THE COGNITION OF COUNTEREXAMPLE IN MATHEMATICS STUDENTS

SHANNON EZZAT AND SCOTT RODNEY

Abstract. Studying Mathematics requires a synthesis of skills from a multitude of academic disciplines; logical reasoning being chief among them. This paper explores mathematical logical preparedness of students entering first year university mathematics courses and also the effectiveness of using logical facility to predict successful course outcomes. We analyze data collected from students enrolled at the University of Winnipeg in a pre-service course for high school teachers. We do find that, being able to successfully answer logical questions, both before and after intervention, are significant in relation to improved student outcomes.

Keywords. Counterexample, mathematical thinking, logical inference, student success

1. Introduction

This paper explores mathematical logic preparedness and its effect on success in the first year university mathematics classroom. It is well known that skill sets among students entering courses in university mathematics vary widely from student to student and thus an understanding of the abilities and effects of logical aptitude is a worthwhile endeavour to help improve student success in mathematics.

This work focuses on two themes: student facility in reasoning about counterexample, and how well assessment of this logical skill can predict successful course outcomes in introductory mathematics courses.

To address these questions, a 3-month study of university students was initiated in MATH-2903, Math for Early/Middle Year Teachers during the winter semester of the 2017-18 academic year. As indicated by its name, this class is designed for pre-service elementary- and middle-school teachers. The prerequisite for the course is any university-level math class from Grade 12 in Manitoba (with a grade above 65% for applied math or any passing grade in precalculus) and does not count towards any math major course, or as a required math course for Bachelor of Science students. The students in this course predominantly have poor to adequate arithmetic and algebraic skills and a substantially negative attitude towards mathematics as a subject.

Our methods are inspired by several works on student ability with logic and counterexample including [BF] which discusses undergraduate students understanding of various concepts in calculus using the tool of diagnostic tests, intervention/primers, and quizzes; [L], which studies the cognition of counterexample of high school students in a similar fashion; and [IS], which studies undergraduate (and staff) success in the Wason Selection Task. We use these tools to measure students understanding of logical statements, and specifically counterexamples to these statements.

A careful description of our study is found in Section 2. Section 3 contains statistical analysis discussing evidence of predictive assessment question types. A summary of our results together with concluding remarks is found in Section 4. Note that in a future note the authors will report results for a similar study from MATH-1105, Differential and Integral Calculus I.

Reasoning logically using counterexample is linked to the idea of conditional statements. We begin with a discussion about conditional statements and the notion of
counterexample itself and its importance in mathematical pedagogy, including a particular instance of a question type designed by Peter Wason \cite{W} called the Wason Selection Task. We refer to this as the Wason question in our tests.

1.1. Conditional Statements. A conditional statement consists of two parts, an antecedent (if-part) and a consequent (then-part). In general, these are of the form “If P then Q”, for some statements P and Q. An example would be the following:

If an integer is even then the rightmost digit of the number is either a 2, 4, 6 or 8.

An important component of logic and mathematics in general, is determining the truth value (True(T) or False(F)) of a given conditional statement. To that end it is sometimes helpful to reword conditionals using a universal quantifier. In general, “If P then Q” becomes “All P are Q” and specific to the example above:

All even integers have the rightmost digit being 2, 4, 6 or 8.

To falsify such a universally quantified statement is to demonstrate the existence of an object that satisfies statement P but does not satisfy statement Q; in our example, 10 is such an object. We call such an object a counterexample to the conditional statement. In fact, a conditional statement is false exactly when such an object exists. For more details, the reader is encouraged to refer to the introductory chapters of most textbooks in logic or discrete mathematics. It is important to see that the idea of counterexample is not necessarily intuitive. We note one more example from our diagnostic study to illustrate this point:

If two shapes have the same perimeter then they have the same area.

On the diagnostic test, 31 of 62 responded correctly indicating that the statement is false while only 8 of those justified their conclusion using a counterexample. Counterexamples provided by correct respondents most often used rectangles with few referring to circles or triangles.

