Measuring the Effects of a Thinking Classroom on Students' Problem-Solving Ability

Paul Gitchos

A Capstone Presented to the Teachers College Faculty

of Western Governors University

in Partial Fulfillment of the Requirements for the Degree

Master of Science, Curriculum and Instruction

October, 2021

Abstract

The author conducted action research in a grade 12 AP Calculus class to study the effect of the Thinking Classroom instructional model on students' problem-solving ability. Sixteen non-routine problems were selected as an observational framework, and were administered to a randomly selected subset consisting of half the students; a Rasch analysis of these results measured the person-item interaction, revealing the difficulty level of each problem. The remaining students completed eight of the problems selected at random as a pre-assessment. Following an instructional intervention consisting of ten hours of Thinking Classroom instruction, in which students worked collaboratively on rich tasks, the same students completed the other eight problems as a post-assessment. Using the difficulty levels determined previously, a test-equating procedure provided a measure of problem-solving ability for students before and after the instructional intervention. An increase in problem-solving ability was measured for all students in this group, supporting a conclusion that students were better problem-solvers after the Thinking Classroom instruction than they were before.

Table of Contents

Chapter 1: Topic and Problem

Topic

This paper describes an action research study into the effect of the Thinking Classroom instructional model on students' non-routine problem-solving skills, in a grade 12 AP Calculus class. This topic addresses a central challenge in mathematics curriculum and instruction: when students are busy learning new content, in a prescribed curriculum, how can in-class instruction best support the development of their thinking and problem-solving skills?

Problem Statement

A problem I notice in teaching AP Calculus is that many students, even those who demonstrate strong content understanding, fail to improve their non-routine problem-solving skills. As a result, students have difficulty applying their new content knowledge in challenging or unfamiliar contexts. This problem may exist because the AP Calculus course contains a significant amount of new content learning; students often spend the majority of their time both in- and outside of class working to understand new concepts and master new procedural skills, and as a result non-routine problem-solving receives less attention.

Problem Impact and Root Cause

A constructivist theory of learning, and the desire to improve students' problem-solving ability, motivate a focus on rich, challenging tasks. However, in a content-intensive course such as AP Calculus, class time has traditionally been devoted to the introduction or "coverage" of new content. Students learn difficult mathematical procedures, and must practice them to become proficient. In such a setting we risk our classes devolving into the dystopian vision of math education described by Paul Lockhart in his influential essay "A Mathematician's Lament:"

> The main problem with school mathematics is that there are no problems. Oh, I know what passes for problems in math classes, these insipid "exercises." "Here is a type of problem. Here is how to solve it. Yes it will be on the test. Do exercises 1-35 odd for homework." What a sad way to learn mathematics: to be a trained chimpanzee. But a problem, a genuine honest-to-goodness natural human question— that's another thing.... A good problem is something you don't know *how* to solve. (2009, p. 9)

Potential impacts of this situation include not only a lack of growth in problem-solving skills, but student dissatisfaction or lack of interest in mathematics, as well.

When I teach AP Calculus, I sometimes feel pressure to make a trade-off between content coverage and skill mastery, on the one hand, and thinking and problem-solving, on the other. The Thinking Classroom model was designed to improve students' engagement with collaborative problem solving. If students are so engaged, they should improve their problem-solving skills, since, in a Thinking Classroom, "problem-solving becomes not only a means but also an end. A thinking classroom is shot through with rich problems" (Liljedahl, 2016, p. 384). Of particular interest is increasing student success with problems that are *non-routine*; that is, students will possess "neither a known answer nor a previously established (routine) procedure for finding one." (Malone et al., 2007, p. 187). Improving students' ability at non-routine problem-solving is the primary motivation of this study.

Research Question

What is the effect on student problem-solving ability of implementing the Thinking Classroom model in grade 12 AP Calculus?

Justification

In addition to teaching content mastery, math classes should develop students' ability to grapple with and solve problems that are unfamiliar. Teachers consistently assess content knowledge, but do not typically measure students' problem-solving ability. The Thinking Classroom model was developed explicitly to improve students' critical thinking and problem-solving skills. This study sets out to attempt to measure the efficacy of this instructional model.

Chapter 2: Review of the Literature

Introduction

Problem-solving has long been an important focus of math education, and the rise of the constructivist theory of learning has only strengthened and made explicit the value of problem-solving. Along with mathematical content and strategies, a growth mindset contributes to the development of problem-solving ability. Recently, the Thinking Classroom model was designed through an effort to maximize the quantity and quality of problem-solving activities. This instructional model prescribes a set of elements of classroom design and instruction which encourage critical thinking and collaborative problem-solving.

