the types of proofs constructed during their proving attempts. Their successes were somewhat limited by their choices of representations. Algebraic representations and operations supported the students in constructing deductive proofs and generic counterexamples to some extent while numerical representations were limited to falsification of implications by isolated counterexamples.

Students' use of representations seemed to be effective to a certain extent. They were able to construct connected chain of deductive inferences and deduce whether the consequent followed from the antecedent from the results of the mathematical operations. Based on the outcome, they were able to make logical conclusions about the implication. As such, they were able to construct Proof-by-counterexamples and deductive proofs. However, their use of representation for constructing mathematical proofs seemed to interact with their choice of representations. In the situation that they chose numerical representations, they were able to construct Proof-by-counterexample but not able to construct deductive proofs. In the situation

that they chose algebraic representations, they were able to construct deductive proofs to a certain extent.

Students' recognition of mathematical pattern seemed to influence their reasoning of implication. When asked to modify a given implication, students considered various sets of mathematical objects interpreted from the implication and determined whether the implication was true. From the mathematical sets that made the implication true, they came up with a mathematical pattern that described the common mathematical properties of these sets and modified the implication to fit that pattern. Overall, all types of modifications were guided by the mathematical patterns they recognized, as attempts to classify sets of mathematical objects into supersets related to the mathematical patterns.

In sum, students' mathematical proving in the interview was influenced by students' interpretation of implication, their choice and use of representations, and their recognition of mathematical patterns. Of these factors, students' interpretation, their choice of representations and recognition of patterns were conjectured as uncontrolled during the experimental study as these were individual students' attributes which were not monitored. Students' use of representations was conjectured as commonly controlled by all training conditions since the worked examples provided across all conditions used algebraic representations and operations frequently in drawing mathematical conclusions. Students might have improved their use of deductive inferences in mathematical proving by learning from these worked-out examples.

For the subsequent sections, I will first discuss the results from both the experimental study and interview findings. Then, I will highlight the limitations of this study and finally, propose future studies.

Discussion

In this section I will discuss the significance of the results reported above. A few issues pertaining to the findings from both the experimental study and the interviews needed to be addressed. First, explanations were needed for students' improved performances in various aspects of mathematical proving which were not significantly different across training conditions. Based on the findings from both the experimental study and the interviews, I will account for the students' improved performance in their Deductive-proof Construction, Validation of Proof-by-contradiction, Proof Knowledge related to converse and contrapositive after training, as well as absence of improvements in their Proof-by-counterexample Construction and Invalidation of Empirical Proof. Second, the suitability of Wason's tasks as indicators of logical reasoning is explored. Lastly, I will examine how the logic training emphasizing counterexamples might be more effective for developing students' logical reasoning of implications.

Effects of Logic Training on Various Aspects of Mathematical Proving

Students showed improved performances in some aspects of mathematical proving across all conditions. Explanations had been conjectured in terms of the factors controlled or uncontrolled by the experiments. A central issue that needs to be addressed is: Why did students exhibited similarly improved performance (or unaffected performance) on the proving tasks across all three Conditions?

I will propose the following explanation for students' mathematical proving performances. Logic training in all conditions had enhanced students' productive use of deductive inferences in mathematical proving. However, their successes were limited by other

factors, which were conjectured to be students' interpretation of mathematical statements and their mathematical knowledge. Of the factors conjectured based on the interview findings, students' choice and use of representations and their recognition of mathematical patterns were both considered as components of their mathematical knowledge. Figure 19 below shows a schematic of these conjectured factors affecting students' mathematical proving performance, namely, Proof Construction, Proof Validation and Proof Knowledge.

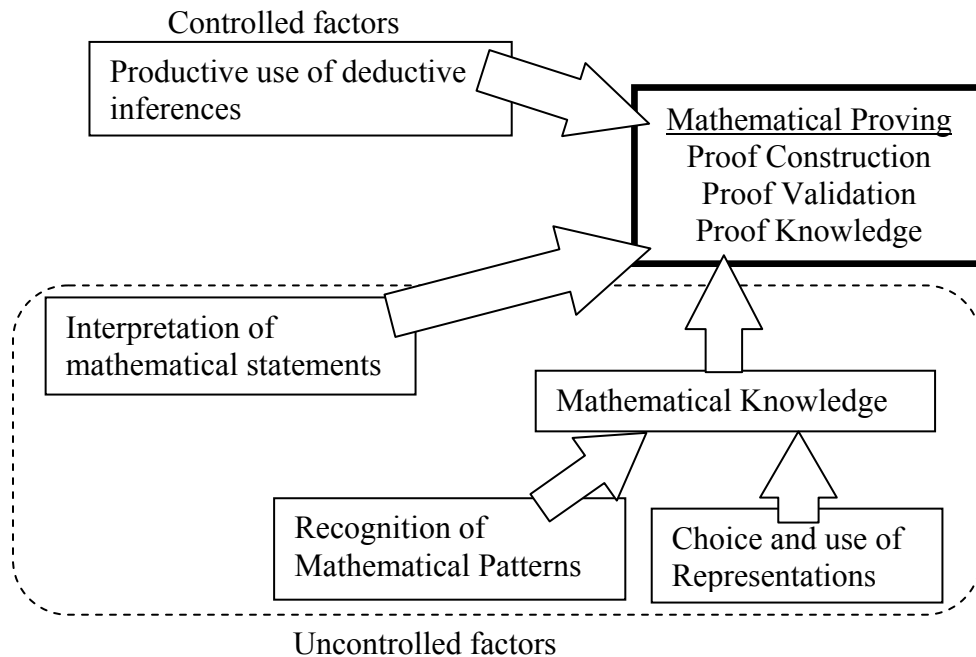


Figure 19: Conjecture Factors affecting Students' Mathematical Proving performance

Next I will argue that this proposed explanation can adequately account for the experimental findings and interview findings of students' mathematical proving.

Effects of Logic training on Productive Use of Deductive Inferences

To show that logic training in all conditions enhanced students' productive use of deductive inferences in their mathematical proving, I will present evidence from the

experimental findings comprising of students' performance in the Deductive –proof Construction, Validation of Proof-by-contradiction, Proof Knowledge related to converse and contrapositive, and evidence from the interview findings comprising of students' proving attempts for the given and the modified implications.

Deductive-proof Construction. The significant improvement found in students' Deductive-proof Construction across conditions reflected increased productive use of deductive inferences according the coding schemes. This was most likely due to the learning effect from the worked-out examples found in all three logic training Conditions. Though the examples emphasized different logical bases in each Condition, these worked-out examples demonstrated the processes of deductive inferences using the same set of representations and operations (usually numerical and algebraic) for proving the same mathematical idea across Conditions. Students' learning from these deductive inferences based on mathematical representations might have enhanced their Deductive-proof Construction to the similar effect across Conditions. However, they were less successful (15 % or less across Conditions) in constructing coherent mathematical proofs for both Deductive-proof Construction items (see Figure 8 and Figure 9).

Validation of Proof-by-contradiction. Students' improved performance in Proof-by-contradiction tasks also substantiated that logic training of all Conditions enhanced students' productive use of deductive inferences. As mentioned previously (see Figure 15, Chapter 4), students improved performance in this aspect were substantially due to an increase in students' responses at level 1, i.e., students evaluated and agreed that the particular statement was indeed a mathematical contradiction. Since the Proof-by-contradiction worked-out the inferential steps that deduced the contradiction statement, students' might have learned to validate the Proof-by-contradiction by checking the inferential steps through applying deductive inferences. This way

of using deductive inferences productively might have been prompted by the deductive inferential processes illustrated in the worked-out examples. In contrast, fewer students went beyond this level of validation to examine how the contradiction would falsify the assumption of the implication is false and led to the conclusion that the implication is true.

Proof Knowledge related to Converse and Contrapositive. Similarly, the improved student performances in determining the logical equivalence or non-equivalence of contrapositive and converse were also plausibly due to the enhancement of students' deductive inferences by logic training across all Conditions. Students' performances in these two aspects were improved only to the extent that students would use deductive inferential processes to check the logical truth of the converse and contrapositive with the implication (see Figure 17 and Figure 18, Chapter 4). Enhancing students' use of deductive inferences in all conditions had generally drawn them away from the illogical reliance on the sentential form to determine logical equivalence, as indicated by the decrease of proportion of students' scoring level 0. Instead, students were more likely to construct proofs for the converse and the contrapositive statements to make such logical decisions.

Students' proving attempts in the interview. Findings from the interview seemed to help explain why students' productive use of deductive inferences was enhanced. Students who used numerical representations were unsuccessful in proving the average of an odd number of consecutive numbers is a whole number. However, some students were able to use algebraic representations and operation to produce a general result for the average. Based on the result, they made deductive inferences about the average to complete their deductive proof. The interview findings concurred with the experimental evidence in supporting that logic training were successful in enhancing students' use of deductive inferences across all Condition, possibly

due to their learning from the worked-out examples. However, their successes also seemed to be limited by the representations they chose, which will be discussed later.

Effects of Students' Interpretation of Implication on Students' Mathematical Proving

Though logic training in all conditions enhanced students' productive use of deductive inferences, their successes in mathematical proving were limited by their interpretations of mathematical statements. To show that, I will present evidence from the experimental findings comprising of students' performance in the Proof-by-counterexample Construction, as well as evidence from the interview findings comprising of students' interpretation of the implications.

Students' Proof-by-counterexample Construction. Students' Proof-by-counterexample Constructions suggested that students interpreted the implication based on their mathematical knowledge, which sometimes contained misconceptions. Yet, students who made such mathematical errors seemed unaware of these errors. They did not incorporate logical reasoning to check their interpretations (e.g., Fig 13, Chapter 4, student made a incorrect conclusion due to her misconception of omitting the prime number '2'). The training materials in all three Conditions did not emphasize checking the interpretation of an implication by logical reasoning. The worked-out examples started with an interpretation of the mathematical objects without devoting much explanation to establish the logical correctness of such interpretations. For example, in one of the worked-out examples involving five consecutive numbers, the algebraic representations " $n, n + 1, n + 2, n + 3, n + 4$ " were immediately presented without a logical procedure of deducing that these representations represented all instances of five consecutive numbers. Instead, logical explanations were devoted to deriving deductive inferences from these

interpreted representations. In this respect, students' successes of proof-by-counterexample seemed to depend on whether their interpretations were logically correct.

Students' Interpretation of Implications in Interview. Interview findings also showed that students' interpretation of mathematical statement play a role in their proving attempts. When students interpreted the set of "at least three whole numbers" in the statement, a number of students considered only three numbers for subsequent inferences. The mathematical error in their interpretation needed to be prompted via external feedback so that they could make a logically correct conclusion.

From the experimental and interview findings, students' considerations of mathematical objects seemed to fall short of a cognizant process of logical reasoning to verify their interpretations. Since this process of mathematical reasoning was not controlled by the experimental design, its effect on students' mathematical proving might have been distributed randomly across all Conditions and undermined their performance, especially on the aspect of Proof-by-counterexample Construction. This perhaps explained why all three logic training approaches did not create significant impact to students' Proof-by-counterexample Construction.

Effects of Mathematical Knowledge on Students' Mathematical Proving

Another factor conjectured to have limited students' successes in mathematical proving was their mathematical knowledge, which comprised of choice and use of representations and recognition of mathematical pattern. Evidence from students' proving attempts during the interview will be used to illustrate students' choice and use of representation. Evidence from the experimental findings comprising of students' performance in the Proof-by-counterexample Construction and Invalidation of Empirical Proof, as well as evidence from their proving

attempts and their modification of implications during the interview will be used to illustrate their recognition of mathematical pattern.

Choice and Use of Representation. During the interview, students who adhered to numerical representations were not successful in producing deductive proofs. Others who used algebraic representations earlier or later were successful in producing deductive proofs for a specific odd number of consecutive numbers. In proving the modified implication for any odd number of consecutive numbers, students who used algebraic representations rarely represented the consecutive terms in a more general form of " $x, x + 1, \dots, x + n$ " that represented *any* odd number of consecutive numbers. As such, they could not come up with a deductive proof for the more general implication. Students' use of representations seemed to be generally effective in their Proof Constructions once they had chosen the appropriate representations. They were able to perform numerical or algebraic operations to support their deductive inferences and Proof Constructions. However, a few students face obstacles in operating on the sum of 1 to n , or made error while operating an algebraic sum of some terms, which affected their success in proving.

Recognition of Mathematical Pattern. Students' performances in Proof-by-counterexample Constructions were also likely due to their recognition of mathematical patterns as justifications for an implication. A counterexample that was not part of the recognized mathematical patterns might not have been identified for Proof-by-counterexample Construction. For example, the only even prime number '2' was omitted while odd prime numbers were considered in Figure 13, Chapter 4. Moreover, the Quadratics items in the pretest and posttest has a mathematical pattern in which the product of two positive whole numbers is always larger than their sum, except when one of the numbers is 1. Omission of this counterexample would affect their performance. Their performances in Invalidation of Empirical Proof were also likely

influenced by the mathematical pattern given in the task. The three examples in the Empirical Proof shared a common mathematical property (odd numbers in Test set 1 or even numbers in Test set 2) and yielded an identical mathematical outcome. At least 30% of the students did not reject the proof after training (see Figure 14, Chapter 4). The well-documented prevalence of students' pattern-based inferences perhaps explained the absence of students' improvements in the Invalidation of Empirical Proofs plausibly (Healy & Hoyles, 2000; Knuth, et al., 2009).

Students' proving attempts during the interview also suggested students' reliance on mathematical patterns as justifications. Some students were able to predict that the factorized form of the average followed a growth pattern as the number of consecutive numbers increased, or that there was a trend of an odd number of consecutive numbers generating a whole-number average. When the representations students used could not support them in making deductive inferences, they used that mathematical pattern as a basis for their justification. In addition, students' modification of implications revealed that they searched for mathematical objects that satisfied an implication by the mathematical patterns they recognized, e.g., an odd number of consecutive numbers.

From the experimental and interview findings, students' choice and use of representations and their recognition considerations of mathematical objects seemed to affect their performance in mathematical proving. Since this process of mathematical reasoning was not controlled by the experimental design, its effect on students' mathematical proving might have been distributed randomly across all Conditions and undermined their performance, especially on the aspects of Proof-by-counterexample Construction and Invalidation of Empirical Proof. This perhaps explained why all three logic training approaches did not create significant impact to students' performances in these two aspects of mathematical proving.

Alternative Explanations

In contest to my proposed explanation, one alternative explanation could be that students applied the demonstrated logical reasoning approaches in the training materials, which was unique to each Condition, to their Proof Constructions. For students in Condition PO and W, this implied that they had to approach the task indirectly by considering the possible counterexamples and rule violations, and reasoning subsequently to a contradiction (e.g., student's work in Fig 4.10 of Chapter 4).

Worthy of note was also the unique outcome of Validation of Proof-by-contradiction pertaining to Condition W (the logic training emphasizing counterexamples). Four students in that condition were able to justify the validity of Proof-by-contradiction based on the rationale that the contradiction had eliminated the possibility of counterexample to the implication. They regarded the proof as an attempt to find a counterexample and the deduced contradiction implied that no such counterexample could be found; elimination of counterexamples proved that the implication is true (see Figure 16, Chapter 4).

However, such instances were infrequent to build up a systematic effect to students' performances due to training conditions. To infer or generate counterexamples seemed less intuitive (e.g., think of numbers which are not whole numbers); most students constructed proofs using a direct approach, i.e., started deriving inferences from the antecedent that lead to the consequent.

Why Students exhibited Similar Performance on the Proving tasks across Conditions

Across the students' performances in each of the aspects of mathematical proving discussed above, a consistent trend seemed to emerge to account for the improved performance (or unaffected performance) in various aspects of mathematical proving found across conditions. For the aspects in which students' improved performance were found, they were using more deductive inferences productively after the logic training. For the Deductive-proof Construction, students showed more productive use of deductive inferences but the successes of constructing a coherent proof were subjected to their choice of representations. For the Validation of Proof-by-contradiction, students also showed substantial application of deductive inferences to evaluate and agree with the contradiction statement but lesser students went beyond to examine how that contradiction leads to the implication being proved. For the Proof Knowledge concerning converse and contrapositive, students were drawn to construct proofs for the converse and contrapositive statements, which indicated more use of deductive inferences, to compare their truth values with the implication.

For the aspects in which students' performance did not improve significantly, their use of deductive inferences seemed to be undermined by their interpretations of the implication and their choice and use of representations. For the Proof-by-counterexample Construction, students' performances were conjectured to depend on their interpretation of implications and their mathematical knowledge and an apparent lack of logical check. In addition, students' reliance on mathematical patterns might have undermined their search for counterexamples. For the Invalidiation of Empirical Proof, students' reliance on mathematical patterns when they interpreted the implication was conjectured to have undermined students' use of deductive inferences.

I have proposed that logic training in all conditions enhanced students' productive use of deductive inferences in their mathematical proving. In addition, I conjectured that their successes were limited by students' interpretation of mathematical statements and their mathematical knowledge. This explanation seemed to account for students' performance in mathematical proving adequately. In particular, they were better able to use deductive inferences productively to construct deductive proofs or indirect proofs that eliminate counterexamples, which explained their improved performances in some aspects of mathematical proving. However, students' interpretation of implication and their mathematical knowledge, including mathematical patterns and choice of representations, seemed to affect their performance in the proving tasks as well.

Wason's Task as a Logical Reasoning Indicator

Students in Condition W had shown improved performance in logical reasoning over the others, which was assessed using Wason's selection tasks, but not in Proof Construction, Proof Validation and Proof Knowledge. One immediate question arises: Why was students' performance in logical reasoning significantly better in Condition W than Conditions C and PO but their performance in mathematical performance was similar across Conditions?