1.2. The Importance of Counterexample. In mathematical reasoning, the notion of counterexample is fundamental. Counterexamples allow us to understand the limits of mathematical results and why each condition in a mathematical logical statement is needed to justify the conclusion. In \cite{K}, Klymchuck writes that, for undergraduate students, understanding by way of counterexamples “can improve students’ conceptual understanding in mathematics, reduce their common misconceptions, provide a broader view on the subject and enhance students’ critical thinking skills”. For students that plan on a career in education Vinsonhaler and Lynch \cite{VL} believe that primary and secondary school teachers “support student reasoning and thinking and promote productive struggle by incorporating counterexamples into the classroom.” Zazkis and Chernoff \cite{ZC} suggest that counterexamples can be particularly convincing when they are “in accord with individuals’ example spaces”.

1.3. The Wason Question. The card selection question was originally reported by Wason \cite{W}. In his experiment, there are cards on a table and the participant is told that there is a number on one side and a letter on the other. The participants are told of a rule, for example “If a card has a D on one side then it has a 3 on the other side” or, reworded as a universally quantified statement “All cards with a D on one side have a 3 on the other side”. Four cards are presented that correspond to each logical case (P true, P false, Q true, Q false) and the participant is asked “to select all those cards, but only those cards, which you would have to turn over in order to discover whether or not the rule has been violated.”

To answer the question, the participant is required to assess the conditional statement four times in relation to each card presented. The correct answer is to select the cards where either P is true or Q is false since these conditions are necessary (but not sufficient) for the card to be a counterexample or exception to the rule. Interestingly, most of the
population does not make this connection. Indeed, Inglis and Simpson in [IS] found from
an online, self-selected sample at the University of Warwick, only 29 percent of math
students, 8 percent of history students and 49 percent of mathematics staff responded
correctly. During our study described below, only 2 students answered this style of
question correctly on the pre-test; see §2.

1.4. Proving a conditional true. More difficult perhaps is the skill of reasoning the
justification of a true conditional. Consider the following sentence:

If the sum of two positive whole numbers is odd then their product is
even.

To justify the truth of this statement the participant must convey that, since the sum of
the numbers is odd, one number must be even while the other odd. This requires more
algebraic ability, and usually a considerable amount of lecturing to convey; indeed, many
universities have a course in the first or second year of the program (usually Discrete
Mathematics or a standalone Introduction to Proofs course) that spend considerable time
teaching the concept of correct proof writing. We note that the diagnostic test did not
include any questions of this type, however as part of the course material, some lecture
time was spent teaching how to prove a conditional statement for simple examples; e.g.
“If $a, b, c$ are three consecutive whole numbers then their sum has a factor of 3”.

As we will discuss in more detail, the ability to understand and construct counterex-
amples to a conditional statement is not inherent in most students entering Math 2903.
These results are not surprising in the context of similar statistical results found in [IS]
and [L]. However, perhaps more interesting is that on average, students who are able
to respond well to these problems have better course outcomes. We also observe that
students who learn these skills through in-course intervention experience better course
outcomes on average. As such, our study suggests that logical skill assessment can be
used as a tool to help improve course outcomes for students. Such assessment is also
effective in identifying struggling students early on so that they can find effective help.
Lastly, it is also quite interesting that a simple intervention can on average have a mea-
surable impact on student success.

2. The Study

The data used for this study comes from Math 2903, Math for Early/Middle Year
Teachers at the University of Winnipeg. We chose students in Math 2903 since it was,
most likely, the students’ first university mathematics class. Indeed, the likelihood that
our student populations in the class had been exposed to logical reasoning on an academic
level (for example, a course on critical thinking) was quite low. This is borne out in the
results.

The study was conducted in four phases. In the first week of classes (Phase I), stu-
dents were given a diagnostic test to benchmark their understanding of logical thinking.
In the third week (Phase II), students were given an intervention (75-minute lab/lecture
on the idea of conditional statements and counterexamples) complete with a worksheet
for practice. During the 6th week of classes (Phase III), a post test was given as part of
the in-class midterm exam. Final grades were tallied (Phase IV) and small sample size
statistical methods were used to analyze aggregate data collected. In what follows, we
give detailed descriptions of each phase of the study but leave Phase IV to §3.

2.1. I - The Diagnostic Test. The Diagnostic test was designed to gauge incoming
student ability to reason by counterexample and was given during the first week of class.
The problems appearing on the test were selected from 2 categories similar to those found
in [L]. We describe them now. We also note that specific questions can be found in §3.
2.1.1. Type I: The Wason selection task. As described in the introduction, this exercise concerns rule verification for a finite set of objects. A complete description of the Wason selection task can be found in both of [L] and [IS]. For this study, we depicted an array of 6 cards each with a number on one side and a letter on the other, see (1).

```
(1)

9 4 K A 3 Q
```

The student was told: “Each card has a number and letter on opposite sides. Every card with a vowel on one side has an even number on the other. Of the six cards below (it can be more than one) which must you flip over in order to determine whether or not the rule has been violated? Try your best to explain your reasoning.” Note that the correct answer to this problem is to flip the A, 9, and 3, as these could be possible counterexamples; for example, the A could have a 3 on the other side, and the 3 or 9 could have an A on the other side, and these cards would violate the rule).