The precise definition and quantitative measurement of students' problem-solving ability pose some unique challenges not found in the assessment and measurement of students' math content knowledge. One possible solution stems from the psychometric field of measurement pioneered by George Rasch. By defining a unidimensional variable and utilizing a Rasch model for measurement, it should be possible to quantify students' problem-solving ability at two distinct points in time, in an effort to measure improvement.

Constructivism and Problem-Solving in Mathematics Education

In the late twentieth century, the classical behaviorist model of mathematics education, wherein students observe direct instruction to learn to execute procedures in pursuit of an extrinsic motivation (typically a grade, in secondary and collegiate settings), was supplanted by a student-centered constructivist theory of learning (Steffe & Kieren, 1994; Thompson, 2013; Faulkenberry & Faulkenberry, 2006). Rather than deliver content via lecture, teachers working in the constructivist paradigm sought to develop classroom activities that would engage student thinking and lead students to construct new mathematical knowledge. The transition was not without controversy; opposing views of the effectiveness of this instructional style, along with the inclusion of constructivist language in NCTM's 1989 *Standards*, led to the "Math Wars" of the 1990's (Schoenfeld, 2003). Nonetheless, by the second decade of the 21st century most mathematics education research findings were consistent with the constructivist theory of learning, leading education historian Patrick W. Thompson to conclude, "Constructivism is now taken for granted" (2013, p. 6).

Mathematical problem-solving resides at the center of constructivist math education. Mathematics educators have long understood the importance of challenging students to think, struggle, and solve problems in unfamiliar contexts; problem-solving "has been of interest to mathematics education researchers for as long as our field has existed" (Liljedahl et al., 2016, p. 1). At least since the 1945 publication of George Polya's *How to Solve It*, math teachers have acknowledged that problem-solving skills can be inculcated in students alongside math content knowledge, and that teaching math and teaching problem-solving are complementary, but not strictly equivalent. Polya was explicit in his description of problem-solving, describing four distinct phases: Understanding the Problem, Devising a Plan, Carrying out the Plan, and Looking Back (2009, p. 5-6). Alan Schoenfeld refined and expanded upon Polya's schema, describing the role of students' prior knowledge and individual points of view; like Polya, he believed that math teachers could and should help their students become better problem solvers (1985). Mayer (1998) characterized problem-solving as consisting of three essential components: cognition ("skill"), metacognition ("metaskill", or knowledge transfer) and motivation ("will"). The Common Core State Standards for Mathematics, published over 60 years after Polya's seminal work, situate problem-solving in a place of highest importance, first among the standards for Mathematical Practice:

> MP1: Make Sense of Problems and Persevere in Solving Them.
>
> Mathematically proficient students start by explaining to themselves the meaning of a problem and looking for entry points to its solution. They analyze givens, constraints, relationships, and goals. They make conjectures about the form and meaning of the solution and plan a solution pathway rather than simply jumping into a solution attempt. They consider analogous problems, and try special cases and simpler forms of the original problem in order to gain insight into its solution. They monitor and evaluate their progress and change course if necessary (2010).

Like constructivism, the importance of problem-solving in mathematics education is now taken for granted.

Problem Solving and Mindset

The development of students' problem-solving ability is related to, and impacted by, students' attitudes and mindsets. Carol Dweck's book *Mindset: the New Psychology of Success*, published in 2007, helped many educators understand the importance of emphasizing every student's potential for improvement, as a key to achievement. Growth mindset, the attitude that everyone can improve through effort and coaching, is related to self-efficacy, the belief that one's own success is possible (Bandura, 1997). Math classrooms which emphasize deep understanding, communicating one's thinking, and improvement over time are more likely to foster growth mindsets in students (Dweck, 2007, p. 219-220). Researchers have since shown a connection between having a growth mindset and succeeding in math, particularly for female students (Degol et al., 2018).

Thinking Classroom

In an effort to make problem-solving the central activity of secondary math classrooms, Peter Liljedahl spent a decade evaluating classroom practices and their relative effects on student engagement and thinking. The goal of his research was to specify conditions for what he terms a Thinking Classroom: "a classroom that is not only conducive to thinking but also occasions thinking, a space that is inhabited by thinking individuals as well as individuals thinking collectively, learning together and constructing knowledge and understanding through activity and discussion" (2016, p. 362). Liljedahl describes a set of elements which he has found to contribute to a Thinking Classroom. These include both details of classroom design, such as visibly random grouping, vertical non-permanent surfaces, and de-fronting the classroom, as well as teaching practices, such as giving task instructions verbally and answering only those student questions which promote continued progress. Of central importance to the Thinking Classroom are engaging problem-solving tasks, which students work on collaboratively. As Liljedahl states in the conclusion, in a Thinking Classroom, "problem-solving becomes not only a means but also an end. A thinking classroom is shot through with rich problems" (p. 384).