One plausible explanation was that Wason's task had low sensitivity to the improvements achieved by students in logical reasoning. The difficulty level of the task might have generated false-negative assessments of their deductive reasoning (Evans & Over, 1996; Stenning & Lambalgen, 2004). In order to perform well in the selection task, students needed to expertise in logical reasoning. Logic training might have helped students make some improvements in logical reasoning of varying magnitude across all conditions. Some were large enough to enable them to perform well in the Wason's task while others made smaller improvements which were

insufficient to perform well in the Wason's task but met the threshold level needed for improving their productive use of deductive inferences.

Moreover, logical reasoning is likely one of the many contributing factors to students' mathematical proving. As my proposed explanation of the experimental outcomes and interview findings suggested, students' successes in mathematical proving were capped by other factors such as their interpretations of statements and mathematical knowledge. Therefore, higher achievements in logical reasoning were unlikely able to make up for the shortfall of other conjectured contributing factors to students' successes in mathematical proving.

An alternative explanation was that performance in Wason's task was simply irrelevant to students' mathematical proving abilities, as suggested by the study's findings on the correlation between students' logical reasoning performance and their various aspects of mathematical proving (see Table 18, Chapter 4). However, this explanation did not square with the experimental findings that students in Condition W improved significantly both on the Wason's tasks and mathematical proving tasks, just not significantly better in mathematical proving tasks than the students in the other Conditions. Furthermore, empirical evidence existed for the correlation between expertise in mathematical proving and performance in Wason's tasks. Undergraduates and mathematicians with expertise in mathematical proving were found to performed better in Wason's task than their non-mathematics-trained counterparts (Inglis & Simpson, 2004; Jackson & Griggs, 1988). The small and insignificant correlation between students' scores in Wason's tasks and proving-related tasks was perhaps due to the false negativity of the Wason's tasks. Furthermore, the association might be weakened by pooling the sample of Condition W exhibiting significant results with the other samples exhibiting

insignificant results from the other two Conditions. Hence the evidence from this study and other related studies weighed against the alternative explanation of irrelevance.

Put together, students trained in Conditions C and PO might have developed some logical reasoning abilities after completing the logic training that helped their mathematical proving performances. The improvements were perhaps not as robust as their counterparts in Condition W to exhibit significant improved performance in Wason's tasks, due to false negativity of the Wason's task. However, high performance in logic reasoning was perhaps insufficient for successful mathematical proving. As our experimental outcomes and interview findings suggested, students' successes in mathematical proving were capped by other factors such as their logical interpretation and mathematical representations of the implication.

Logic Training Emphasizing Counterexamples for Students' logical reasoning

The training approach emphasizing counterexamples significantly improved students' logical reasoning than the approach emphasizing rule violations. A plausible explanation is that the training approach emphasizing counterexamples prompted students to think about how counterexamples, being instances that do not satisfy an implication, affect the logical truth of an implication. As argued similarly, explication of the counterexamples enhanced students' reasoning about the truth of an implication (R. Platt & Griggs, 1993). When students were actively searching for a possible counterexample in their interpretation of the implication, they were likely to make logically valid conclusions about an implication. The approach emphasizing rule violations did not produce a significant effect possibly because evoking reasoning schemas associated to rule violations were more sensitive to real-life context than abstract context (Cheng, et al., 1986). Since the mathematical context portrayed in the implications was more similar to an

abstract context than to a real-life context, the training approach emphasizing counterexample was more effective than the approach emphasizing rule violations.

Cheng et al.'s (1986) found that training emphasizing violations of obligatory rules were equally effective as training emphasizing counterexamples in students' logical reasoning of implications. Based on this finding, Stylianides & Stylianides (2008) advocated the training approaches of emphasizing rule violation for developing students' logical reasoning in mathematics. Lawson (1990) had also shown that logic training emphasizing counterexamples were effective for scientific reasoning.

However, our findings about students' logical reasoning yielded contrary empirical evidence. Logical reasoning emphasizing counterexamples was significantly more effective for students' logical reasoning of mathematical implication than emphasizing rule violations and truth tables, as evidenced by the experimental outcomes of students' logical reasoning performance. Unlike the implications used in Cheng et al.'s (1986), the majority of the training items used involved mathematical content, which were more abstract than real-life contexts. Thus, evoking schemas of permission and obligation for reasoning about violation of implications might become less sensitive in mathematical context and hence less effective than emphasizing counterexamples for enhancing logical reasoning of mathematical implications.

Implications for Education and Research

The discussion and the findings of this study drive a few implications with regards to the role of logic training and the use of counterexamples in students' logical reasoning and mathematical proving. I will explore the educational implications with regards to what and how the logic training and the counterexamples can be used for supporting students' logical reasoning

and mathematical proving. I will then explore the research implications with regards to the issues concerning the relationship between students' mathematical knowledge, logical reasoning and mathematical proving.

The Role of Logic Training in Teaching and Learning of Mathematics

The results of the study suggested that logic training has its place in developing students' logic reasoning and mathematical proving. The approaches of logic training used in this study, be it based on logic truth tables, emphasizing rule violations or emphasizing counterexamples, seemed equally helpful in enhancing students' deductive inferential processes during construction of deductive proofs to justify mathematical implications, evaluation of the logical equivalence between similar mathematical statements (converse and contrapositive), and reasoning by contradiction.

As argued in other studies (Epp, 2003; Mueller, 1975; J. L. Platt, 1967; A. Stylianides & Stylianides, 2009b), logic training seemed to provide students the logical foundation of proving mathematical statements through enhancing students' productive use of deductive inferences. Contrary to Stylianides & Stylianides' (2008) recommendation of using selection task emphasizing rule violations in real-life contexts, this study recommended the use of Wason's version of selection task as part of the logic training for enhancing logical reasoning; emphasizing counterexamples improves students' logical reasoning of "If...then..." in mathematical context.

Mathematics teaching and learning should consider logical training in mathematics classroom, be it based on truth tables or other approaches such as the use of Wason's tasks. Many mathematical concepts and relationships can be stated in as mathematical implications.

Since logic training enhanced students' productive use of deductive inferences, it would support students' mathematical justifications of these concepts and relationships in class (Ball & Bass, 2003; Simon, 2000; Simon & Blume, 1996).

The Role of Counterexamples in Mathematical Reasoning and Proving

The results of the study also suggested that using counterexamples is also an effective approach to enhance students' logical reasoning and mathematical proving. Representing the counterexamples and deducing the possibility of counterexamples supported students' construction of mathematical proofs in the same way as students who were trained using the same approach of truth tables and proof demonstration. In some instances, students picked up the indirect approach of deducing the implication by the falsification of counterexample to the implication, which added on to their proving strategies.

Another role for the use of counterexample might be in supporting students' understanding of Proof-by-contradiction. Students often had trouble comprehending the Proof-by-contradiction because the proof assumed the negation of implication is true and constructed contradiction instead of proving why the implication is true directly (Antonini & Mariotti, 2008). They could not comprehend the rationale of proving an implication by proving its negation to be false. The use of counterexamples might help explain that why Proof-by-contradiction is mathematically meaningful. By reframing the Proof-by-contradiction as a process of searching for counterexamples to the implication, the students might be able to appreciate the rationale of assuming the implication is false and that the deduced contradiction in the Proof-by-contradiction meant that such search is mathematical impossible (similar to the student's validation provided in Figure 16 of Chapter 4).

Setting Realistic Expectation of Logic Training

The findings of this study also painted a clearer picture of what teachers and students could expect of logic training in mathematics classroom, if the same sort of training were implemented in classrooms. Students' awareness to deductive inferences might have been heightened. They might have understood the relation between the truth of an implication and the counterexample and be able to construct deductive proofs. However, their performances were capped by their prior mathematical knowledge, specifically, their interpretation and representation of the implication. In other words, based on the findings of this study, students' better logical reasoning is useful but perhaps inadequate for facilitating improved mathematical Proof Constructions. Conversely, attributing students' low success in mathematical Proof Constructions to lack of logical reasoning competence might seem a quick jump to conclusion; individual factors such as prior knowledge are at play.

Note that one of the training approaches was based on a conventional approach used in university courses. It first introduced the logical truth tables, the related laws of logical deduction, and subsequently, demonstration of its applications to mathematical proving. Hence the above argument also applies to university classrooms: logic training based on the conventional approaches is necessary but perhaps inadequate for facilitating competence in students' Proof Construction. Lack of logical reasoning competence is one of the many factors at play.

Likewise, students' better logical reasoning need not necessarily translate into better Proof Validation and Proof Knowledge. This study's outcomes shared some striking resemblance with other long-term studies of middle-school students' mathematical proving. For example, a

high proportion of students' still regarded the converse of an implication as logically equivalent to the original proposition (Healy & Hoyles, 2000). A lot of students also accepted empirical proof as valid (Healy & Hoyles, 2000; Knuth, et al., 2009). It seemed that students experienced difficulties in separating truth from validity (Durand-Guerrier, 2008) and logic training of the sort seen in this study is inadequate in address this issue. Additional instructions are needed for resolving this difficulty.

Rethinking about Instructions of Logical Reasoning for Mathematics Classrooms

Perhaps a bigger point and a strong critique of the current state of logic training (including the study's training approaches) is this: An instructional theory of logical reasoning for mathematics in classroom is much needed to structure the mathematical activities and to attain the instructional goals pertaining to mathematical reasoning and proofs. At the minimum, the theory should map out the various cognitive components of students' logical reasoning process and suggest the pedagogical activities most likely effective for enhancing a particular components. Pedagogical recommendations made by Epp (2003) and Stylianides & Stylianides concerning the use of selection tasks and concrete examples are specific parts of the instruction theory for enhancing students' use of logical principles in mathematics. Additional components of the instructional theory include students' interpretation of implication and their mathematics knowledge as other important components, as conjectured by this study.

Much of the focus in the instructions on logical reasoning had been placed on the rules of inferences often illustrated by logical truth tables and supplemented by informal or real-life applications (Epp, 1994; Mueller, 1975). At the same time, it is hope that such instructions could develop students' independent ability to justify their mathematical conceptions and deepen

mathematical understanding through logical reasoning and proving (Ball & Bass, 2003; Simon & Blume, 1996). However, this study hinted at some gaps that required educators' attentions when thinking about logical reasoning in mathematics.

The first gap concerned students' ability to interpret mathematical statements logically without over generalizing or under generalizing. As illustrated by the interview findings, if the students were unaware of his/her misinterpretation of a implication (e.g., interpreted "at least three" to be "three" or "prime numbers" as "odd prime numbers" in this study), they would be unlikely to be able to rectify their reasoning error, which was not due to errors made in the process of proving but rather at the interpretation stage. Developing students' abilities to scrutinize their interpretation under logical reasoning before making subsequent deductive inferences seemed crucial.

Related to the interpretation issue, the second gap concerned students' prior knowledge of mathematical patterns (e.g., recognized that the pattern that prime numbers were all odd numbers except for '2'). Students were prone to generate mathematical sets based on the mathematical patterns and to regard that as convincing grounds of justifications (Coe & Ruthven, 1994; Knuth, et al., 2009; Recio & Godino, 2001; A. Stylianides & Stylianides, 2009b). Given the pattern-based thinking is an inevitable part of human inferences, harnessing this thinking process to support logical reasoning of their interpretations seemed a sensible instructional consideration. One positive illustration would be students generalize the pattern they observed and used it to construct a deductive mathematical proof (Pedemonte & Buchbinder, 2011).

In sum, an instructional theory of logical reasoning for mathematics needs to address the various cognitive components of students' logical reasoning. While the often focused component is students' logical reasoning of the interpreted mathematical statement, other often neglected

components needed to be identified. Enhancing students' logical reasoning in terms of these components is central to development of their mathematical reasoning and proving.

Role of Mathematical Knowledge in Mathematical Reasoning and Proving

Interview findings showed that students' successes of mathematical Proof Constructions were limited by their mathematical knowledge. Students' choices of representations, which were part of their mathematical knowledge, mattered as much as their logical reasoning, perhaps so because the representations were able to classify many instances mathematically into manageable finite number of possibilities for logical reasoning, e.g., three consecutive numbers expressed as x , $x + 1$, $x + 2$. Students who were able to harness the mathematical usefulness of the representations were able to construct deductive proofs. Students' who had no access to more useful representations most likely resorted to observed patterns as their mathematical justifications (Harel & Sowder, 1998; Knuth, et al., 2009). In supporting students' proving, mathematics instructions in classroom need to highlight the advantages of different representations in reducing infinitely many instances into finitely many sets. Comparisons of the use of at least two different representations of a few implications might be a source for classroom discussions to highlight the usefulness of each representation for each implication.

Research implications

The results of the study suggested that some additional conjectured factors should be examined with regards to students' logical reasoning and mathematical proving, namely, students' interpretation of the implication and students' choice of representations. Pertaining to students' interpretation, one issue concerns how logic training can enhance students' logical

reasoning of their interpretation worth research attentions. Students' logical interpretations or detection of their interpretation errors are two possible aspects worth researching. Pertaining to students' choice of representations, one issue concerns how their mathematical proving performance can be enhanced by making more strategic choices or developing greater flexibility in representations. More specifically, research is needed to inquire whether equipping students with more knowledge about representations and strategies of choosing representations will increase their chances of constructing coherent mathematical proofs.

At a macro level, the research issue pertaining to students' logical reasoning and mathematical proving needed to be widened to include inquiries about the role of students' prior mathematical knowledge. There are different aspects of students' prior knowledge conjectured as relevant by this study, one of which being students' knowledge of mathematical representations and the other being students' recognition of mathematical patterns. However, there might be other aspects that remained to be identified. Inquiries about how certain aspects of students' prior mathematical knowledge affect students' logical reasoning and mathematical proving, and identification of others, are likely to be important areas of research. Coupled with the consideration of logic training for students' interpretation of statements and training for productive use of deductive inferences, the research of mathematical reasoning and proving would be more enriched.

Limitations and Future Studies

Finally, a few words about the limitations of the study are mandated before ending the chapter.

Specificity of Singapore Students

Singapore students participated in this study. From recent TIMSS (Mullis, Martin, & Foy, 2008) studies, Singapore students were found to performed well above other countries in mathematics. They were also highly motivated in terms of their longer hours of engagement in homework, as compared to their counterparts in other countries. Moreover, they were also responsive to the tasks requested of them in this study. These attributes certainly increased the feasibility of using self-paced learning in the study. For students of other countries with lower motivation and mathematical knowledge, the challenge of conducting self-paced learning increased.

Nevertheless, Singapore students also exhibited similar barriers in their mathematical reasoning and proving. Like students in U.K. and U.S., they were not competent in invalidating Empirical proofs and recognizing that the implication and its converse were logically non-equivalent (Hoyles & Küchemann, 2003; Knuth, et al., 2009). Follow-up studies involving other countries students, such as U.S. and U.K. students, might turn up similar or different findings about the effect of logic training on their logical reasoning and mathematical proving.

Sample size and Significance of findings

The number of students in each condition was set at 20, based on an estimation of the statistical power. Significant improvements with large effect sizes were found in students' logical reasoning from the experimental findings. However, unlike other studies which involved numbers of students in the order of magnitude of hundreds, the sample size of this study may not be large enough to provide a stronger substantiation to the conclusions derived for the limited

benefits of logic training; statistical power of the inferred students' non-significant improvements might be limited by the sample size.

More tasks needed for reliability

First, the lower-than-convention reliability coefficient of the test instrument may be a concern. Items grouped under the same assessment component might not be measuring the same attribute. Though less satisfactory, the analysis for the two components of Proof Construction (deductive-proof and proof-by-counterexample) got around this issue by a principal factor analysis of the principal component scores. But a relatively low reliability of proof-by-counterexample construction (pretest: 0.420, posttest: 0.360) obscured the significance of the results. The reliabilities were low possibly because the items were of very different levels of difficulties and additional items of intermediate levels of difficulties should be introduced. Second, components of the proof validation and proof knowledge were assessed by single item in the tests. More isomorphic items need to be introduced for further inquiries involving these two aspects of mathematical proving.

Conviction and Validation: Personal vs. Social

There are definitely uncontrolled factors on students' performance, which might be related to social issues, such as influence of external knowledge authorities (Segal, 1999; Simon & Blume, 1996). In this study, the social effects on their mathematical proving were not monitored. A couple of studies, however, had demonstrated that students' choice of acceptable proofs might vary depending on whether criteria leaned more towards personal or social effects (Healy & Hoyles, 2000; Segal, 1999; A. Stylianides & Stylianides, 2009b). Students may find

empirical proofs more acceptable than formal proofs for reasons of personal conviction.

However, if the proof were to gain teachers' approval and validation, students might reverse their choices. In this study, I had stressed personal conviction over gaining my approval repeatedly to minimize such effects.

Coding scheme for Proof Constructions

The coding scheme for Deductive-proof Construction aimed to evaluate students' use of deductive inferences in varying degrees. As such, there were other aspects of students' proof which were not captured. For example, the indirect approach shown in Figure 10 of Chapter 4 was rather innovative and exhibited a certain sophisticated level of mathematical thought, as feedback by a mathematics teacher. Yet the mathematical quality was not reflected but rather classified as a proof with inadequate logical reasoning.

Another limitation of the coding scheme was to capture the so-called generic examples which had been championed for its instructional potential by mathematics educators (Balacheff, 1988; Mason & Pimm, 1984; Zazkis & Chernoff, 2008). Generic examples were proofs using specific examples to model a generic deductive process that have proven the mathematical proposition. The specific symbols were meant as a representative placeholder for a class of mathematical objects. Part of the challenge was to identify from the written data that a particular example was used by a student to as a means to point to generic mathematical properties. Students' written responses were not always elaborative and clearly tractable. To infer that generic mathematical properties were being referred to by the student might result in reading too much into the data. As such, the category of generic examples was dropped and replaced by other levels whose criteria were more grounded in text. Nevertheless, if one were to pinpoint the

occurrence of generic examples, students' proofs being coded as level 2 would be the most promising group. To reliably identify the generic examples, interview data of students' thinking behind these specific examples is required.

Training tasks

The tasks used in this study were, of course, not encompassing the entire possible range of mathematical tasks. In this study, the training tasks used only arithmetic, algebraic, Quadratics and Elementary Number Theory contents to pose the mathematical situations and implications. Geometry contents were left out due to practical constraints of carrying out the study. As such, we do not know whether the conclusions of this study can be extended to the case of geometric proving.