2.1.2. Type II: Provide a counterexample to a given conditional statement. Here the student is asked to falsify a conditional statement by constructing a counterexample. For example, “If two shapes have the same perimeter then they have the same area” is seen to be false by considering two rectangles with dimensions $2 \times 4$ and $5 \times 1$. This forms the required counterexample.

2.2. II - The Intervention. The intervention consisted of a 75 minute primer on conditional statements that occurred during the second week of classes. During the intervention, students were introduced to the concept of a conditional statement, initially via a truth table, and the relationship between universally quantified statements and conditionals. Students were then shown an applied example of a liquor inspector enforcing the rule described by the conditional statement “If you are drinking alcohol then you must be 18 years of age or older” to reinforce the idea of when a conditional statement is false. This was then connected to the notion of counterexample. The intervention lecture closed with a discussion of the Wason selection task where students were shown the relationship between the truth values of the conditional sentence defining the rule and the values on each of the cards. This was followed by a group work session where students reinforced these ideas by completing a practice worksheet with their peers.

In addition to this 75 minute primer, as part of the course work students were given one lesson on verifying a conditional statement via an algebraic proof. The students were also given a small number of beginner practice problems, and one problem on an assignment to complete. Note that this type of question was not included on the original diagnostic test as algebraic verification is a technique that is more specialized and the instructor of the course (the first author) believed that most students would not have been able to answer the question appropriately.

2.3. III - The Post Test. The post test was given as part of the MATH-2903 class midterm test. The students were given one Wason Selection Task question, as well as one question where the students would have to identify which of two conditional statements had a counterexample, provide a counterexample for that statement, and algebraically justify the other conditional statement.
3. Data and Results

The aim of the data analysis found in this section is to compile and compare results of the diagnostic test (DT) and post test (PT). Our results, presented in tables 1-6, analyze relationships between success/retention/difficulty with specific problems on our tests with mean final course grades. In order to present this efficiently, we group and name specific problems of interest together with our raw data. Our data consists of evaluated student performance in terms of the following categories.

(i) **CR**: Correct answer and reasoning. This evaluation is given if the student provides a clear, complete solution to the problem providing a counterexample if necessary.
(ii) **C**: Correct answer with incorrect reasoning. The student was able to answer the problem correctly but lacked sufficient justification such as, for example, an erroneous or missing counterexample.
(iii) **I**: Incorrect answer. The student has answered incorrectly with or without justification.
(iv) **B**: The student left the problem blank.

3.1. The Diagnostic Test. The diagnostic test consisted of 5 questions of varying type that we state now followed by student response data.

- **DT:1** If a number \( n \) is less than 1, then the expression \( n^2 - 1 \) is negative.
- **DT:2** If two shapes have the same perimeter, then they have the same area.
- **DT:3** If the sum of two numbers is odd, then their product is even.
- **DT:4** If two numbers \( x, y \) have positive sum, then both \( x \) and \( y \) are positive.
- **DT:5** This is the Wason selection task problem exactly as given in section §\ref{sec:wason} with cards as presented in figure \( \text{[1]} \).

| Question | CR | C  | I  | B  |
|----------|----|----|----|----|
| DT:1     | 19 | 14 | 19 | 10 |
| DT:2     | 8  | 23 | 20 | 11 |
| DT:3     | 18 | 18 | 9  | 17 |
| DT:4     | 26 | 26 | 6  | 4  |

Table 1. Summary of student answers to counterexample questions on diagnostic test

For the Wason selection task problem DT:5, we analyze student responses more finely using common observed card selections. Category “A93” refers to the student selecting the correct cards, the ace, nine, and three. The other categories, “A”, “A4”, “All”, “Blank” and “Other” are self evident.

| Answers | A93 | A  | A4 | All | Blank | Other |
|---------|-----|----|----|-----|-------|-------|
| Question 2 | 3   | 2  | 26 | 12  | 10    | 9     |

Table 2. Summary of student answers to Wason Selection Task question on diagnostic test

**Remark 3.1.** We note that very few students correctly answered the question (A93) on the diagnostic test. Furthermore the most common answer was to select the A and four, (A4). This choice is the common fallacy that the conditional “if \( P \) then \( Q \)” is equivalent...
to “P and Q”. This result aligns with previous work by Inglis and Simpson [IS], and Wason [W].