Defining and Measuring Problem-Solving Ability

In "A Mathematician's Lament," Paul Lockhart criticizes math education by contrasting the repetitive practice of procedural math skills with genuine problem-solving, asserting that "a good problem is something you don't know *how* to solve" (2009, p. 9). More formally, Mayer (1992) has defined problem-solving as a cognitive process directed toward an end goal, when the procedure or method is not known in advance by the solver. For both Lockhart and Mayer, as for Polya, genuine problem-solving requires engagement with problems that are non-routine; that is, the solver possesses “neither a known answer nor a previously established (routine) procedure for finding one. (An implication is that the student has not previously attempted the problem or one very similar to it)” (Malone et al., 2007, p. 187).

In contrast to the assessment of a procedural skill (e.g., the number of two-digit by two-digit multiplication problems, or first derivatives, done correctly in 5 minutes), the requirement that problems be non-routine creates a unique challenge in the measurement of problem-solving skill. In alignment with our belief in the importance of a growth mindset, we would like to measure the change in a student’s problem-solving ability over time, but in order to be non-routine the pre-assessment and post-assessment problems must all be different! This is a problem of measurement, and a potential solution arises from the work of Danish mathematician George Rasch (1960) and extended by Andrich (1978) and Masters (1982). For the purposes of measurement, one can construct a unidimensional variable representing non-routine problem-solving ability, and specify as an observational framework a set of non-routine math problems of varying difficulty levels. As Andrich explains, “every human performance, action, or belief is complex and involves a multitude of component abilities… Nevertheless, there are circumstances in which it is considered useful to think of concepts in unidimensional terms” (1988, p. 9). With this constructed variable and observational framework defined, a Rasch model relates the problem-solvers (“persons”) and non-routine problems (“items”) using a single numerical scale; it transforms observations (student performances on a set of problems) into measurements (person ability and item difficulty). This mathematical technique provides the ability to calibrate non-routine problems in order to measure individual student problem-solving skill at multiple points in time.

Conclusion

As mathematics educators, we want our students to acquire new math content knowledge, and we also want them to improve their problem-solving skills. As students learn more math, they gain new problem-solving tools; with the proper guidance they can continue to strengthen their critical thinking and problem-solving ability, as well. The Thinking Classroom model centers rich, challenging tasks to create an instructional environment which positively impacts students' problem-solving skills. The belief that students' problem-solving skills can be developed over time resides at the intersection of a constructivist theory of math learning and a growth mindset. Using a Rasch model, it is possible to measure the problem-solving ability of students at multiple points in time. Our ultimate goal is not merely for students to master all of the school math content, but for them to grow as autonomous, resilient, confident problem solvers.

Chapter 3: Research Methodology

Research Question

What is the effect on student problem-solving ability of implementing the Thinking Classroom model in grade 12 AP Calculus?

Participants

Participants in the action research study were grade 12 students enrolled in my AP Calculus class in the fall of 2021; there were 18 students in the class. These students passed a prerequisite Precalculus course last year, indicating their past success at learning math content.

Data Collection Instruments and Methods

Quantifying problem-solving ability, and gathering student data both before and after an instructional intervention, are appropriate to determine the impact of the intervention on student problem-solving ability. The research instruments were composed of one pre- and one post-assessment, each consisting of eight non-routine math problems for students to solve. The problems were created or chosen so that participants had already acquired the necessary math content knowledge. Between the pre- and post-assessments, students participated in ten hours of class time conducted using the Thinking Classroom model.

Because it was necessary to present students with problems that were unfamiliar, the pre- and post-assessment problem sets needed to be comprised of different problems, and yet yield comparable data. Using a polytomous Rasch model, problems were calibrated using data from a subset of students, and assigned difficulty scores. This allowed for "test-equating," to generate pre- and post-assessment measurements of each student.

Using a randomization technique, the participants were divided into two approximately even-sized groups, A and B. Group A's responses were used to calibrate the problems via a Rasch model for measurement. Group B students were the subjects of the study.

The list of 16 non-routine problems (see Appendix B1) was divided randomly into two sets of 8 problems; these were the pre- and post-assessments. Some of these problems are from Malone et al. (2007) and some are from the Ohio digital mathematics project (2010).

For both the pre- and post-assessment, each students' work on each problem was evaluated using a four-point rubric, adapted from Malone et al. (2007). The rubric is reproduced in Appendix B2.

The rubric scores from group A students' pre- and post-assessments were analyzed using Rasch measurement to generate problem difficulty levels ("locations"). Then, group B students' rubric scores, along with the calibrated difficulty levels of the problems, were used to generate a problem-solving ability score for each student in group B from before and after the intervention.

Data Analysis Techniques

The data collection procedures described above were designed to yield a problem-solving ability score for each student in group B from before and after the unit of instruction. A comparison of these scores were used to address the research question: What effect did the instructional intervention have on students' problem-solving ability?