Furthermore, the logical reasoning and mathematical proving were inquired using the implication statement as a platform. Although almost all mathematical propositions can be logically translated to implications, we do not know the effects of logic training would be similar if the study was carried out for other logical forms of mathematical statements, e.g., the syllogistic form of "All A are B " were used instead. Research studies in cognitive science had found difference of students' logical reasoning between the implication and the syllogistic statements (Leighton, 2006).

Laboratory-based Instructions and Classroom Instructions

An issue to note is the different nature of laboratory-based instruction and classroom instructions. The laboratory-based instructions used in this study came in the form of self-paced problem solving with interactions between the students being controlled and discouraged.

Interactions between the students and the “teacher” (the researcher) were also minimal. This mode of instructions is quite distant from instructions in mathematics classrooms of Singapore and in U.S. at large. Granted that some classroom moments in Singapore (and perhaps U.S.) do encourage students’ independent learning through problem solving, many common features of day-to-day classroom instructions in Singapore and U.S. classrooms were excluded from the laboratory instructions. The most glaring absence was the small group or whole class discussions about the mathematical content at hand. Absent were also teachers’ guidance provided to students in understanding the contents and elaborations to unpack the content to a level within students’ reach. These limitations needed to be taken into consideration when one considers the findings in this study for classroom uses.

Maintenance of Training Effects and Latent Effects

Consider that the posttest was administered the next day or one more day after the training sessions, the improvements students exhibited in their logical reasoning and mathematical proving might be short-lived. Alternatively, their performance might hold out over a longer period for some or all training approaches. Unfortunately, this study did not administer a retention assessment to monitor students’ improvement beyond the posttest. In addition, this study did not monitor any latent effects due to different training approaches. Students’ learning from the training materials might require a longer period of time to take effect and lead to significant differences across conditions. Introducing a retention test two weeks after the posttest would provide further information about the effects (non-effects) of logic training, e.g., indicate whether students’ improvement was rooted in surface features or deep structure problem solving processes (Chi, Feltovich, & Glaser, 1981).

Future studies

Further inquiries may be directed towards other factors that influence students' success in these tasks related to mathematical proving, given the similar effects (or absence of effects) observed across all conditions. Students' increased use of deductive inferences did not always translate to successful attempts in various tasks and might still produce mathematically flawed proofs or justifications. Studies about the role of students' prior mathematical knowledge might inform the relationship between students' logical reasoning and their mathematical knowledge in successful mathematical proving. Another series of promising study might be the investigation of logic training that improves students' logical interpretation of the implication as well as logical derivation of conclusions from the interpretation. Comparison of students trained for logical interpretations with students who are not might reveal the importance of more comprehensive logic training and inform classroom instructions. In carrying out these proposed studies, limitations listed above should be taken into account to improve their explanatory power.

Conclusion

Logic training emphasizing counterexamples played a more effective role in improving students' logical reasoning. Logic training also improved students' mathematical proving through enhancing their productive use of deductive inferences. However, students' successes in mathematical proving were conjectured to be limited by their interpretation of mathematical statements and mathematical knowledge, which includes recognition of mathematical patterns, and choice and use of mathematical representations.

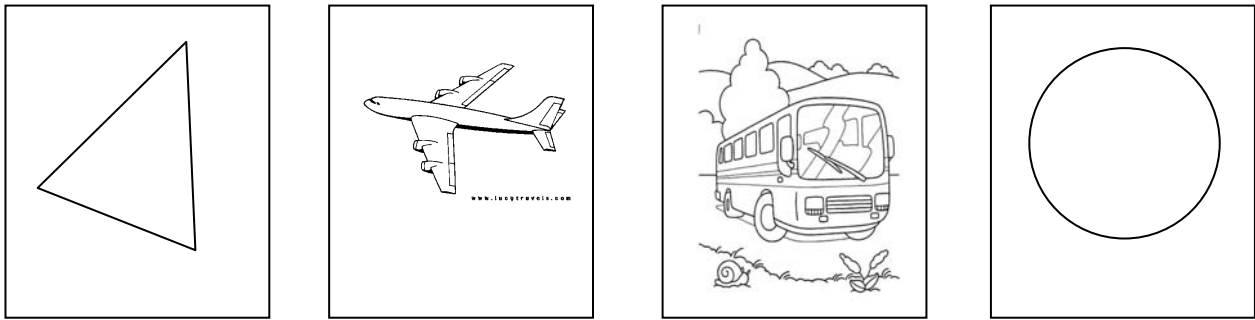
APPENDICES

Appendix A: Test Set 1

1. Below is depicted a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is a shape on one of its sides and a picture of a transport vehicle on the other.

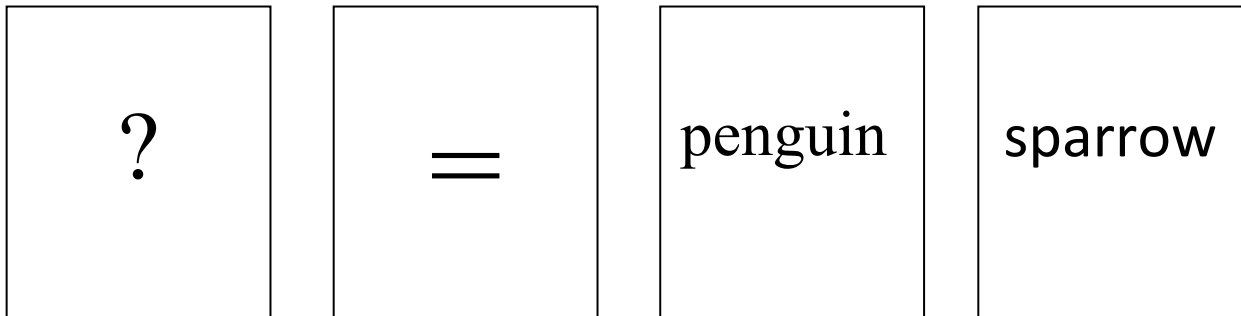
Also below there is a rule which applies only to the four cards. Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the card(s) you want to turn.

Rule: *If there is a shape with straight edges on one side, then there is a land transport vehicle on the other side.*



-
2. Below is depicted a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is a bird on one of its sides and a symbol on the other. Also below there is a rule which applies only to the four cards. Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the card(s) you want to turn.

Rule: *If there is a flying bird on one side, then there is a punctuation mark on the other side.*



3. Below is depicted a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is a piece of cloth on one of its sides and a stamp on the other. Also below there is a rule which applies only to the four cards. Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the card(s) you want to turn.

Rule: *If there is a triangular stamp on one side, then there is a piece of red cloth on the other side.*

Triangular stamp	Circular stamp	Red cloth	Blue cloth
---------------------	-------------------	-----------	------------

4. Below is depicted a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is the year of building a house on one of its sides and the construction material used to build the house on the other. Also below there is a rule which applies only to the four cards. Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the card(s) you want to turn.

Rule: *If there is a house built before 1969 on one side, then there is a house built by timber on the other side.*

House B is built by bricks	House C is built in 1981	House A is built in 1960	House D is built by timber
----------------------------------	--------------------------------	--------------------------------	----------------------------------

5. Three times a whole number x is added to the square of x . For example, 3 times 2 is added to the square of 2, or, $3 \times (2) + 2^2$. Decide whether the following rule is true or false:

If $3x$ is added to x^2 , then the sum is an even number.

Justify why your conclusion must be true or false using the most convincing argument.

6. A prime number is a whole number that has exactly two factors, 1 and the number itself (Note that 1 is not a prime number since it has only one factor). Two positive numbers, which may or may not be prime numbers, are multiplied together. Decide whether the following rule is true or false:

If two prime numbers are multiplied together, then the product is an odd number.

Justify why your conclusion must be true or false using the most convincing argument.

7. The graph of the quadratic equation $y = ax^2 + bx + c$ looks like either one of the graphs below, depending on whether a is greater or less than 0. The x -intercepts are the values where the graph 'cuts' the x -axis. (The y -axis is not shown in the graphs because its position depends on the x -intercepts).

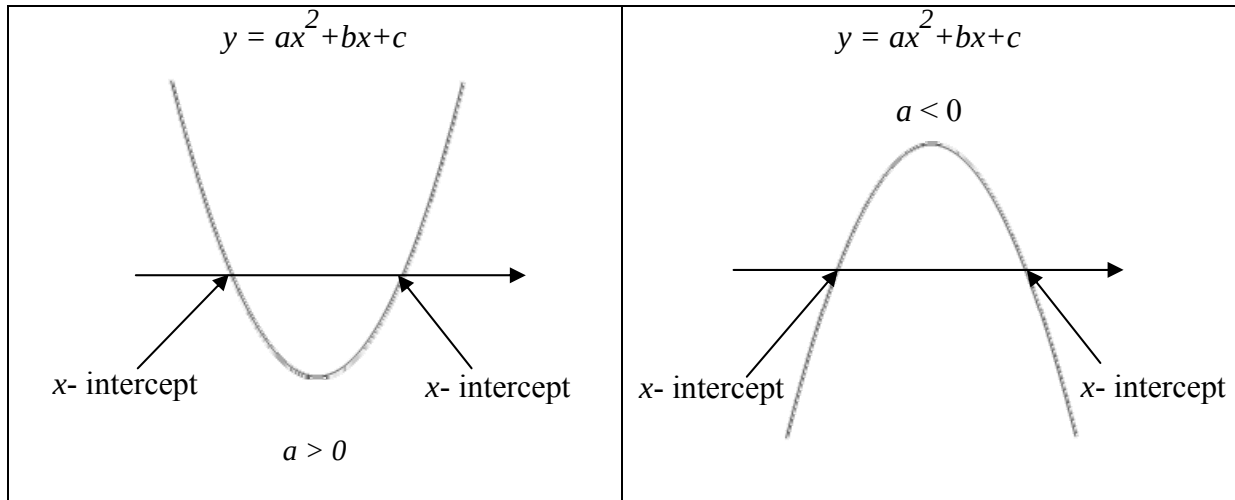


Figure 20: Graph of Quadratic Equation

Here, c is the y -intercept of the graph (where the graph 'cuts' the y -axis) because when $x = 0$, $y = c$.

Decide whether the following rule is true or false:

If a is positive and c is negative, then the x -intercepts of the graphs are one positive number and one negative number.

Justify why your conclusion must be true or false using the most convincing argument.

8. The quadratic expression $x^2 + Mx + N$, where M and N are positive whole numbers, may be factorized into the form $(x + a)(x + b)$, where a and b are also positive whole numbers. For example, $x^2 + 7x + 12$ can be factorized into $(x + 3)(x + 4)$ but not $x^2 + 7x + 11$. Decide whether the following rule is true or false:

If $x^2 + Mx + N$ can be factorized, then $M < N + 1$.

Justify why your conclusion must be true or false using the most convincing argument.

9. A series of numbers are generated by inserting even numbers, starting from 2, into the expression $n^2 + 7n + 7$. Each result is checked to see if it is a composite number, that is, it is divisible by other numbers on top of 1 and itself. The first three instances are shown below.

Take $n = 2$, $n^2 + 7n + 7 = (2)^2 + 7 \times 2 + 7 = 25$; $25 = 5 \times 5$ is a composite number

Take $n = 4$, $n^2 + 7n + 7 = (4)^2 + 7 \times 4 + 7 = 51$; $51 = 17 \times 3$ is a composite number

Take $n = 6$, $n^2 + 7n + 7 = (6)^2 + 7 \times 6 + 7 = 85$; $85 = 17 \times 5$ is a composite number

A mathematical statement is proposed:

“If n is an even number, then $n^2 + 7n + 7$ is a composite number.”

Can you conclude that the mathematical statement is true because of the three instances above? Why or why not?

10. Joe and Eva are thinking about pairs of whole numbers (which may be different) in their mind.

Joe says: *If the product of two whole numbers is even, then their sum is odd.*

Eva says: *If the sum of two whole numbers is odd, then their product is even.*

Are Joe's and Eva's statements saying the same mathematical idea? Please provide justifications for your answer.

11. Consider the mathematical statement “Let N be an integer. If N^2 is an even number, then N is an even number.”

Kathy says that the truth of the above statement is the same as the truth of this statement:

“Let N be an integer. If N is an odd integer, then N^2 is an odd number.”

Based on what you know about odd and even numbers in general, do you agree with Kathy’s conclusion? Please provide justifications for your answer.

12. Consider the mathematical statement “Let x and n be two real numbers. If $x > 0$ and $n > 0$, then $\frac{x}{n} + \frac{n}{x} \geq 2$.”

To determine whether the statement is true, Gabriel reasons in the following way:

Suppose there exists a pair of real numbers a and b such that both are greater than 0 but $\frac{a}{b} + \frac{b}{a} < 2$. Let's simplify this inequality using algebraic operations.

Combining the fractions, $\frac{a^2 + b^2}{ab} < 2$

Multiply ab on both sides, $a^2 + b^2 < 2ab$

Subtracting $2ab$ from both sides, the inequality becomes $a^2 + b^2 - 2ab < 0$

Factorizing the left hand side, the inequality becomes $(a - b)^2 < 0$

But $(a - b)^2 < 0$ cannot be satisfied by any value of a and b because the square of any real number is either 0 or positive.

So the supposition is false. This tells us that the statement “If $x > 0$ and $n > 0$, then

$\frac{x}{n} + \frac{n}{x} \geq 2$ ” is true.

Based on what you know about inequalities in general, do you agree with Gabriel's way of making his conclusion? Please provide justifications for your answer.

Appendix B: Summary of Isomorphic Items in Test Sets 1 and 2

Table 30: Isomorphic implications in Test Set 1 and 2

Item	Implications in Set 1	Implications in Set 2	Nature of item	Topic
1	<i>If there is a shape with straight edges on one side, then there is a land transport vehicle on the other side.</i>	<i>If there is a sea animal on one side, then there is a straight line on the other side</i>	Logical Reasoning of Implications	Non-math
2	<i>If there is a flying bird on one side, then there is a punctuation mark on the other side</i>	<i>If there is a soft drink on one side, then there is a fast food restaurant on the other side.</i>		
3	<i>If there is a triangular stamp on one side, then there is a piece of red cloth on the other side.</i>	<i>If there is a compass pointing to the North on one side, then there is a coin showing heads on the other side.</i>		
4	<i>If there is a house built before 1969 on one side, then there is a house built by timber on the other side.</i>	<i>If there is a car turning left on one side, then there is a red traffic light on the other side.</i>		
5	<i>[x is a whole number] If $3x$ is added to x^2, then the sum is an even number.</i>	<i>If a whole number x is added to its square, x^2, then the result is an even number.</i>	Deductive-proof Construction	Elementary Number Theory
6	<i>If two prime numbers are multiplied together, then the product is an odd number.</i>	<i>If two prime numbers are added together, then the sum is an even number.</i>	Proof-by-counterexample Construction	Elementary Number Theory
7*	<i>[For graphs of $y=ax^2+bx+c$] If a is positive and c is negative, then the x-intercepts of the graphs are one positive number and one negative number.</i>	<i>If $x^2 + Mx - N$ can be factorized, then $M \leq N - 1$.</i>	Deductive-proof Construction	Quadratics
8*	<i>If $x^2 + Mx + N$ can be factorized, then $M < N + 1$.</i>	<i>[For graphs of $y=ax^2+bx+c$] If the x-intercepts of the graphs</i>	Proof-by-counterexample Construction	Quadratics

Table 30 (cont'd)

		<i>are both positive numbers, then both a and c are positive.</i>		
9	“If n is an even number, then $n^2 + 7n + 7$ is a composite number.”	If n is an odd number, then $n^2 + n + 1$ is a prime number.	Invalidation of Empirical Proof	Elementary Number Theory
10	Joe says: If the product of two whole numbers is even, then their sum is odd. Eva says: If the sum of two whole numbers is odd, then their product is even.	Gabriel says: If the product of two whole numbers is odd, then their sum is even. Dewey says: If the sum of two whole numbers is even, then their product is odd.	Logical non-equivalence between an implication and its converse	Elementary Number Theory
11	“Let N be an integer. If N^2 is an even number, then N is an odd integer” “Let N be an integer. If N is an odd integer, then N^2 is an odd number”	“Let N be an integer. If N^2 is an even number, then N is an odd integer” “Let N be an integer. If N is an even integer, then N^2 is an even number.”	Logical equivalence between an implication and its contrapositive	Elementary Number Theory
12	Let x and n be two real numbers. If $x > 0$ and $n > 0$, then $\frac{x}{n} + \frac{n}{x} \geq 2$	Let a and b be two numbers. If $a > 0$ and $b > 0$, then $\frac{1}{2}(a + b) \geq \sqrt{ab}$	Validation of Proof- by-contradiction	Algebra

*Note that Item 7 and Item 8 switched their order of appearance in Test set 2

Appendix C: Implications used in the Training Materials of all Conditions

Table 31: Implications used across all Conditions

Item	Control (Baseline):	Treatment*	
		Permission and Obligation (PO) Formulation	Wason (W) Formulation
1	If there is a vowel on one side, then there is an even number on the other side.	If a person is drinking beer, then the person must be over 19 years old.	If there is a vowel on one side, then there is an even number on the other side.
2	If there is a tree on one side, then there is a sea animal on the other side.	If the amount of a sale is over \$30, then the section manager must have approved it.	If there is a tree on one side, then there is a sea animal on the other side.
3	If the problem has the word 'more' to relate two quantities, then the answer is the sum of two quantities.	If the problem uses the word 'more' to relate two quantities, then I must use addition with the two quantities to get the answer.	If the problem has the word 'more' to relate two quantities, then the answer is the sum of two quantities.
4	If the numerator of the fraction subtraction is a prime number, then both fractions are the simplest fractions.	If the numerator of the fraction subtraction is a prime number, then I must have chosen two simplest fractions for subtraction.	If the numerator of the fraction subtraction is a prime number, then both fractions are the simplest fractions.
5	If a 2-digit number is divisible by 4, then the last digit of the multiple is an even number.	If the 2-digit number is divisible by 4, then the last digit of the code must be an even number.	If a 2-digit number is divisible by 4, then the last digit of its code is an even number.
6	If both the numbers added together are even, then their sum is an even number.	If any two even numbers are added together, then my friend must get an even number for the sum.	If both the numbers added together are even, then their sum is an even number.
7	If any three positive whole numbers multiplied together are consecutive in order, then their product is divisible by 6.	If I choose three consecutive numbers to be multiplied together, then the product must be divisible by 6.	If any three positive whole numbers multiplied together are consecutive in order, then their product is divisible by 6.
8	If a positive number and its reciprocal are added together, then the answer is at least 2.	If a positive number and its reciprocal are added together, then I must get a answer that is at least 2.	If a positive number and its reciprocal are added together, then the answer is at least 2.