3.2. The Post Test: The post test consisted of three questions PT:1, PT:2, PT:3 on the first midterm of Math 2903. Question PT:1 was the Wason selection task problem as in §2.1.1, see (I), with the altered rule

“If a card has an odd number on one side then it has a vowel on the other”.

The correct answer is to select the nine, three, K and Q (93KQ) as in (I).

Questions PT:2 and PT:3 appeared as two parts of a single problem that we state now.

One of the following conditional statements is always true and the other has a counterexample. ***Circle (a) or (b)*** of the the true statement and justify that it is true (using the direct method). Find a counterexample for the other statement (and explain why it is a counterexample).

- PT:2 (a) If m and n are whole numbers and \( m \times n \) is even, then both m and n are even.
- PT:3 (b) If \( n \) is an even number then \( n^2 \) is a multiple of 4.

The results of the post-test questions are shown in the table below. Students are grouped into two categories: correct, and otherwise.

| Question | Correct Answer | Correct | Incorrect |
|----------|----------------|---------|-----------|
| PT:1     | Selected (93KQ)| 21      | 38        |
| PT:2     | Provided a counterexample | 41 | 18 |
| PT:3     | Provided a correct proof | 12 | 47 |

Table 3. Number of students answering post-test questions correctly

Remark 3.2. In contrast with Remark 3.1 and noting that this specific question type was covered in the stage II intervention, it is very encouraging that a far larger proportion of students answered correctly. However, checking the truth of “P and Q” is still frequent.

3.3. Relationship between counterexample knowledge and final grade. Below are tables with mean final grades in Math 2903 grouped by their results on the diagnostic and post-tests, respectively. We used independent two-sample t-tests (equal variances not assumed) to determine if these groups means differ significantly. We note that this analysis is moot for problem PT:5 as most students answered incorrectly.

| Question | MFG (Correct/Justified) | MFG (Otherwise) | p-value |
|----------|-------------------------|-----------------|---------|
| DT:1     | 73.9                    | 66.5            | 0.072   |
| DT:2     | 81.2                    | 66.7            | 0.019   |
| DT:3     | 75.4                    | 66.1            | 0.029   |
| DT:4     | 70.0                    | 67.9            | 0.582   |
| DT:5     | *                       | *               | *       |

Table 4. Comparison of mean final grade (MFG) of students by diagnostic question success
Table 5. Comparison of mean final grade (MFG) of students by post-test question success

| Question | MFG (Correct/Justified) | MFG (Otherwise) | p-value |
|----------|-------------------------|-----------------|---------|
| PT:1     | 78.0                    | 64.2            | <0.001  |
| PT:2     | 70.0                    | 67.1            | 0.471   |
| PT:3     | 82.3                    | 65.8            | <0.001  |

To aid in answering the question of whether learning logic and counterexample affects student success, for each of the parts of DT:1-DT:4 on the Math 2903 diagnostic test we looked at the mean final grades for the students who did not answer the question completely correctly (not both correct and justified). We then compared the mean final grades of those answering the PT:2 on the post-test question correctly (CR: correct and justified) and those that did not. We again performed an independent two-sample t-test (equal variances not assumed) to determine if these groups means differ significantly. The results are in the table below.

Table 6. Comparison of mean final grade (MFG) of students with incorrect diagnostic test answers by post-test question success

| Incorrect | MFG (PT:2 CR) | MFG (Otherwise) | p-value |
|-----------|---------------|-----------------|---------|
| DT:1      | 69.1          | 61.6            | 0.104   |
| DT:2      | 67.2          | 65.9            | 0.76    |
| DT:3      | 68.7          | 60.2            | 0.053   |
| DT:4      | 68.9          | 66.1            | 0.610   |

4. Conclusions

We set out to study the effect of minimal logical training on student outcomes. Our data shows interesting trends associated to the effect of understanding the idea of counterexamples in logic. Through our investigations we found that certain problems types can be used as checks for students less likely to find success in their mathematics course. We now examine these ideas specifically in relation to our data and then once again in the context of our perspective on learning mathematics in general.