Treatment*: Does not include prompts to search for/construct counterexamples

Appendix D: Training Materials for Conventional Approach (Condition C)

[Selected number of items included for illustration, other items are posed similarly]

In this study we are interested in how people interpret and reason about a very important type of logical statement, called the implications. Even though implication statements are really very simple, people often make errors in dealing with them. These instructions are intended to help you understand implications. Read through these instructions carefully; they should help you solve some mathematical thinking problems.

An implication statement consists of two statements which are often joined by the connective “*If... then.*” The implication statement can be expressed in the standard form

$$\text{If } P, \text{ then } Q$$

where the letters “*P*” and “*Q*” each represent a statement. This implication statement means “*If statement P is true, then statement Q is also true.*” For example, let *P* stand for “It is raining,” and *Q* stand for “The field is wet.” Then the implication becomes “*If it is raining, then the field is wet.*”

The table below lists how different truth values of statements *P* and *Q* affect the statement “*If P, then Q.*”

In the first line, when <i>P</i> is true and <i>Q</i> is also true, the implication “ <i>If P then Q</i> ” is true.		In the second line, when <i>P</i> is true and <i>Q</i> is false, the implication is false, since <i>P</i> does not imply <i>Q</i> .	
<u><i>P</i></u>	<u><i>Q</i></u>	<u><i>If P, then Q</i></u>	
True	True	True	
True	False	False	
False	False	True	
False	True	True	
In the third line, when <i>P</i> and <i>Q</i> are both false, the implication is true. For “ <i>If P then Q</i> ” to be true when <i>Q</i> is false, <i>P</i> cannot be true or it will go against the truth values shown in the second line.		In the last line, when <i>P</i> is false and <i>Q</i> is true, it does not mean that “ <i>If P then Q</i> ” is false, so the truth value is “true.”	

Figure 21: Truth table of “*If P then Q*”

Whenever one can find that an instance of *P* being true and *Q* being false, it is a counterexample to the implication “*If P then Q*” and so “*If P then Q*” is false.

Practice No.1

Let us try out the truth table again using an example of implications about the cards, “*If there is a vowel on one side, then there is an even number on the other side.*” P would be “there is a vowel on one side” and Q would be “there is an even number on the other side.” Fill in the blanks in the table using the truth values “True” or “False”:

there is a vowel on one side	there is an even number on the other side	<i>If there is a vowel on one side, then there is an even number on the other side.</i>
True	True	_____
True	_____	False
_____	False	True
False	True	_____

*Answer: 1st row – True, 2nd row – False, 3rd row – False, 4th row – True

The above table tells us that the implication “*If there is a vowel on one side, then there is an even number on the other side*” is true for almost all combinations of the truth values of statements P and Q except in the second line, when “there is a vowel on one side” is true and “there is an even number on the other side” is false. Whenever there is such a counterexample, the implication is not logically true.

Practice No. 2

Let us try out the truth table again using another implication about the sides of a card, “*If there is a tree on one side, then there is a sea animal on the other side.*” P would be “there is a tree on one side” and Q would be “there is a sea animal on the other side.” Fill in the blanks in the table using the truth values “True” or “False”:

there is a tree on one side	there is a sea animal on the other side	<i>If there is a tree on one side, then there is a sea animal on the other side.</i>
_____	False	False
True	_____	True
False	True	_____
_____	False	True

*Answer: 1st row – True, 2nd row – True, 3rd row – True, 4th row – False

The above table tells us that the implication “*If there is a tree on one side, then there is a sea animal on the other side*” is true for almost all combinations of P and Q except in the first line, when “there is a tree on one side” is true and “there is a sea animal on the other side” is false. Whenever there is such a counterexample, the implication is not logically true.

Now that you have some basic understanding about mathematical implications, you are ready for more practices. Please read the following problems and attempt them carefully.

Practice No.3

A simple word problem is usually solved by using an addition or subtraction. Decide whether the following implication statement is true or false:

If the problem has the word 'more' to relate two quantities, then the answer is the sum of two quantities.

Justify why your conclusion must be true or false using the most convincing argument.

Solution for Practice No.3

The above statement is false. To see this, we need an example.

One example would be the problem that uses the word "more" and requires subtraction as described below.

"John has 4 sweets and has one more sweet than Heidi. How many sweets does Heidi has?"

As you can see, the correct answer is obtained from the subtraction $4 - 1 = 3$ sweets even though the problem uses the word 'more.' So the implication is not logically true.

Practice No.8

A reciprocal of a number n is 1 divided by that number n . Decide whether the following implication statement is true or false:

If a positive number and its reciprocal are added together, then the answer is at least 2.
Justify why your conclusion must be true or false using the most convincing argument.

Solution for Practice No.8

As stated clearly in the question, the reciprocal of a number is 1 divided by that number.

Let's call the positive number to be p . The reciprocal is then $\frac{1}{p}$. Adding them together, we have

the sum $p + \frac{1}{p} = \frac{p}{1} + \frac{1}{p} = \frac{p^2 + 1}{p}$. To say that a number is at least 2 is the same as saying the

number minus 2 is at least 0. We will try out this idea to find out if the sum $\frac{p^2 + 1}{p}$ is at least 2.

Subtracting $\frac{p^2 + 1}{p}$ by 2, we have $\frac{p^2 + 1}{p} - 2$. Let's simplify this using algebraic operations.

$$\begin{aligned}\frac{p^2 + 1}{p} - 2 &= \frac{p^2 + 1}{p} - \frac{2p}{p} \\ &= \frac{p^2 - 2p + 1}{p} \\ &= \frac{(p-1)^2}{p}\end{aligned}$$

The denominator is p , which is a positive number as given. The numerator $(p-1)^2$ is the square of a number $p-1$. This is either 0 or a positive number, and cannot be a negative number.

So the fraction $\frac{(p-1)^2}{p}$ is either a zero divided by p , or a positive number divided by p .

Either way, it is not a negative number.

Thus we can say that the difference $\frac{p^2 + 1}{p} - 2$ is at least 0.

Which means that $\frac{p^2 + 1}{p} \geq 2$.

Therefore $p + \frac{1}{p}$ is at least 2 when p is a positive number.

So the implication statement is logically true.

Appendix E: Training Materials emphasizing Counterexample (Condition W)

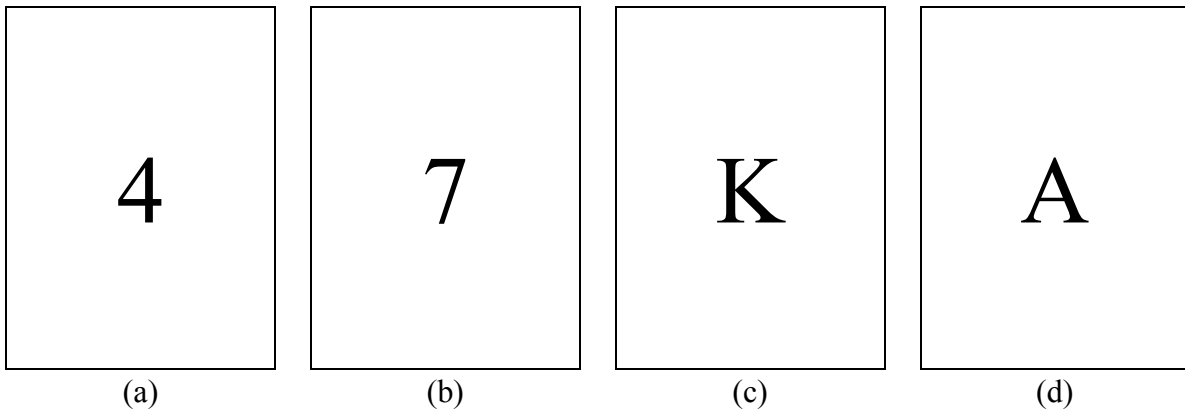
[Full version is enclosed]

Practice No.1

Below is shown a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is a letter on one of its sides and a number on the other.

Also below there is a rule which applies only to the four cards. Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the card(s) you want to turn.

Rule: *If there is a vowel on one side, then there is an even number on the other side.*



Solution for Practice No.1

The conditional rule in the question is *“If there is a vowel on one side, then there is an even number on the other side.”*

The rule is false when you can find cards that have a vowel on one side and an odd number on the other. The cards that need to be turned over for deciding are the cards “A” and “7” as they may be such instances that make the rule false.

The card “A” obviously needs to be turned over because if the other side shows an odd number, we have an instance of the rule being false. If the other side shows an even number, we have an instance verifying the rule but it is unknown whether the rule is false in other instances.

The card “7” also needs to be turned over because the other side may show a vowel. When that is the case, we have an instance of the rule being false.

If there is a consonant on one side, like the card “K”, whatever number is on the other side would not matter because it would not be an instance of the rule being false.

If an even number is shown, like the card “4”, it does not matter which letter is on the other side because it would not be an instance of the rule being false.

[the contents of the remaining Practices are shown as follows]

Practice No.2

Below is shown a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is a plant on one of its sides and an animal on the other.

Also below there is a rule which applies only to the four cards. Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the card(s) you want to turn.

Rule: *If there is a tree on one side, then there is a sea animal on the other side.*

Pine Tree	Lion	Whale	Hibiscus
(a)	(b)	(c)	(d)

Solution for Practice No.2

The conditional rule in the question is "*If there is a tree on one side, then there is a sea animal on the other side.*"

The rule is false when you can find cards that have a tree on one side and a land animal on the other. The cards that need to be turned over for deciding are the cards "Pine Tree" and "Lion" as they may be such instances that make the rule false.

The card "Pine Tree" obviously needs to be turned over because if the other side shows a land animal, we have an instance of the rule being false. If the other side shows a sea animal, we have an instance verifying the rule but it is unknown whether the rule is false in other instances.

The card "Lion" also needs to be turned over because the other side may show a tree. When that is the case, we have an instance of the rule being false.

If there is a non-tree plant on one side, like the card "Hibiscus", whatever animal is on the other side would not matter because it would not be an instance of the rule being false.

If a sea animal is shown, like the card "Whale", it does not matter which plant is on the other side because it would not be an instance of the rule being false.

Practice No.3

Below is shown a set of four worksheets, of which you can see only the exposed face but not the hidden back. On each sheet, there is a addition or subtraction word problem on one of its sides and a corresponding solution on the other.

Also below there is a rule which applies only to the four sheets. Your task is to decide which (if any) of these four sheets you must turn in order to decide if the rule is true. Don't turn unnecessary sheets. Tick the sheet(s) you want to turn.

Rule: *If the problem has the word 'more' to relate two quantities, then the answer is the sum of two quantities.*

John has 6 candies. Mary has 5 more candies than John. How many candies does Mary have?	For this problem, your child did $3 + 5 = 8$ books	For this problem, your child did $7 - 4 = 3$ stickers	John has 9 toys. Mary has 2 fewer toys than John. How many toys does Mary have?
(a)	(b)	(c)	(d)

Solution for Practice No.3

The conditional rule in the question is "*If the problem has the word 'more' to relate two quantities, then the answer is the sum of two quantities.*"

The rule is false when you can find sheets that have the word 'more' to relate two quantities on one side and the answer is not the sum of two quantities on the other. The sheets that need to be turned over for deciding are the sheets (a) and (c) as they may be such instances that make the rule false.

Sheet (a), which uses the word “more”, obviously needs to be turned over because if the other side shows a subtraction of two quantities, we have an instance of the rule being false. If the other side shows an addition of two quantities, we have an instance verifying the rule but it is unknown whether the rule is false in other instances.

Sheet (c), which shows that subtraction is used, also needs to be turned over because the other side may show the problem has the word 'more' to relate two quantities . When that is the case, we have an instance of the rule being false.

If the problem does not have the word “more”, like sheet (d), whatever solution would not matter because it would not be an instance of the rule being false.

If the answer is the sum of the quantities, like sheet (b), it does not matter whether the problem has the word ‘more’ because it would not be an instance of the rule being false.

Practice No.4

Two fractions are formed using four different positive whole numbers. They are then subtracted together. The numerator of the answer is checked to see if it is a prime number. Below is shown a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, the pair of fractions are shown on one of its sides and the answer of subtraction on the other.

Also below there is a rule which applies only to the four cards. Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don’t turn unnecessary cards. Tick the card(s) you want to turn.

Rule: *If the numerator of the fraction subtraction is a prime number, then both fractions are the simplest fractions.*

For $\frac{469}{1274} - \frac{198}{523}$, one of them is not the simplest fraction	For $\frac{945}{1378} - \frac{198}{523}$, both are simplest fractions	The answer for the subtraction is $\frac{21}{60}$	The answer for the subtraction is $\frac{3}{105}$
(a)	(b)	(c)	(d)

Solution for Practice No.4

The conditional rule in the question is “*If the numerator of the fraction subtraction is a prime number, then both fractions are the simplest fractions.*”

The rule is false when you can find cards that have a prime number as the numerator of the answer on one side and one of the fractions subtracted is not a simplest fraction on the other. The cards that need to be turned over for deciding are the cards (d) and (a) as they may be such instances that make the rule false.

The card (d) obviously needs to be turned over because if the other side shows that at least one fraction is not the simplest, we have an instance of the rule being false. If the other side shows both fractions are the simplest fractions, we have an instance verifying the rule but it is unknown whether the rule is false in other instances.

The card (a) also needs to be turned over because the other side may show a prime number as the numerator of the answer. When that is the case, we have an instance of the rule being false.

If the numerator of the answer is not a prime number, like the card (c), whether any of the fraction subtracted is the simplest would not matter because it would not be an instance of the rule being false.

If both fractions are the simplest fractions, like the card (b), it does not matter whether the numerator of the answer is a prime number because it would not be an instance of the rule being false.

Practice No.5

Below is shown a set of four labels, of which you can see only the exposed face but not the hidden back. On each label, there is a 2-digit number on one of its sides and a code number, which is a random multiple of the whole number, on the other.

Also below there is a rule which applies only to the four labels. Your task is to decide which (if any) of these four labels you must turn in order to decide if the rule is true. Don't turn unnecessary labels. Tick the label(s) you want to turn.

Rule: *If a 2-digit number is divisible by 4, then the last digit of its code is an even number.*

The code is 1376	The 2 digit number is 23	The 2 digit number is 36	The code is 147
(a)	(b)	(c)	(d)

Solution for Practice No.5

The conditional rule in the question is “*If a 2-digit number is divisible by 4, then the last digit of its code is an even number.*”

The rule is false when you can find labels that have a whole number which is divisible by 4 on one side and its code number does not end with an even digit on the other. The labels that need to be turned over for deciding are the labels showing “The 2-digit number is 36” and “The code is 147” as they may be such instances that make the rule false.

The label “The 2-digit number is 36” obviously needs to be turned over because if the other side shows an odd digit as the last digit of the code number, we have an instance of the rule being false. If the other side the last digit to be an even number, we have an instance verifying the rule but it is unknown whether the rule is false in other instances.

The label “The code is 147” also needs to be turned over because the other side may show a 2-digit whole number which is divisible by 4. When that is the case, we have an instance of the rule being false.

If the 2-digit whole number is not divisible by 4, like the label “The 2-digit number is 23”, whatever the code number is would not matter because it would not be an instance of the rule being false.

If the last digit of the code is even, like the label “The code is 1376”, it does not matter which 2-digit whole number is entered because it would not be an instance of the rule being false.

Practice No.6

Below is shown a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, two positive whole numbers are written on one of its sides and their sum on the other.

Also below there is a rule which applies only to the four cards. Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the card(s) you want to turn.

Rule: *If both the numbers added together are even, then their sum is an even number.*

14 + 37	12 + 26	36	43
(a)	(b)	(c)	(d)

Solution for Practice No.6

The conditional rule in the question is “*If the two numbers added together are both even, then their sum is an even number.*”

The rule is false when you can find cards that have two even numbers adding together on one side and their sum is odd on the other. The cards that need to be turned over for deciding are the cards “12 + 26” and “43” as they may be such instances that make the rule false.

The card “12 + 26” obviously needs to be turned over because if the other side shows an odd number as the sum, we have an instance of the rule being false. If the other side shows an even number, we have an instance verifying the rule but it is unknown whether the rule is false in other instances.

The card “43” also needs to be turned over because the other side may show two even numbers adding together. When that is the case, we have an instance of the rule being false.

If any of the numbers added is odd, like the card “14 + 37”, whatever the sum is would not matter because it would not be an instance of the rule being false.

If the sum is an even number, like the card “36”, it does not matter which two numbers are adding together because it would not be an instance of the rule being false.

Practice No.7

Below is shown a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, there is a set of three randomly chosen positive whole numbers on one of its sides and their product on the other.

Also below there is a rule which applies only to the four cards. Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the card(s) you want to turn.

Rule: *If any three positive whole numbers multiplied together are consecutive in order, then their product is divisible by 6.*

The product of multiplication is 210	The product of multiplication is 315	The three random whole numbers are: 4, 6 and 7	The three random whole numbers are: 7, 8 and 9
(a)	(b)	(c)	(d)

Solution for Practice No.7

The conditional rule in the question is “*If any three positive whole numbers multiplied together are consecutive in order, then their product is divisible by 6.*”

The rule is false when you can find cards that have three consecutive numbers on one side and their product is not divisible by 6 on the other. The cards that need to be turned over for deciding are the cards “The three random whole numbers are: 7, 8 and 9” and “The product of multiplication is 315” as they may be such instances that make the rule false.

The card “The three random whole numbers are: 7, 8 and 9” obviously needs to be turned over because if the other side shows that their product is not divisible by 6, we have an instance

of the rule being false. If the other side that shows their product is divisible by 6, we have an instance verifying the rule but it is unknown whether the rule is false in other instances.

The card “The product of multiplication is 315” also needs to be turned over because the other side may show three consecutive numbers. When that is the case, we have an instance of the rule being false.

If the three positive whole numbers are not consecutive, like the card “The three random whole numbers are: 4, 6 and 7”, whatever the product is divisible by 6 would not matter because it would not be an instance of the rule being false.

If the product is divisible by 6, like the card “The product of multiplication is 210”, it does not matter whether the three numbers are consecutive because it would not be an instance of the rule being false.

Practice No.8

A reciprocal of a number n is 1 divided by that number n . Below is shown a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, a positive number n and its reciprocal $\frac{1}{n}$ are added or subtracted together on one of its sides and the answer on the other.

Also below there is a rule which applies only to the four cards. Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the card(s) you want to turn.

Rule: *If a positive number and its reciprocal are added together, then the answer is at least 2.*

$$1.3 + \frac{1}{1.3}$$

(a)

The answer is 2.5.

(b)

$$1.57 - \frac{1}{1.57}$$

(c)

$$\text{The answer is } 1\frac{2}{3}.$$

(d)

Solution for Practice No.8

The conditional rule in the question is “*If a positive number and its reciprocal are added together, then the answer is at least 2.*”

The rule is false when you can find cards that have a positive number and its reciprocal adding together on one side and the answer is less than 2 on the other. The cards that need to be turned over for deciding are the cards “ $1.3 + \frac{1}{1.3}$ ” and “The answer is $1\frac{2}{3}$ ” as they may be such instances that make the rule false.

The card “ $1.3 + \frac{1}{1.3}$ ” obviously needs to be turned over because if the other side shows the answer to be less than 2, we have an instance of the rule being false. If the other side shows the result to be at least 2, we have an instance verifying the rule but it is unknown whether the rule is false in other instances.

The card “The answer is $1\frac{2}{3}$ ” also needs to be turned over because the other side may show a positive number and its reciprocal added together. When that is the case, we have an instance of the rule being false.

If a positive number and its reciprocal are subtracted together, like the card “ $1.57 - \frac{1}{1.57}$ ”, whatever the answer is would not matter because it would not be an instance of the rule being false.

If the answer is at least 2, like the card “The answer is 2.5”, it does not matter whether the positive number and its reciprocal are adding together because it would not be an instance of the rule being false.

Practice No.9

Below is shown a set of four cards, of which you can see only the exposed face but not the hidden back. For each card, there is a word written on one of its sides and a number on the other. Also below there is a rule which applies only to the four cards.

Rule: *If there is a single-digit number on one side, then there is a three-letter word on the other side.*

i) A counterexample makes the rule false when a single-digit number is on one side and a word other than three-letters long is on the other side. Can you think of such a counterexample?
_____ Yes. _____ No.

If “Yes”, what might the example be?

If “No”, based on what you know about the numbers and words in general, why is it impossible to find a counterexample to the rule?

ii) Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don’t turn unnecessary cards. Tick the card(s) you want to turn.

9	eat	12	jump
(a)	(b)	(c)	(d)

Solution for Practice No.9 part (i)

To find a counterexample to the statement, a card has a single-digit number paired with a word that does not have three letters.

The choice of the number may be from 0 to 9.

The word can have any number of letters except three letters. One choice will be the word "am". As you can see, a number "3" paired with the word "am" does not satisfy the statement.

Solution for Practice No.9 part (ii)

The conditional rule in the question is "*If there is a single-digit number on one side, then there is a three-letter word on the other side.*"

The rule is false when you can find cards that have a single-digit number and a word that does not have three letters. The cards that need to be turned over are the cards "9" and "jump" as they may be such instances that make the rule false.

The card "9" obviously needs to be turned over because if the other side shows a word that is not made up by three letters, we have an instance of the rule being false.

The card "jump" also needs to be turned over because the other side may show a single-digit number. When that is the case, we have an instance of the rule being false.

If there is a number of more than one digit, like the card "12", whatever word is on the other side would not matter because it would not be an instance of the rule being false.

If there is a three-letter word, like the card "eat", it does not matter whether a single-digit number is on the other side because it would not be an instance of the rule being false.

Practice No.10

Below is shown a set of four cards, of which you can see only the exposed face but not the hidden back. For each card, there is land vehicle on one of its sides and the name of a driver on the other.

Also below there is a rule which applies only to the four cards.

Rule: *If there is a female driver on one side, then there is a four-wheeled vehicle on the other side.*

i) A counterexample makes the rule false when a female driver is on one side and a vehicle that is not four-wheeled is on the other side. Can you think of such a counterexample?

_____ Yes. _____ No.

If “Yes”, what might the example be?

If “No”, based on what you know about the drivers and vehicles in general, why is it impossible to find a counterexample to the rule?

ii) Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don’t turn unnecessary cards. Tick the card(s) you want to turn.

Motorbike	Jeep	Aaron	Jane
(a)	(b)	(c)	(d)

Solution for Practice No.10 part (i)

To find a counterexample to the rule, a female driver is paired with a vehicle that does not have four wheels. So a driver's name like "Alice" will indicate a female driver.

The vehicle can have other numbers of wheels except four. One choice of such a vehicle will be a tricycle.

As you can see, a female driver named "Alice" paired with a tricycle does not satisfy the statement.

Solution for Practice No.10 part (ii)

The conditional rule in the question is "*If there is a female driver on one side, then there is a four-wheeled vehicle on the other side.*"

The rule is false when you can find cards that have a female driver on one side and a vehicle which is not four-wheeled on the other. The cards that need to be turned over are the cards "Jane" and "Motorbike" as they may be such instances that make the rule false.

The card "Jane" obviously needs to be turned over because if the other side shows a vehicle which is not four-wheeled, we have an instance of the rule being false.

The card "Motorbike" also needs to be turned over because the other side may show a female driver. When that is the case, we have an instance of the rule being false.

If there is a male driver, like the card "Aaron", whatever vehicle is on the other side would not matter because it would not be an instance of the rule being false.

If there is a four-wheeled vehicle, like the card "Jeep", it does not matter whether the driver is a female because it would not be an instance of the rule being false.

Practice No.11

Two whole numbers (denoted as A and B) are added and rounded using two different methods. In the first method, the numbers A and B are rounded upwards or downwards to the nearest ten first and then added together. In the second method, the numbers are added together first and the sum is then rounded to the nearest ten. Below is shown a set of four worksheets, of which you can see only the exposed face but not the hidden back. On each worksheet, the rounding of

numbers are shown on one of its sides and the answers calculated using both methods are shown on the other.

Also below there is a rule which applies only to the four worksheets.

Rule: *If both numbers are rounded to the nearest ten by rounding up, then the same answer is calculated from both methods.*

i) A counterexample makes the rule false when both numbers are rounded up to the nearest ten and different answers are calculated from both methods. Can you think of such a counterexample? ☐ Yes. ☐ No.

If "Yes," what might be the pair of numbers?

If "No," based on what you know about the rounding of numbers to the nearest ten in general, why is it impossible to find a counterexample to the rule?

ii) Your task is to decide which (if any) of these four worksheets you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the worksheet(s) you want to turn.

A is rounded up to 40, B is rounded down to 30	A is rounded up to 30, B is rounded up to 20	90 is the same answer calculated from both methods	60 is calculated from the first method and 70 from the second method
(a)	(b)	(c)	(d)

Solution for Practice No.11 part (i)

You are looking for a counterexample where the pair of numbers must both be rounded up in order to get to the nearest ten and yet different answers are obtained from both methods. How can we find such a pair of numbers? If we think about which numbers are rounded upwards to the nearest ten, their last digits are either "5", "6", "7", "8" or "9".

Rounding up both numbers before adding makes the sum even larger. The increase may be as big as a value of 10 if each number has its last digit as '5' and rounded up before adding. This may lead to a different answer than using the second method. One probable choice of A and B are thus two numbers whose last digits are "5".

Let's say we choose 35 and 55. Using the first method, $35 + 55$ is first rounded up to 40 and 60 and their sum is equal to 100. Using the second method, $35 + 55$ is first added to become 90, and remains the same after rounding to the nearest ten. As you can see, using these two numbers in both methods leads to different answers, which is a counterexample that makes the rule false.

Solution for Practice No.11 part (ii)

The conditional rule in the question is *"If both numbers are rounded to the nearest ten by rounding up, then the same answer is calculated from both methods."*

The rule is false when you can find worksheets that have both numbers rounded to the nearest ten by rounding up on one side but different answers obtained from using both methods on the other side. The worksheets that need to be turned over are the worksheets (b) and (d) as they may be such instances that make the rule false.

The worksheet (b) obviously needs to be turned over because if the other side shows different answers are being calculated from using both methods, we have an instance of the rule being false.

The worksheet (d) also needs to be turned over because the other side may show both numbers being rounded up to the nearest ten. When that is the case, we have an instance of the rule being false.

If any of the numbers are not rounded to the nearest ten by rounding upwards, like the worksheet (a), whether the answers are the same would not matter because it would not be an instance of the rule being false.

If the same answer is calculated from using both methods, like the worksheet (c), it does not matter whether both numbers are rounded up to the nearest ten because it would not be an instance of the rule being false.

Practice No.12

A positive whole number N and its square, N^2 , give some remainders other than 0 when divided by 5. For example, 7 gives a remainder of 2 and 7^2 gives a remainder of 4 when they are divided by 5. Below is shown a set of four cards, of which you can see only the exposed face but not the hidden back. For each card, the remainder when N is divided by 5 is written on one of its sides and the remainder when N^2 is divided by 5 is written on the other.

Also below there is a rule which applies only to the four cards.

Rule: *If N^2 divided by 5 give a remainder 1, then N divided by 5 also gives a remainder 1.*

A counterexample makes the rule false when N^2 divided by 5 gives a remainder 1 but N divided by 5 does not give a remainder 1. Can you think of such a counterexample? _____ Yes.

_____ No.

If “Yes,” what might the number N be?

If “No,” based on what you know about remainders and division by 5, why is it impossible to find a counterexample to the rule?

ii) Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the card(s) you want to turn.

$N = 4576$. The remainder is 1 when divided by 5	$N^2 = 4023238041$. The remainder is 1 when divided by 5	$N = 2749$. The remainder is 4 when divided by 5	$N^2 = 84327489$. The remainder is 4 when divided by 5
(a)	(b)	(c)	(d)

Solution for Practice No.12 part (i)

As stated clearly in the question, the counterexample you are looking for is a positive whole number N that does not give remainder 1 but its square, N^2 , gives a remainder 1 when divided by 5. How can we find such a number? Let's think about what it means to say “ N^2 divided by 5 gives a remainder 1” and “ N divided by 5 does not give a remainder 1.”

Given any whole number, when the remainder is 1 after being divided by 5, the last digit of the number is either ‘1’ or ‘6’. N does not give a remainder 1, so the last digit of N is not ‘1’ or ‘6’.

But N^2 gives a remainder 1, so the last digit of N^2 is ‘1’ or ‘6’.

Now you have a clearer idea about what N looks like. It is a positive whole number that ends with the digit ‘0’, ‘2’, ‘3’, ‘4’, ‘5’, ‘7’, ‘8’ or ‘9’ but its square, N^2 , ends with digit ‘1’ or ‘6’.

Keeping track of how the last digit of N changes when it is squared, we can narrow down what N might be:

When the last digit of N is '0' or '5', the last digit of N^2 is '0' or '5'.

When the last digit of N is '2' or '8', the last digit of N^2 is '4'.

When the last digit of N is '3' or '7', the last digit of N^2 is '9'.

When the last digit of N is '4' or '9', the last digit of N^2 is '1' or '6'.

Only the last case matches what we are looking for - that last digit of N^2 is '1' or '6'.

Now what happens if you choose a number from the last group, say $N = 14$, and $N^2 = 196$? When 196 is divided by 5, the remainder is 1 but the remainder is 4 when 14 is divided by 5. This is a counterexample that makes the rule false. In fact, any number whose last digit is '4' or '9' is also a counterexample that make the rule false.

Solution for Practice No.12 part (ii)

The conditional rule in the question is "*If N^2 divided by 5 give a remainder 1, then N divided by 5 also gives a remainder 1.*"

The rule is false when you can find cards that show N^2 with a remainder 1 on one side and N does not give a remainder 1 on the other. The cards that need to be turned over are the cards (b) and (c) as they may be such instances that make the rule false.

The card (b) obviously needs to be turned over because if the other side shows other remainders when N is divided by 5, we have an instance of the rule being false.

The card (c) also needs to be turned over because the other side may show that the remainder of N^2 to be 1. When that is the case, we have an instance of the rule being false.

If the remainder is not 1 when N^2 is divided by 5, like the card (d), whether N gives a remainder 1 on the other side would not matter because it would not be an instance of the rule being false.

If the remainder is 1 when N is divided by 5, like the card (a), it does not matter what remainder N^2 gives because it would not be an instance of the rule being false.

Practice No.13

Below is shown a set of four cards, of which you can see only the exposed face but not the hidden back. For each card, there is a set of five positive whole numbers written on one of its sides and their average written on the other.

Also below there is a rule which applies only to the four cards.

Rule: *If the five whole numbers are consecutive in order, then their average is a whole number.*

i) A counterexample makes the rule false when a set of five numbers are consecutive and their average is not a whole number. Can you think of such a counterexample?

____ Yes. ____ No.

If "Yes," what might the five consecutive numbers be?

If "No," based on what you know about the average of numbers and consecutive numbers in general, why is it impossible to find a counterexample to the rule?

ii) Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the card(s) you want to turn.

33, 34, 35, 36, 37	67, 20, 42, 54, 36	Their average is 45.2	Their average is 25
(a)	(b)	(c)	(d)

Solution for Practice No.13 part (i)

To find a counterexample, you want to look for five consecutive positive whole numbers whose average is not a whole number. Let's call the first whole number as n , the five consecutive numbers are namely, n , $n + 1$, $n + 2$, $n + 3$ and $n + 4$.

Since the average is not a whole number after dividing the sum of these five numbers by 5, it means that the sum of n , $n + 1$, $n + 2$, $n + 3$ and $n + 4$ is not a multiple of 5. That is, $n + (n + 1) + (n + 2) + (n + 3) + (n + 4) = 5n + 10$ is not a multiple of 5. So we have a counterexample that makes the rule false when we find a number n such that $5n + 10$ is not a multiple of 5.

But $5n + 10 = 5(n + 2)$, which actually means 5 multiplies a whole number (n is a whole number so $n + 2$ is also a whole number). It's impossible to pick a number n without having $5n + 10$ being a multiple of 5. Thus we cannot find five consecutive whole numbers whose average is not a whole number. Because we have no counterexample that makes the rule *false*, the rule is *true*.

Solution for Practice No.13 part (ii)

The conditional rule in the question is "If the five whole numbers are consecutive in order, then their average is a whole number."

The rule is false when you can find cards that have five consecutive numbers on one side and their average is not a whole number on the other. The cards that need to be turned over are the cards "33, 34, 35, 36, 37" and "Their average is 45.2" as they may be such instances that make the rule false.

The card "33, 34, 35, 36, 37" obviously needs to be turned over because if the other side shows their average is not a whole number, we have an instance of the rule being false.

The card "Their average is 45.2" also needs to be turned over because the other side may have five consecutive numbers. When that is the case, we have an instance of the rule being false.

If the five numbers are not consecutive, like the card "67, 20, 42, 54, 36", whatever their average is would not matter because it would not be an instance of the rule being false.

If the average is a whole number, like the card "Their average is 25", it does not matter whether the five numbers are consecutive because it would not be an instance of the rule being false.

Practice No.14

Below is shown a set of four cards, of which you can see only the exposed face but not the hidden back. For each card, there is an even number (denoted by N) on one of its sides and its square root on the other.

Also below there is a rule which applies only to the four cards.

Rule: *If the square root of N is a whole number, then the last digit of N is '0', '4' or '6'.*

i) A counterexample makes the rule false when the square root of an even number N is a whole number and the last digit of N is "2" or "8". Can you think of such a counterexample?

_____ Yes. _____ No.

If "Yes," what might the even number N be?

If "No," based on what you know about even numbers and perfect square numbers in general, why is it impossible to find a counterexample to the rule?

ii) Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the card(s) you want to turn.

The square root of N is 29.39387691339814	The even number N is 2,592	The even number N is 196	The square root of N is 26
--	------------------------------	----------------------------	------------------------------

(a)

(b)

(c)

(d)

Solution for Practice No.14 part (i)

A counterexample will be an even number N whose last digit is "2" or "8" and $\sqrt{N}$ is a whole number. Now the last digit of $\sqrt{N}$ is not '1', '3', '5', '7' or '9' because that would mean N is an odd number. So $\sqrt{N}$ is an even number and its last digit may be "2", "4", "6", "8" or "0". But when $\sqrt{N}$ is squared to become N , the last digit becomes "2" or "8".

The last digit of $\sqrt{N}$ is not "0" since after squaring $\sqrt{N}$, N will end with "0".

Also, the last digit of $\sqrt{N}$ is not "2" or "8" since after squaring $\sqrt{N}$, N will end with "4".

Similarly, the last digit of $\sqrt{N}$ is not "4" or "6" since after squaring $\sqrt{N}$, N will end with "6".

So $\sqrt{N}$ is not an even number whose last digit is "0", "2", "4", "6" or "8". But that means we have no other choices! As you can see, we cannot find an even number N with the last digit being "2" or "8" so that the square root $\sqrt{N}$ is an even number. Because we have no counterexample that makes the rule *false*, the rule is *true*.