From the data above, we can reasonably conclude two important things: first, that very few students have a strong idea of counterexample entering Math 2903, though many do have some intuition about whether a statement is true or false, without the skills to definitively show why. Indeed, from Table 2 it can be seen that only 2 out of 60 students chose the correct answer, and the plurality of students checking both P true and Q true. This is in line with the findings presented in IS. Additionally, only 8 of 62 students answered DT:2 correctly.

From the midterm data, it is worth noting that these pre-service teachers can learn logical reasoning and counterexample. Given the data for the post-test, it is clear that students that do successfully learn the idea of logic and counterexamples have more success in the course overall than students that do not. From Table 3 we note that while the Wason question (PT:1) was worth approximately 1 per cent of the students’ final grade, students that do answer the question correctly do, on average, 13.8 percentage points better in the course than those that do not. We emphasize that this question tests exactly one logical idea, without any computational or algebraic skills needed. Indeed, the Wason Selection Task seems to be a highly discriminant question. More research is
needed to determine whether this logical knowledge is a driver of higher course grades, or whether stronger students simply learn these ideas more quickly than other students.

Question PT2, where the student needed to supply a counterexample, was not very predictive. We believe this is the case since the question may have been too intuitive, though students who did not understand counterexamples in the beginning of the course and did indeed answer this question correctly seemed to have a significant, or near-significant, difference in mean grades as opposed to students who did not learn the idea of counterexample. We conjecture that it is important to understand the idea of counterexample at its core and not just intuitively for easy cases, as a central idea of mathematics is using logical and mathematical techniques to understand non-intuitive problems. Indeed, if we chose a slightly more challenging counterexample problem it is possible that we would have seen more significant differences in Table 6.

Question PT3 was highly predictive. We are not surprised by this data; the skills to write a correct proof mean that you understand the statement and the associated objects it deals with in generality, and have the computational and algebraic skills necessary to complete this proof.

In our comparison of mean final grades of students that did not answer the first four diagnostic questions correctly, we find that, while the mean final grades of the students who did answer PT:2 correctly was higher for all diagnostic test questions, we find that at $\alpha = 0.05$ none of these means were significantly higher than the groups of students that answered PT:2 incorrectly. However, we would conjecture that we may have found a significant difference in mean final grade if PT:2 was a more difficult question.

Putting this all together, we believe that the data supports the current research that counterexample plays a very strong role in mathematical understanding in early undergraduate mathematics. Indeed, a student gaining understanding of the idea significantly improves student success in their respective course.

REFERENCES

[BF] C. Bardelle and P. L. Ferrari, Definitions and examples in elementary calculus: the case of monotonicity of functions, ZDM Mathematics Education 2011 43:233-246, DOI 10.1007/s11858-101-0303-4

[W] P. C. Wason, Reasoning about a rule, Quarterly Journal of Experimental Psychology 1968 Vol. 20 Is. 3:273-281

[K] S. Klymchuk, Using Counter-Examples to Enhance Learners’ Understanding of Undergraduate Mathematics 2009, Good Practice Publication Grants New Zealand: Ako Aotearoa National Centre on Tertiary Teaching Excellence (Online)

[VL] R. Vinsonhaler and A. G. Lynch, Students' Understanding of Counterexamples 2020, Mathematics Teacher: Learning and Teaching PK-12 MTLT, Vol. 113 Iss. 3:223-228

[DO] B. Divjak and D. Oreski, Prediction of Academic Performance Using Discriminant Analysis, Proceedings of the ITI 2009 31st Int. Conf. on Information Technology Interfaces, June 22-25, 2009, Cavtat, Croatia

[IS] A. Inglis and M. Simpson, Mathematics and the selection task, Proceedings of the 28th Conference of the International Group for the Psychology of Mathematics Education, 2004

[KBT] F. Bagheri, O. Kara, and T. Tolun, Factors Affecting Students’ Grades In Principles Of Economics, Amer. J. Bus. Ed. 2009 Vol. 2 No.7:25-34

[L] K. Lee, Student’s logical reasoning and mathematical proving of implications, PhD thesis, Michigan State University, 2011

[ZC] R. Zazkis, E.J. Chernoff, What makes a counterexample exemplary?, Educational Studies in Mathematics, Volume 68:195-208