Solution for Practice No.14 part (ii)

The conditional rule is "If the square root of N is a whole number, then the last digit of N is '0', '4' or '6'."

The rule is false when you can find cards that have a whole number for the square root of N on one side and the last digit of N is '2' or '8' on the other. The cards that need to be turned over are the cards "The square root of N is 26" and "The even number N is 2,592" as they may be such instances that make the rule false.

The card "The square root of N is 26" obviously needs to be turned over because if the other side shows the last digit of N to be '2' or '8', we have an instance of the rule being false.

The card "The even number N is 2,592" also needs to be turned over because the other side may show the square root of N to be a whole number. When that is the case, we have an instance of the rule being false.

If the square root of N is not a whole number, like the card "The square root of N is 29.39387691339814", whatever the even number N is on the other side would not matter because it would not be an instance of the rule being false.

If the even number N ends in either '0', '4' or '6', like the card "The even number N is 196", it does not matter whether the square root is a whole number because it would not be an instance of the rule being false.

Practice No.15

A and B are two positive whole numbers. Their product AB is compared with their sum $A + B$. Below is shown a set of four cards, of which you can see only the exposed face but not the hidden back. For each card, the numbers chosen for A and B are shown on one of its sides and their product and sum are shown on the other.

Also below there is a rule which applies only to the four cards.

Rule: If $A > 2$ and $B > 2$, then $AB > A + B$.

- i) A counterexample makes the rule false when both A and B are greater than 2 and $AB \leq A + B$. Can you think of such a counterexample? ____ Yes. ____ No.
If "Yes", what might the pair of numbers A and B be?

If “No”, based on what you know about products and sums in general, why is it impossible to find a counterexample to the rule?

ii) Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don't turn unnecessary cards. Tick the card(s) you want to turn.

<i>The product</i> $A \times B$ is 432 and the sum $A + B$ is 62	<i>A is a whole number between 10 and 50, B is a whole number between 45 and 200</i>	<i>A is a whole number between 100 and 500, B is 1</i>	<i>The product</i> $A \times B$ is 4 and the sum $A + B$ is 5
(a)	(b)	(c)	(d)

Solution for Practice No.15 part (i)

You are looking for a counterexample where both A and B are greater than 2 and $AB \leq A + B$. How can we find such a pair of numbers for A and B ? Let's think about what happens when $AB \leq A + B$ with both A and B greater than 2.

For a number greater than 2, it can be written as $2 + n$, where n is a certain positive number.

Since both A and B are greater than 2, we can write A and B as $2 + n$ and $2 + m$, where m and n are some positive numbers.

Now $AB \leq A + B$ can be rewritten as $(2 + m)(2 + n) \leq (4 + m + n)$.

Let's try to narrow down what m and n may be to find a counterexample for A and B . We can do this by simplifying the inequality.

After removing the brackets on both sides, the above inequality becomes

$$4 + 2m + 2n + mn \leq 4 + m + n$$

Subtracting the sum $4 + m + n$ from both sides, it becomes $m + n + mn \leq 0$.

Now we have a clearer idea about the counterexample we are looking for. A and B are numbers $2 + n$ and $2 + m$ such that $m + n + mn \leq 0$.

However, $m + n + mn$ is a sum of positive numbers (m and n are positive, so does mn). So it's impossible to have $m + n + mn$ to be less than or equal to 0! Thus we cannot find numbers that are greater than 2 for A and B so that $AB \leq A + B$. Because we have no counterexample that makes the rule *false*, the rule is *true*.

Solution for Practice No.15(ii)

The conditional rule in the question is "If $A > 2$ and $B > 2$, then $AB > A + B$."

The rule is false when you can find cards that have both A and B are greater than 2 on one side and AB is less than or equal to $A + B$ on the other. The cards that need to be turned over are the cards (b) and (d) as they may be such instances that make the rule false.

The card (b) obviously needs to be turned over because if the other side shows that the product is less than the sum, we have an instance of the rule being false.

The card (d) also needs to be turned over because the other side may show both A and B to be greater than 2. When that is the case, we have an instance of the rule being false.

If one of the numbers A and B is less than or equals to 2, like the card (c), whether AB is greater than $A + B$ would not matter because it would not be an instance of the rule being false.

If $AB > A + B$, like the card (a), it does not matter whether both A and B are greater than 2 because it would not be an instance of the rule being false.

Practice No.16

A quadratic equation $ax^2 + bx + c = 0$ is formed by non-zero whole numbers a , b and c . The solutions for x are solved by $x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$ and the discriminant of the equation, $b^2 - 4ac$, is calculated. Below is depicted a set of four cards, of which you can see only the exposed face but not the hidden back. On each card, the discriminant of the quadratic equation, $b^2 - 4ac$, is computed on one of its sides and the solutions for x in the equation are written on the other. Also below there is a rule which applies only to the four cards.

Rule: *If the solutions for x in the equation are whole numbers, then the discriminant $b^2 - 4ac$ is a perfect square.*

i) A counterexample makes the rule false when the solutions of a quadratic equation are whole numbers and its discriminant is not a perfect square. Can you think of such a counterexample? ☐ Yes. ☐ No.

If “Yes”, what might the quadratic equation be?

If “No”, based on what you know about quadratic equations in general, why is it impossible to find a counterexample to the rule?

ii) Your task is to decide which (if any) of these four cards you must turn in order to decide if the rule is true. Don’t turn unnecessary cards. Tick the card(s) you want to turn.

The discriminant is $(-1)^2 - 4 \times 6 \times (-7) = 169$, which is the square of 13.

(a)

The solutions for x are 2 and -3.

(b)

The solutions for x are $\frac{2}{3}$ and $1\frac{3}{4}$.

(c)

The discriminant is $(-12)^2 - 4 \times 2 \times 7 = 88$, which is not a perfect square.

(d)

Solution for Practice No.16 part (i)

You are looking for a counterexample where the quadratic equation $ax^2 + bx + c = 0$ has whole number solutions, say M and N , and the discriminant $b^2 - 4ac$ is not a perfect square. How can we find such a quadratic equation? Let’s think about the solutions M and N and how they are related to the discriminant $b^2 - 4ac$.

According to the quadratic formula, the solution(s) to the quadratic equation $ax^2 + bx + c = 0$

are $\frac{-b + \sqrt{b^2 - 4ac}}{2a}$ or $\frac{-b - \sqrt{b^2 - 4ac}}{2a}$.

So M is either equal to $\frac{-b + \sqrt{b^2 - 4ac}}{2a}$ or $\frac{-b - \sqrt{b^2 - 4ac}}{2a}$

Multiplying both sides by $2a$ and then adding b ,

$$2aM = -b + \sqrt{b^2 - 4ac} \text{ or } -b - \sqrt{b^2 - 4ac}$$

$$2aM + b = \sqrt{b^2 - 4ac} \text{ or } -\sqrt{b^2 - 4ac}$$

Since $b^2 - 4ac$ is not a perfect square, its square root, $\sqrt{b^2 - 4ac}$ is not a whole number. Hence $2aM + b$ is not a whole number. But it's impossible to pick whole numbers a and b without making $2aM + b$ become a whole number. Thus we cannot find non-zero whole numbers a , b and c for the quadratic equation $ax^2 + bx + c = 0$ whose solutions M and N are whole numbers and the discriminant $b^2 - 4ac$ is not a perfect square. Because we have no counterexample that makes the rule *false*, the rule is *true*.

Solution for Practice No.16(ii)

The conditional rule in the question is "*If the solutions for x in the equation are whole numbers, then the discriminant $b^2 - 4ac$ is a perfect square.*"

To decide that the rule is false, you need to look for cards that have whole numbers as the solution for x on one side and the discriminant not being a perfect square on the other. The cards that need to be turned over for deciding are the cards (b) and (d) as they may be such instances that make the rule false.

The card (b) obviously needs to be turned over because if the other side shows that the discriminant is not a perfect square, we have an instance of the rule being false. If the other side shows that the discriminant is a perfect square, we have an instance verifying a rule but it is unknown whether the rule is false in other instances.

The card (d) also needs to be turned over because the other side may show non-whole number solutions for x . When that is the case, we have an instance of the rule being false.

If any of the solutions for x is not a whole number, like the card (c), whatever the discriminant is would not matter because it would not be an instance of the rule being false.

If the discriminant is a perfect square, like the card (a), it does not matter whether the solutions for x are whole numbers because it would not be an instance of the rule being false.

Practice No. 17

Consider the mathematical statement:

"If two numbers multiplied together are both odd, then the product is an odd number."

i) A counterexample makes the rule false when you multiply two odd numbers together and the product is an even number. Can you think of such a counterexample?

_____ Yes. _____ No.

If "Yes", what are these numbers?

If "No", based on what you know about multiplications and odd numbers in general, why is it impossible to find an example of the statement being false?

ii) Arthur, Bonnie, Ceri, Duncan, Eric and Yvonne were trying to determine whether the following statement is true or false:

"If two numbers multiplied together are both odd, then the product is an odd number."

Arthur's answer:

When an odd number is divided by 2, it will give a number p as the quotient and a remainder of 1. So an odd number can be written as $2p + 1$ (p is the quotient).

The multiplication of 2 odd numbers can then be written as $(2p + 1) \times (2q + 1) = 4pq + 2p + 2q + 1$

The first three numbers have a common factor of 2 so their sum is an even number.

After adding 1 to the three even numbers, the result is an odd number.

So Arthur says it's true.

Duncan's answer:

Odd numbers end in 1, 3, 5, 7 or 9. When you multiply any two of these, the answer will still end in 1, 3, 5, 7 or 9.

So Duncan says it's true.

Yvonne's answer

```

      * * * * *
      * * * * *
***** x ***** = * * * * *
      * * * * *
      * * * * *
```

So Yvonne says it's true.

Bonnie's answer:

$1 \times 1 = 1$	$3 \times 3 = 9$	$5 \times 5 = 25$
$1 \times 3 = 3$	$3 \times 5 = 15$	$5 \times 7 = 35$

So Bonnie says it's true.

Ceri's answer:

Odd numbers are numbers that do not have the factor 2. When you multiply two odd numbers, both without the factor 2, the product will not have the factor 2. The product is then an odd number.

So the Ceri says it's true.

Eric's answer:

Let N = any whole number

Let M = any whole number

$N \times M = P$

$P \div M = N$

$P \div N = M$

$(P \times P) \div (N \times M) = N \times M = \text{an odd number.}$

So Eric says it's true.

From the above, choose one argument that best explains why it is impossible to find a counterexample that make the above statement false. Explain why you choose that argument.

Solution to Practice No. 17

Bonnie's answer and Yvonne's answer provide some examples that verify the statement.

However, they did not explain why other examples can never be counterexamples that make the statement false. So we do not know whether the statement is true.

Though Eric's answer describes all possible odd numbers and carries out different operations involving the numbers N and M correctly, he simply derives the operations from the statement " $M \times N = P$ " and later restates it as " $M \times N = \text{odd}$ ". His answer never provides any reason for why other examples can never be counterexamples that make the statement false. So we do not know whether the statement is true.

Duncan's answer describes all possible types of the odd numbers based on what is immediately known about odd numbers. He further justified that the multiplication of the last digits of odd

numbers would still give the last digits of the product as 1, 3, 5, 7, or 9. His answer has provided reasons for why all examples can never be counterexamples that make the statement false. So we know the statement is true.

Ceri's answer describes all possible odd numbers as numbers that do not have the factor 2. She further justified how multiplying odd numbers without the factor 2 can never produce a number with the factor 2. Her answer has provided reasons for why all examples can never be counterexamples that make the statement false. So we know the statement is true.

Arthur's answer provides explanations for writing any pair of odd numbers as $2p + 1$ and $2q + 1$ and carried out the multiplication of $(2p + 1)(2q + 1)$ correctly. He further justified that the first three numbers in the result of multiplication are even. His answer has provided reasons for why all examples can never be counterexamples that make the statement false.

Practice No. 18

Consider the mathematical statement:

"If any three whole numbers multiplied together are consecutive numbers, then the product is divisible by 6."

i) A counterexample makes the rule false when three consecutive numbers are multiplied together and the product is not divisible by 6. Can you think of such a counterexample?

_____ Yes. _____ No.

If "Yes", what are the three numbers?

If "No", based on what you know about multiplications and consecutive numbers in general, why is it impossible to find an example of the statement being false?

ii) Kate, Leon, Maria and Nisha were asked to determine whether the following statement is true or false:

"If any three whole numbers multiplied together are consecutive numbers, then the product is divisible by 6."

Kate's answer:

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 in the three times table.

Also, at least one number will be even and all even numbers are multiples of 2. If you multiply the three consecutive numbers together, the answer must have at least one factor of 3 and one factor of 2.

So Kate says it's true.

Leon's answer:

$$1 \times 2 \times 3 = 6$$

$$2 \times 3 \times 4 = 24$$

$$4 \times 5 \times 6 = 120$$

$$6 \times 7 \times 8 = 336$$

So Leon says it's true.

Maria's answer:

M is any whole number

$$\begin{aligned} M \times (M + 1) \times (M + 2) &= (M^2 + M) \times (M + 2) \\ &= M^3 + 3M^2 + 2M \end{aligned}$$

Cancelling the M 's gives $1 + 3 + 2 = 6$

So Maria says it's true.

Nisha's answer

Of the three consecutive numbers, x , $x + 1$ and $x + 2$, either x or $x + 1$ is an even number, so $x(x + 1)(x + 2)$ is a multiple of 2.

The first number x is either a multiple of 3 or not a multiple of 3.

If x is a multiple of 3 then $x(x + 1)(x + 2)$ is a multiple of 3

If x is not a multiple of 3, then x has a remainder of 1 or 2 when divided by 3

If x has a remainder of 1 then $x + 2$ is a multiple of 3

If x has a remainder of 2 then $x + 1$ is a multiple of 3

So $x(x + 1)(x + 2)$ is a multiple of 3 since either $x + 1$ or $x + 2$ is a multiple of 3

Regardless of whether x is a multiple of 3, $x(x + 1)(x + 2)$ will be a multiple of 3.

Since $x(x + 1)(x + 2)$ is a multiple of 2 and also 3, it is a multiple of 6.

So Nisha says it's true.

From the above, choose one argument that best explains why it is impossible to find a counterexample that make the above statement false. Explain why you choose that argument.

Solution to Practice No.18

Leon's answer provides some examples that verify the statement. However, he did not explain why other examples can never be counterexamples that make the statement false. So we do not know whether the statement is true.

Though Maria's answer describes all possible selection of three consecutive numbers as M , $M + 1$ and $M + 2$ and carried out the multiplication correctly, she simply cancels the M 's and changes $M^3 + 3M^2 + 2M$ into $1 + 3 + 2$. This is the same as substituting $M = 1$ into the result. Her answer never provides any reason for why substituting other values of M can never be counterexamples that make the statement false. So we do not know whether the statement is true.

Kate's answer describes all possible multiples of 6 based on what is immediately known about factors of 6. She further justified why any three consecutive numbers would include a multiple of 3 and an even number without fail. Her answer has provided reasons for why all examples can never be counterexamples that make the statement false. So we know the statement is true. Nisha's answer provides explanations for writing any triplet of numbers to be x , $x + 1$ and $x + 2$. She further justified that either x or $x + 1$ is an even number and that either x , $x + 1$ or $x + 2$ is a multiple of 3. She then explains that multiplying the numbers x , $x + 1$ and $x + 2$ together will produce a multiple of 6. Her answer has provided reasons for why all examples can never be counterexamples that make the statement false. So we know the statement is true.

Appendix F: Training Materials emphasizing Rule Violation (Condition PO)

[Selected number of items included for illustration, other items are posed similarly]

Practice No.1

Imagine you are a police officer on duty. It is your job to ensure that people conform to a law about alcohol drinking in a restaurant. The cards below show information about four people sitting at a table. On one side of a card is the person's age and on the other side of the card is what the person is drinking. The law you want to uphold is: *"If a person is drinking beer, then the person must be over 19 years old."*

Which of the card(s) below would you have to turn over to make sure that the people have violated the law? Turn over only those which you need to check to be sure. Tick the card(s) you would have to turn over.

22 years old	16 years old	Drinking Soda	Drinking Beer
(a)	(b)	(c)	(d)

Solution for Practice No.1

The rule in the question is: *"If a person is drinking beer, then the person must be over 19 years old."*

The people have violated the above rule when you can find cards that have a person drinking beer on one side but is not over 19 years old on the other. The cards that need to be turned over for checking are "Drinking Beer" and "16 years of age" as they may be instances of such violations.

The card "Drinking Beer" obviously needs to be turned over because if the other side shows the person's age to be under 19, we have an instance of violation. If the other side shows the person's age to be over 19, we have an instance of complying with the rule but it is unknown whether the rule is violated in other instances.

The card "16 years of age" also needs to be turned over because the other side may show a person drinking beer. When that is the case, we have an instance of violation.

If the person is not drinking beer, like the card "Drinking Soda", whatever is on the other side would not matter because this would not become an instance of violation.

And if the person is over 19 years old, like the card "22 years of age", it does not matter whether the person is drinking beer because it would not become an instance of violation.

Practice No.2

As part of your job as an assistant at Sears, you have the job of checking sales receipts to make sure that they are approved by the sales clerk in accordance to the store's policies. The amount is written on the front of the receipt, while the section manager's approval is signed on the back of the receipt. One store policy that the sales clerk has to follow is: *"If the amount of a sale is over \$30, then the section manager must have approved it."*

Which of the receipt(s) below would you have to turn over to make sure that the sales clerk has followed the policy? Turn over only those which you need to check to be sure. Tick the receipt(s) you would have to turn over.

The amount is \$50	The receipt is not signed	The receipt is signed	The amount is \$25
(a)	(b)	(c)	(d)

Solution for Practice No.2

The policy rule in the question is: *"If the amount of a sale is over \$30, then the section manager must have approved it."*

The sales clerk has violated the above rule when you can find receipts that have over \$30 on one side but have not been approved by the section manager on the other. The receipts that need to be turned over for checking are "The amount is \$50" and "The receipt is not signed" as they may be exhibiting such violations.

The card "The amount is \$50" obviously needs to be turned over because if the other side shows that the receipt is not signed, we have an instance of violation. If the other side shows the receipt is signed, we have an instance of complying with the rule but it is unknown whether the rule is violated in other instances.

The card "The receipt is not signed" also needs to be turned over because the other side may show an amount of over \$30. When that is the case, we have an instance of violation.

If the amount of a receipt is not over \$30, like the card "The amount is \$25", whatever is on the other side would not matter because this would not become an instance of violation.

And if the receipt is signed, it does not matter whether the amount is over \$30 because it would not become an instance of violation.

Practice No.11

You are a textbook author sorting through worked examples of rounding additions of numbers to the nearest ten. You calculated different pairs of whole numbers (denoted as A and B) using two different methods. In the first method, you round the numbers A and B upwards or downwards to the nearest ten first and then add them together. In the second method, you add the numbers first and then round the sum to the nearest ten. The calculation steps of both methods are written on one side and their results are compared on the other. To help students see that rounding numbers before or after addition can produce different answers, you want to identify the worked examples that violate the rule: *"If both numbers are rounded to the nearest ten by rounding up, then I must obtain the same answer using both methods."*

i) An instance violates the above rule when both numbers are rounded up to the nearest ten and different answers are obtained using both methods. Can you think of such an instance of violation? ☐ Yes. ☐ No.

If "Yes," what might be the pair of numbers?

If "No," based on what you know about the rounding of numbers to the nearest ten in general, why is it impossible to find an instance that violates the above rule?

ii) Which of the page(s) below would you have to turn over to make sure that the worked examples violate the rule? Turn over only those which you need to check. Tick the page(s) you would have to turn over.

A is rounded up to 40, B is rounded down to 30

(a)

A is rounded up to 30, B is rounded up to 20

(b)

I obtained the same answer 90 from both methods

(c)

I obtained 60 from the first method and 70 from the second method

(d)

Solution for Practice No.11 part (i)

You are looking for an instance where the pair of numbers must both be rounded up in order to get to the nearest ten and yet different answers are obtained from both methods. How can we find such a pair of numbers? If we think about which numbers are rounded upwards to the nearest ten, their last digits must be "5", "6", "7", "8" or "9".

Rounding up both numbers before adding makes the sum even larger. The increase may be as big as a value of 10 if each number has its last digit as '5' and rounded up before adding. This may lead to a different answer than using the second method. One probable choice of A and B are thus two numbers whose last digits are "5".

Let's say we choose 35 and 55. Using the first method, $35 + 55$ is first rounded up to 40 and 60 and their sum is equal to 100. Using the second method, $35 + 55$ is first added to become 90, and remains the same after rounding to the nearest ten. As you can see, using these two numbers in both methods leads to different answers, which is an instance violating the rule.

Solution for Practice No.11 (ii)

The conditional rule in the question is *"If both numbers are rounded to the nearest ten by rounding up, then I must obtain the same answer using both methods."*

A worked example has violated the above rule when you can find pages that show both numbers are rounded to the nearest ten by rounding up on one side but you obtain different answers from using both methods on the other side. The pages that need to be turned over for checking are pages (b) and (d) as they may be exhibiting such violations.

Page (b) obviously needs to be turned over because if the other side shows that you obtain different answers from using both methods, we have an instance of the rule being violated. If the other side shows that you obtain the same answer, we have an instance of the rule being followed but it is unknown whether the rule is violated in other instances.

Page (d) also needs to be turned over because the other side may show both numbers being rounded up to the nearest ten. When that is the case, we have an instance of the rule being violated.

If any of the numbers are not rounded to the nearest ten by rounding upwards, like the page (a), whether you obtain the same answer from both methods would not matter because it would not be an instance of the rule being violated.

And if you obtain the same answer using both methods, like the page (c), it does not matter whether both numbers are rounded up to the nearest ten because it would not be an instance of the rule being violated.

Practice No.13

You are helping your friend in checking some flash cards she made for investigating the average of whole numbers. She wrote any five positive whole numbers on one side of the flash cards and calculate their average on the other side. You want to make sure that her cards follow the mathematical rule, *"If the five whole numbers are consecutive in order, then my friend must get a whole number for the average."*

- i) An instance violates the above rule when a set of five numbers are consecutive and their average is not a whole number. Can you think of such an instance of violation? ☐ Yes.
☐ No.

If "Yes," what might the five consecutive numbers be?

If "No," based on what you know about the average of numbers and consecutive numbers in general, why is it impossible to find an instance that violates the above rule?

- ii) Which of the four card(s) below would you need to turn over to check if your friend's work has violated the rule? Turn only those which you need to check. Tick the card(s) you want to turn.

33, 34, 35, 36, 37	67, 20, 42, 54, 36	The average is 45.2	The average is 25
(a)	(b)	(c)	(d)

Solution for Practice No.13 part (i)

To find an instance that violates the rule, you want to look for five consecutive positive whole numbers whose average is not a whole number. Let's call the first whole number as n , the five consecutive numbers are namely, n , $n + 1$, $n + 2$, $n + 3$ and $n + 4$.

Since the average is not a whole number after the sum of these five numbers is divided by 5, it means that the sum of n , $n + 1$, $n + 2$, $n + 3$ and $n + 4$ is not a multiple of 5. That is, $n + (n + 1) + (n + 2) + (n + 3) + (n + 4) = 5n + 10$ is not a multiple of 5. So we have an instance of violation when we find a number n such that $5n + 10$ is not a multiple of 5.

But $5n + 10 = 5(n + 2)$, which actually means 5 multiplies a whole number (n is a whole number so $n + 2$ is also a whole number). It's impossible to pick a number n without letting $5n + 10$ become a multiple of 5. Thus we cannot find five consecutive whole numbers whose average is not a whole number. Because we have no instance that *violates* the rule, the rule is *complied*.

Solution for Practice No.13 (ii)

The conditional rule in the question is "*If the five whole numbers are consecutive in order, then my friend must get a whole number for the average.*"

Your friend's work violates the above rule when you can find cards that have five consecutive numbers on one side but the average calculated by your friend is not a whole number on the other. The cards that need to be turned over for checking are cards "33, 34, 35, 36, 37" and "The average is 45.2" as they may be exhibiting such violations.

Card "33, 34, 35, 36, 37" obviously needs to be turned over because if the other side shows their average is not a whole number, we have an instance of the rule being violated. If the other side shows their average to be a whole number, we have an instance of the rule being followed but it is unknown whether the rule is violated in other instances.

Card "The average is 45.2" also needs to be turned over because the other side may have five consecutive whole numbers. When that is the case, we have an instance of the rule being violated.

If the five numbers are not consecutive, like the card "67, 20, 42, 54, 36", whatever average computed by your friend would not matter because it would not be an instance of the rule being violated.

And if the average is a whole number, like the card "The average is 25", it does not matter whether the five numbers are consecutive because it would not be an instance of the rule being violated.

Appendix G: Interview Task and Sample of Questions

Given task:

A set of at least 3 whole numbers are randomly chosen and their average is calculated.

Decide whether the following implication statement is true or false:

If the set of whole numbers are consecutive in order, then their average is a whole number.

Justify why your conclusion must be true or false using the most convincing argument.

Follow-up action and questions:

- 1) Present the task to the student
- 2) How did you get the conclusion based on this statement?
- 3) What do you mean by [student's conclusion]?
- 4) How did you determine that this statement is false?
- 5) How did you come up with that example?
- 6) Given the statement had these counterexamples that you found, is there any way you can change the statement into a true statement?
- 7) What would you change it into?
- 8) Why did you change the statement in this way?
- 9) Can you show if this new statement is true, like the way you showed in [referring to a valid proof in the student's work]?
- 10) Do you think there are still any counterexamples that would make this statement false?
Can you explain why?

Appendix H: Student' Proving Attempts for the Given Implication

Given implication: [For a set of at least three numbers] If the set of whole numbers are all consecutive in order, then the average is a whole number.

Table 32: Attempts made by Students to prove or falsify the Implication

ID	Trial No. (True/False) and description of Proof constructed	Level of proof	Prompted for other sets of numbers
S1	Trial 1 (True) By calculating a few sets of consecutive numbers	1	Prompted after Trial 1; Focused on three whole numbers only
	Trial 2 (False) Based on the example 1, 2, 3 and 4, the average is 2.5	4	
S2	Trial 1 (True) Based on the total of $x, x + 1, \dots, x + 16$ and divide by 17 with x being the first whole number; Calculated the average of 0 to 16 to be a whole number.	6	No prompts given; focused on more than three numbers on her own
	Trial 2 (False) Calculated the average of $x, x + 1, x + 3, x + 4$ to be $x + 1.5$, which was a decimal number.	5	
	Trial 3 (True) Calculated the average of 2, 3, 4, 5, 6 to get a whole number	1	
	Trial 4 (False) Calculated the average of 2, 3, 4, 5 to get a decimal number	4	
S3	Trial 1 (False) Calculated the average of 7 to 10 and obtain decimal numbers as averages.	4	No prompts given; focused on more than three numbers on her own
	Calculated 1, 2, 3, 4, 5, 6 and obtained 3.4 The latter example was corrected to include 4 and average 3.5.	4	
S4	Trial 1 (False) Started with four consecutive numbers, from x to $x + 3$, and concluded that the average $x + 1.5$ is not a whole number.	5	No prompts given; focused on more than three numbers on his own

Table 32 (cont'd)

	Trial 2 (True) Tried three consecutive numbers, from x to $x + 2$, and obtained the average as $x + 1$. Could not conclude that the implication is true or false. Trial 3 (True) Tried five consecutive numbers, from x to $x+4$, and obtained the average as $x+2$.	6 6	
S5	Trial 1 (False) Calculated the average of 1 to 4 and obtained 2.5.	4	No prompts given
S6	Trial 1 (True) Based on the total of three consecutive numbers, $x + (x+1) + (x+2)$, and deduced $3x+3$ is divisible by 3 Trial 2 (False) Considered the total of four consecutive numbers, $x + (x+1) + (x+2) + (x + 3)$, and deduced that $(4x+6)$ is not divisible by 4. Trial 3 (True) Considered the total of five consecutive numbers, $x + (x+1) + (x+2) + (x + 3) + (x + 4)$, and deduced that $(5x+10)$ is divisible by 5.	6 5 6	Prompted after Trial 1; Focused on three whole numbers only
S7	Trial 1 (True) Based on the assumption that average number of an odd number of consecutive numbers is the middle number of the set and the middle number is a whole number, she concluded that the average is a whole number. Trial 2 (True) Using numbers 1, 2 and 3 to illustrate why the the center number is the average number and thus a whole number. Trial 3 (True) Using the center number as a reference, there are as many integers to the right and to the left of it and that one number on the right is 'balanced' with one number of the left because they are the same difference away from the center number.	3 1 4	No prompts given; aware of more than three numbers on her own self.

Table 32 (cont'd)

	Trial 4 (False) Calculated the average of 1, 2, 3, 4 and obtained decimal number 2.5. In addition, calculated the average of 2, 3, 4, 5, and obtain 3.5	4	
S8	Trial 1 (True) Used the smallest example of three consecutive numbers (1, 2 and 3)	1	Prompted after trial 1; Focused on three consecutive numbers only before prompt
	Trial 2 (False) Considered 4 consecutive numbers 1 to 4 and find that 10 cannot be divided by 4	4	
S9	Trial 1 (True) Use examples of three consecutive numbers, 1 to 3, 2 to 4, and 4 to 6 to verify the implication.	1	No prompts given
	Trial 2 (True) Based on the total of three consecutive numbers, $n + (n + 1) + (n + 2)$, and deduced $(3n + 3)/3 = n + 1$	6	
	Trial 3 (True) Based on four consecutive numbers, but considered the average as $[(n + 1) + (n + 2) + (n + 3)]$ and divide it by 3 by error.	3	
	Trial 4 (False) Based on the total of four consecutive numbers, $n + (n + 1) + (n + 2) + (n + 3)$, and deduced $(4n + 6)/4$ as the average	5	
	Trial 5 (False) Based on four consecutive numbers, 1 to 4, calculate the average to be $2\frac{1}{2}$, which is not a fraction.	4	
S10	Trial 1 (True) Based on two examples, 1 to 3 and 4 to 6, concluded that a multiple of 3 will always be present for the total of three consecutive numbers.	2	Prompted after Trial 2; focused on three consecutive numbers only; thought that the same conclusion holds for 4 numbers
	Trial 2 (True) Based on the total of three consecutive numbers, $n + (n + 1) + (n + 2)$, and deduced $(3n + 3) = 3(n + 1)$ and its division by 3 yields $n + 1$, which is a whole number.	6	

Table 32 (cont'd)

	Trial 3 (True) Based on four consecutive numbers, n to $(n + 3)$, but considered the total as divisible by 4 by erroneously assuming that the number pattern will repeat itself. Trial 4 (False) Based on four consecutive numbers, 100 to 103, finds the total 406 is not divisible by 4.	1 4	
S11	Trial 1 (True) By calculating a few examples of three consecutive numbers Trial 2 (True) Based on the total of three consecutive numbers, $x + (x+1) + (x+2)$, and deduced $3x+3$ is divisible by 3 Trial 3 (False) Considered the total of four consecutive numbers, $x + (x+1) + (x+2) + (x + 3)$, and deduced that $(4x+6)$ is not divisible by 4.	2 6 5	Prompted after trial 2; Focused on three consecutive numbers only before prompt
S12	Trial 1 (True) Choose the smallest example 1, 2, 3 and calculate its average to be 2. Conclude the implication as true based on the smallest example. Trial 2 (True) Consider a set of three non-consecutive numbers 1, 4, 6, and found that the average is not a whole number. Inferred that the implication is true since it is not true for its opposite example. Trial 3 (False) Considered the smallest example 1, 2, 3, 4 and found that the average is not a whole number. Concluded the implication as not true. Trial 4 (True) Considered the smallest example 1 to 5 and found that the average is a whole number. Concluded that the implication is true for five consecutive numbers.	2 1 4 2	Prompted after Trial 2; Consider four or more consecutive numbers

Table 32 (cont'd)

	Trial 5 (False) Considered the smallest example 1 to 6 and found that the average is not a whole number. Concluded that the implication is false for six consecutive numbers.	4	
--	--	---	--

Appendix I: Student' Modification of Implications and Proving Attempts

Given implication: [For a set of at least three numbers] If the set of whole numbers are all consecutive in order, then the average is a whole number.

Table 33: Modifications of Implications made by Students

ID	Modification made	Trial No. (True/False) and description of Proof constructed	Level of proof
S1	If the set of <u>three</u> whole numbers are <u>all</u> consecutive in order, then the average is a whole number. If the set of numbers <u>is an odd number of consecutive numbers</u> , then the average is a whole number.	Trial 1 (True) Based on examples of consecutive numbers from 1, 2, 3 to 4, 5, 6, whose sums are all divisible by 3	1
		Trial 2 (True) Based on two more “random choices” of three consecutive numbers, 55, 56, 57, and 99, 100, 101, and infer that the same will apply to other triplets of consecutive numbers. Trial 3 (True) Based on examples of two, three, four and five consecutive numbers, infer that the same will apply to any odd number of consecutive numbers, unless someone proved her wrong.	2 1
S2	If the set of whole numbers are consecutive in order <u>and the number of numbers are odd</u> , then their average is a whole number	Trial 1 (True) Observed that the constant number in the algebraic total of the consecutive numbers followed a pattern of “odd, odd, even, even.” Attempted to infer the average was a whole number from division of odd and even numbers. Trial 2 (True) Observed that the average followed a pattern of increasing by 1, in the form of $x + 1$, $x + 2$, $x + 3$, as the set of three, five, seven consecutive numbers were considered.	1 2
S3	A set of (<u>must be</u>) <u>3</u> whole numbers are randomly chosen...	Trial 1 (True) Tried three examples of three consecutive whole numbers (1, 2, 3; 4, 5, 6; 7, 8, 9) on	1

Table 33 (cont'd)

	If the set of whole numbers are consecutive in order, then their average is a whole number. [prompted to do other numbers]	calculator and obtain whole numbers as average.	
	A set of <u>(must be) 3 or 5</u> whole numbers are randomly chosen..., [the rest remains the same] A set of whole numbers <u>which must be in odd number form</u>, ...[the rest remains the same]	Trial 2 (True) Calculated the average of the set 1 to 5 and obtained a whole number Trial 3 (True) Calculated the average of the sets, 1 to 5, 1 to 7, 1 to 9, and 1 to 11, and obtained a whole number. '80 % sure' about the truth of the modified implication.	1 1
S4	If the set of whole numbers are consecutive in order <u>and of an even number, like 4 to 6 numbers</u>, then [same consequent] If the set of whole numbers are consecutive in order <u>and are of an odd number, like 3 or 5 numbers</u>, then [same consequent] If the set of whole numbers are consecutive in order <u>and of an even number, like 4 to 6 numbers</u>, then their average is <u>not</u> a whole number.	Trial 1 (True) Referred to the earlier proof that calculated the average of five consecutive numbers, from x to $x + 4$. Trial 2 (True) Calculated the average of six consecutive numbers, from x to $x + 5$, and obtained $x + 2.5$. Trial 3 (True) Based on the observation that summing the first and last number equals to summing the second and second-last number (and so on), re-write the first and the last number in the set as x and $x + n$, and proposed that the sum of the consecutive number is $[x + (x+n)]x n$. Attempted to deduce that the average is a whole number when n is odd but was unsuccessful.	3 3 3
S5	If the set of whole numbers are consecutive in order, then their average is a whole number <u>or a decimal number</u>.	Trial 1 (True) Based on the set of numbers 1 to 4, the average is 2.5. Claimed that the sum of three consecutive numbers is a 'factor' [multiple] of three.	1

Table 33 (cont'd)

	If the set of n consecutive whole numbers, then their average would be a whole number if the sum is a factor of n , or a decimal number if the sum is not a factor of n .	Trial 2 (True) Based on the set of numbers 1 to 5, the total is 15, divisible by 5, and the average is 3. Trial 3 (True) Use x, $x+1$, $x+2$ to represent three consecutive numbers. Subtracted 3 from $3x+3$ and then divided by x to get 3. Trial 4 (True) For three, four or n consecutive numbers, either their sum is or is not a 'factor' [multiple] of three, four or n. If it is, then the average is a whole number, If it is not, then the average is a decimal number.	1 5 1
S6	If the set of integers <u>are odd and</u> are consecutive in order, then their average is a whole number.	Trial 1 (True) Use x, $x+1$, $x+2$, ..., and so on to represent consecutive numbers. Calculated the total of three, five, seven, nine, and eleven consecutive numbers and verified that the total of each set of numbers is divisible by the number of the set. Hence the average is a whole number. Trial 2 (True) Deduced that the total of a set of consecutive numbers is in the form of $ax+b$, and observed that the constant b is a multiple of the coefficient of x, a. Trial 3 (True) Calculate the average of nine consecutive numbers using x, $x+1$, etc. and obtained $9x+36$, which is divisible by 9.	3 2 5
S7	<u>Let the total number of integers be an odd number and the integers must be greater than 1.</u> If the set of integers are consecutive in order, then their average is a whole number.	Trial 1 (True) Based on earlier proof of why the implication was true for odd number of numbers. Trial 2 (True) Based on the example of $1 + 2 + 3 = 6$, show that the sum of consecutive numbers will produce the center number 2 after dividing 3.	3 1

Table 33 (cont'd)

		Trial 3 (True) The last digits of all consecutive numbers are any digits from 0 to 9. Used examples of 71, 72, 73 and 74, 75, 76, to show that the average of the last digits of the consecutive numbers is a whole number.	2
S8	If a set of three whole numbers are added in consecutive order, then their average is a whole number. [after prompting, thinking about 5 whole numbers, seven whole numbers] If the set add up [to] become an odd number, then the average would be a whole number. If the number of consecutive numbers is an odd number, then their average is a whole number.	Trial 1 (True) Based on examples of three consecutive numbers. If the sum of the three numbers is odd, then the average is a whole number. Trial 2 (True) Use $x+(x+1)+(x+2)=3x+3$ and divide by 3 = $x+1$. Trial 3 (True) Use $x+(x+1)+(x+2)+(x+3)+(x+4)=5x+10$ and divide by 5 = $x+2$. Trial 4 (True) Use $x+(x+1)+\dots+(x+6)=7x+21$ and divide by 7 = $x+3$. The average of the consecutive numbers will increase by 1	2 6 3 3
S9	If the set of <u>3</u> whole numbers are consecutive in order, their average is a whole number. If the set of whole numbers are consecutive in order, their average <u>may</u> be a whole number. If <u>there are five consecutive</u> whole numbers, then the average is a whole number When it's a even number, like two, four, six, eight consecutive numbers, you cannot get a middle number [for the average]. If an odd number of numbers are consecutive, then the	Trial 1 (True) Use $n, n+1$ to $n+4$ and calculate their average as $n+2$, which is a whole number. Trial 2 (True) Based on earlier examples such as four consecutive numbers. Trial 3 (True) Choose random set of consecutive numbers, 111 to 115, 10 to 14, and 11 to 17, and	6 3 2

Table 33 (cont'd)

	average is the <u>middle number</u> .	calculate each average. Propose that average is the middle number of the set and the set of even number of consecutive numbers had no middle numbers.	
S10	If the set of 3 whole numbers are consecutive in order, then their average will be a whole number. [after prompting, thinking about 5 whole numbers, seven whole numbers] If the set of odd number of whole numbers are consecutive in order, then their average will be a whole number.	Trial 1 (True) Based on the earlier proof using $n + (n + 1) + (n + 2)$ and that n is a whole number. Trial 2 (True) Representing the first number as n, deduced that the sum of three, five and seven consecutive numbers starting from n are $3n + 3$, $5n + 10$, and $7n + 21$. Since each of the total is a multiple of the number of numbers, the average of three, five and seven consecutive numbers are whole numbers. Trial 3 (True) In the calculation of the sum of odd number of consecutive numbers, the total will be of the algebraic $an + b$. b, the constant in the total will be related to the coefficient of n, a.	6 3 3
S11	If the set has an odd number of consecutive numbers, then their average is a whole number.	Trial 1 (True) Based on examples of three and four consecutive numbers, expressed in algebraic symbols. Trial 2 (True) Calculate the average of 1 to 11 and get 6. Proposed a formula “the sum of constants / constants = whole number” to justify the conclusion Trial 3 (True) Proposed an algebraic formula “$(1+2+3+4+5+\dots+n)/n = \text{whole number.}$”	2 2 2

Table 33 (cont'd)

S12	If the set of the whole number is either three numbers, multiples of three or odd numbers, [then their average is a whole number]. If the set consists of 3 whole numbers and any amount of odd numbers, their average is a whole number.	No proof Trial 1 (True) Use numerical examples, $1+2+3+\dots$, of an odd number of consecutive numbers to verify the implication. Trial 2 (True) Observed that the sum of $1+2+3+\dots$ is a multiple of the number of the set, when the number is odd, e. g., for three, five, seven, nine and eleven numbers. Based on this observation, concluded that the implication is true Trial 3 (True) The sum of $1+2+3+\dots$ followed a pattern when the numbers in the sum are odd. For three consecutive numbers, the total is 3 times 2. For five numbers, the total is 5 times 3. For seven numbers, the total would be 7 times 4. The multiplicand increased by 1 as the odd number increased. The pattern showed that implication is true.	1 2 2
-----	---	--	----------------------------

REFERENCES

REFERENCES

- Agresti, A. (2010). *Analysis of Ordinal Categorical Data* (2nd ed.). New York: Wiley.
- Akritis, M. G. (1990). The Rank Transform Method in Some Two-Factor Designs. *Journal of the American Statistical Association*, 85(409), 73-78.
- Alcock, L., & Weber, K. (2005). Proof validation in real analysis: Inferring and checking warrants. *Journal of Mathematical Behavior*, 24, 125-134.
- Antonini, S., & Mariotti, M. (2008). Indirect proof: what is specific to this way of proving? *ZDM*, 40(3), 401-412.
- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from Examples: Instructional Principles from the Worked Examples Research. *Review of Educational Research*, 70(2), 181-214.
- Balacheff, N. (1988). Aspects of Proof in pupils' practice of school mathematics. In D. Pimm (Ed.), *Mathematics, teachers and children* (pp. 216-235). London: Hodder & Stoughton.
- Ball, D. L., & Bass, H. (2003). Making Mathematics Reasonable in School. In J. Kilpatrick, W. G. Martin & D. Schifter (Eds.), *A research companion to principles and standards for school mathematics* (pp. 27-44). Reston, VA: National Council of Teachers of Mathematics.
- Benacerraf, P., & Putnam, H. (1964). *Philosophy of mathematics; selected readings*. Englewood Cliffs, N.J.: Prentice-Hall.
- Boero, P. (1999). Argumentation and mathematical proof: A complex, productive, unavoidable relationship in mathematics and mathematics education. *International Newsletter on the Teaching and Learning of Mathematical Proof*, 7/8.
- Chazan, D. (1993). High School Geometry Students' justification for their views of empirical evidence and mathematical proofs. *Educational Studies in Mathematics*, 24, 359-387.
- Cheng, P. W., & Holyoak, K. J. (1985). Pragmatic reasoning schemas. *Cognitive Psychology*, 17, 391-416.
- Cheng, P. W., Holyoak, K. J., Nisbett, R. E., & Oliver, L. M. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*(18), 293-328.
- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Coe, R., & Ruthven, K. (1994). Proof Practices and Constructs of Advanced Mathematics Students. *British Educational Research Journal*, 20(1), 41-53.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: teaching the crafts of reading, writing, and mathematics. In L. B. Resnick (Ed.), *Knowing, learning, and instruction* (pp. 453-494). Hillsdale, NJ: Erlbaum.
- Conover, W. J., & Iman, R. L. (1981). Rank Transformations as a Bridge Between Parametric and Nonparametric Statistics. *The American Statistician*, 35(3), 124-129.
- Deer, G. W. (1969). The Effects of Teaching an Explicit Unit in Logic on Students' Ability to Prove Theorems in Geometry. Unpublished doctoral dissertation Florida State University.
- Durand-Guerrier, V. (2003). Which notion of implication is the right one? From logical considerations to a didactic perspective. *Educational Studies in Mathematics*, 53, 5-34.
- Durand-Guerrier, V. (2008). Truth versus validity in mathematical proof. *ZDM*, 40(3), 373-384.
- Epp, S. S. (1994). The Role of proof in problem solving. In A. H. Schoenfeld (Ed.), *Mathematical Thinking and Problem Solving* (pp. 257-269). Hove: Lawrence Erlbaum Associates.
- Epp, S. S. (2003). The Role of Logic in Teaching Proof. *The American Mathematical Monthly*, 110(10), 886-899.
- Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.
- Goetting, M. (1995). *The college students' understanding of mathematical proof*. Unpublished doctoral dissertation, University of Maryland, College Park.
- Griggs, R. A., & Cox, J. R. (1982). The elusive thematic-materials effect in Wason's selection task. *British Journal of Psychology*, 73(407-420).
- Harel, G., & Sowder, L. (1998). Students' Proof Schemes. In A. H. Schoenfeld, J. Kaput & E. Dubinsky (Eds.), *Research in College Mathematics Education III* (pp. 234 - 283). Providence, R.I.: American Mathematical Society.
- Healy, L., & Hoyles, C. (2000). A Study of Proof Conceptions in Algebra. *Journal for Research in Mathematics Education*, 31(4), 396-428.
- Hilbert, T. S., Renkl, A., Kessler, S., & Reiss, K. (2008). Learning to prove in geometry: Learning from heuristic examples and how it can be supported. *Learning and Instruction*, 18(1), 54-65.
- Hoyles, C., & Küchemann, D. (2002). Students' understanding of logical implication. *Educational Studies in Mathematics*, 51, 193-223.

- Hoyles, C., & Küchemann, D. (2003). Students' understanding of logical implication. *Educational Studies in Mathematics*, 51, 193-223.
- Inglis, M., & Simpson, A. (2004). Mathematicians and the Selection Task. In M. J. Hoines & A. B. Fuglestad (Eds.), *Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 3, pp. 89-96). Bergen, Norway.
- Inglis, M., & Simpson, A. (2006). The role of mathematical context in evaluating conditional statements. In J. Novotna, H. Moraov, M. Kratka & N. Stehlikova (Eds.), *Proceedings of the 30th International Conference on the Psychology of Mathematics Education* (Vol. 3, pp. 337-344). Prague, Czech Republic.
- Jackson, S. L., & Griggs, R. A. (1988). Education and the selection task. *Bulletin of the Psychonomic Society*, 26, 327-330.
- Jahnke, H. (2008). Theorems that admit exceptions, including a remark on Toulmin. *ZDM*, 40(3), 363-371.
- Johnson-Laird, P. N., & Byrne, R. M. J. (1991). *Deduction*. Hove ; Hillsdale, USA: Erlbaum.
- Knuth, E. J., Choppin, J., & Bieda, K. (2009). Middle school students' productions of mathematical justification. In M. Blanton, D. Stylianou & E. Knuth (Eds.), *Teaching and learning proof across the grades: A K-16 perspective* (pp. 153–212). NY: Routledge.
- Ko, Y.-Y., & Knuth, E. J. (2009). Undergraduate mathematics majors' writing performance producing proofs and counterexamples about continuous functions. *Journal of Mathematical Behavior*, 28, 68-77.
- Lawson, A. E. (1990). Use of reasoning to a contradiction in grades three to college. *Journal of Research in Science Teaching*, 27(6), 541-551.
- Leighton, J. P. (2006). Teaching and Assessing Deductive Reasoning Skills. *Journal of Experimental Education*, 74(2), 109-136.
- Lin, F. (2005). Modeling Students' Learning on Mathematical Proof and Refutation. In H. L. Chick & J. L. Vincent (Eds.), *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 1, pp. 3-18). Melbourne: PME.
- Mason, J., & Pimm, D. (1984). Generic Examples: Seeing the General in the Particular. *Educational Studies in Mathematics*, 15(3), 277-289.
- Miyazaki, M. (2000). Levels of Proof in Lower Secondary School Mathematics. *Educational Studies in Mathematics*, 41(1), 47.
- Moore, R. C. (1994). Making the transition to formal proof. *Educational Studies in Mathematics*, 27(3), 249-266.

- Mueller, D. J. (1975). *Logic and the ability to prove theorems in geometry*. Unpublished Thesis (Ph D) - Florida State University, Mueller,, [Tallahassee, Fla.].
- Mullis, I. V. S., Martin, M. O., & Foy, P. (2008). *TIMSS 2007 international mathematics report: Findings from IEA's Trends in International Mathematics and Science Study at the fourth and eighth grades*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Lynch School of Education, Boston College.
- Pedemonte, B., & Buchbinder, O. (2011). Examining the role of examples in proving processes through a cognitive lens: the case of triangular numbers. *ZDM*, 1-11.
- Platt, J. L. (1967). *The Effect of the Use of Mathematical Logic in High School Geometry: An Experimental Study*. Unpublished doctoral dissertation, Colorado State College.
- Platt, R., & Griggs, R. A. (1993). Facilitation in the abstract selection task; The effects of attentional and instructional factors. *Quarterly Journal of Experimental Psychology-A*, 46(4), 591-613.
- Pollard, P., & Evans, J. S. B. T. (1987). On the relationship between content and context effects in reasoning. *American Journal of Psychology*, 100, 41-60.
- Quine, W. V. O. (1950). *Methods of Logic*. New-York.: Holt, Rinehart & Winston.
- Recio, A., & Godino, J. (2001). Institutional and personal meanings of mathematical proof. *Educational Studies in Mathematics*, 48(1), 83.
- Rissland, E. L. (1991). Example-based reasoning. In J. F. Voss & D. N. Perkins (Eds.), *Informal reasoning and education* (pp. 187-208): Hillsdale, NJ, Engleandates, Inc, 1991, xvii, 498.
- Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, Fla.: Academic Press.
- Segal, J. (1999). Learning about mathematical proof: Conviction and validity. *Journal of Mathematical Behavior*, 18(2), 191-210.
- Selden, A., & Selden, J. (2003). Validations of Proofs Considered as Texts: Can Undergraduates Tell Whether an Argument Proves a Theorem? *Journal for Research in Mathematics Education*, 34(1), 4-36.
- Sharon, L. S. (1989). Van Hiele Levels and Achievement in Writing Geometry Proofs. *Journal for Research in Mathematics Education*, 20(3), 309-321.
- Simon, M. (2000). Reconsidering mathematical validation in the classroom. In *Proceedings of 24th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 4, pp. 161-168). Hiroshima, Japan.
- Simon, M., & Blume, G. W. (1996). Justification in the mathematics classroom: A study of prospective elementary teachers. *The Journal of Mathematical Behavior*, 15(1), 3.

- Stenning, K., & Lambalgen, M. v. (2004). A little logic goes a long way: basing experiment on semantic theory in the cognitive science of conditional reasoning. *Cognitive Science*, 28, 481-529.
- Stenning, K., & Monaghan, P. (2004). Strategies and Knowledge Representation. In R. J. Sternberg & J. P. Leighton (Eds.), *The nature of reasoning* (pp. 129-168). Cambridge, U.K. ; New York: Cambridge University Press.
- Stylianides, A. (2007). Proof and Proving in School Mathematics. *Journal for Research in Mathematics Education*, 38(3), 289-321.
- Stylianides, A., & Stylianides, G. (2009a). Proof constructions and evaluations. *Educational Studies in Mathematics*, Retrieved from <http://www.springerlink.com/content/v2611945638x1763>. doi:10.1007/s10649-009-9191-3
- Stylianides, A., & Stylianides, G. (2009b). Proof constructions and evaluations. *Educational Studies in Mathematics*, 72(2), 237-253
- Stylianides, A., Stylianides, G., & Philippou, G. (2004). Undergraduate students' understanding of the contraposition equivalence rule in symbolic and verbal contexts. *Educational Studies in Mathematics*, 55(1), 133-162.
- Stylianides, A. J. (2007). Proof and Proving in School Mathematics. *Journal for Research in Mathematics Education*, 38(3), 289-321.
- Stylianides, G., & Stylianides, A. (2008). Proof in School Mathematics: Insights from Psychological Research into Students' Ability for Deductive Reasoning. *Mathematical Thinking and Learning*, 10(2), 103-133.
- Tarski, A. (1956). The concept of truth in formalized languages. In Tarski (Ed.), *Logic, Semantics, Metamathematics: papers from 1923 to 1938*. Oxford: Clarendon Press.
- Thompson, D. R. (1996). Learning and teaching indirect proof. *The Mathematics Teacher*, 89(6), 474-482.
- van Dormolen, J. (1977). Learning to understand what giving a proof really means. *Educational Studies in Mathematics*, 8, 27-34.
- Wason, P. C. (1968). Reasoning about a rule. *Quarterly Journal of Experimental Psychology*(20), 273-281.
- Weber, K. (2001). Student difficulty in constructing proofs: The need for strategic knowledge. *Educational Studies in Mathematics*, 48(1), 101 - 119.
- Weber, K. (2010). Mathematics Majors' Perceptions of Conviction, Validity, and Proof. *Mathematical Thinking and Learning*, 12(4), 306 - 336.

- Weber, K., & Alcock, L. (2004). Semantic and Syntactic Proof Productions.(Author abstract). *Educational Studies in Mathematics*, 56(3), 209(226).
- Zazkis, R., & Chernoff, E. (2008). What makes a counterexample exemplary? *Educational Studies in Mathematics*, 68(3), 195-208.
- Zhu, X., & Simon, H.-A. (1987). Learning mathematics from examples and by doing.