Timeline

I administered the pre-assessment, which required about one hour of class time, to groups A and B, and the post-assessment to group A only, during the last week of August, 2021. Following the instructional intervention, I administered the post-assessment to group B during September, 2021.

Resources

The Thinking Classroom model requires a sufficient number of vertical non-permanent surfaces (such as whiteboards) to accommodate all students in groups of three.

Analyzing the data from this study using Rasch measurement was accomplished using the RUMM 2030 software, provided courtesy of David Andrich (RUMM Laboratory, 2021).

Conclusion

A set of sixteen non-routine math problems was calibrated using Rasch measurement, and divided randomly into two sets to create a pre- and a post-assessment. The pre-assessment was administered to participants. After an instructional intervention consisting of ten hours of class time conducted using the Thinking Classroom model, the post-assessment was administered to participants. The research study yielded a problem-solving ability score for each student from before and after the instructional intervention, which were compared to determine the intervention's impact.

Chapter 4: Results

Summary of Research

In order to determine the effect on student problem-solving ability of implementing the Thinking Classroom model in grade 12 AP Calculus, I administered one set of non-routine problems to students at the beginning of the school year, and another set of problems after 10 hours of Thinking Classroom instruction. Both sets of problems were administered at the start of the year to a different set of students (Group A), whose responses were used to calibrate problem difficulty.

The Rasch analysis of Group A's responses provided a measurement of difficulty ("item location") for each problem, ranging from -4.188 to 6.405; zero is assigned as the defacto mean of item location. A higher location indicates a more difficult problem. One problem was determined by the analysis to be outside the parameters of measurement ("Extreme") because of an unusually low location, and was excluded. A table showing item location for the remaining 15 problems appears in Figure 1.

Figure 1. Item Location (*RUMM 2030*)

Item	Type	Location	SE	FitResid	DF	ChiSq	DF	Prob
I0002	Poly	-4.188	0.434	1.134	4.47	15.446	2	0.000443
I0003	Poly	-3.148	0.289	0.094	4.47	0.676	2	0.713069
I0004	Poly	-3.194	0.324	0.016	4.47	4.633	2	0.098602
I0005	Poly	-2.849	0.312	0.081	4.47	2.089	2	0.351795
I0006	Poly	-0.854	0.219	0.017	4.47	0.360	2	0.835345
I0007	Poly	-0.867	0.265	0.455	4.47	2.492	2	0.287618
I0008	Poly	-0.721	0.304	0.036	4.47	1.738	2	0.419446
I0009	Poly	2.295	0.598	0.806	4.47	1.155	2	0.561374
I0010	Poly	-0.352	0.250	1.061	4.47	0.314	2	0.854716
I0011	Poly	1.730	0.553	0.526	4.47	0.546	2	0.761024
I0012	Poly	-0.748	0.259	0.067	4.47	1.378	2	0.502179
I0013	Poly	-1.210	0.249	0.002	4.47	2.538	2	0.281056
I0014	Poly	6.405	0.949	1.092	4.47	8.972	2	0.011265
I0015	Poly	1.719	0.556	0.698	4.47	0.882	2	0.643256
I0016	Poly	5.982	0.698	-0.009	4.47	0.539	2	0.763747

The Rasch model for measurement quantifies the problem-solving skill of students ("person location") using the same scale as problem difficulty ("item location"). The equivalence of the scales is summarized by the following statement, which is axiomatic: If a person's location and an item's location are equal, the model posits that it would be equally likely for the person to succeed or fail at the item. (The succeed/fail model, which is dichotomous, is extended in the polytomous case here, since there are five possible rubric scores for each problem.) As the person's location increases relative to an item's location, the probability of the person succeeding at the item increases.

An essential goal of the Rasch model is "the careful construction and maintenance of invariant linear measures" (Linacre J.M. & Wright B.D, p. 54). The unit of measure within the Rasch model is the log-odds ratio, or logit, defined as $\log_e(\frac{Probability\ of\ success}{Probability\ of\ failure})$. If a person has an equal chance of success or failure at an item, this results in a person location of $\log_e(1)=0$, or equal to that of the item, as stated above. A person who has a 75% chance of success would have a location of $\log_e(\frac{.75}{.25})=1.099$ logits greater than the item location.

The locations of students in Group A ranged from -1.357 to 3.241, with a mean of -0.254 and a standard deviation of 1.351. No persons were found to be Extreme by the analysis. (This would have indicated a near-perfect or near-zero total score.) The distribution of person and item locations (9 Group A students, and 15 non-routine problems) is shown in Figure 2. Summary statistics for the Rasch analysis are shown in Figure 3. Overall fit of the Rasch model for this set of persons and items was determined to be Excellent.

Figure 2. Person-Item Location Distribution (*RUMM 2030*)

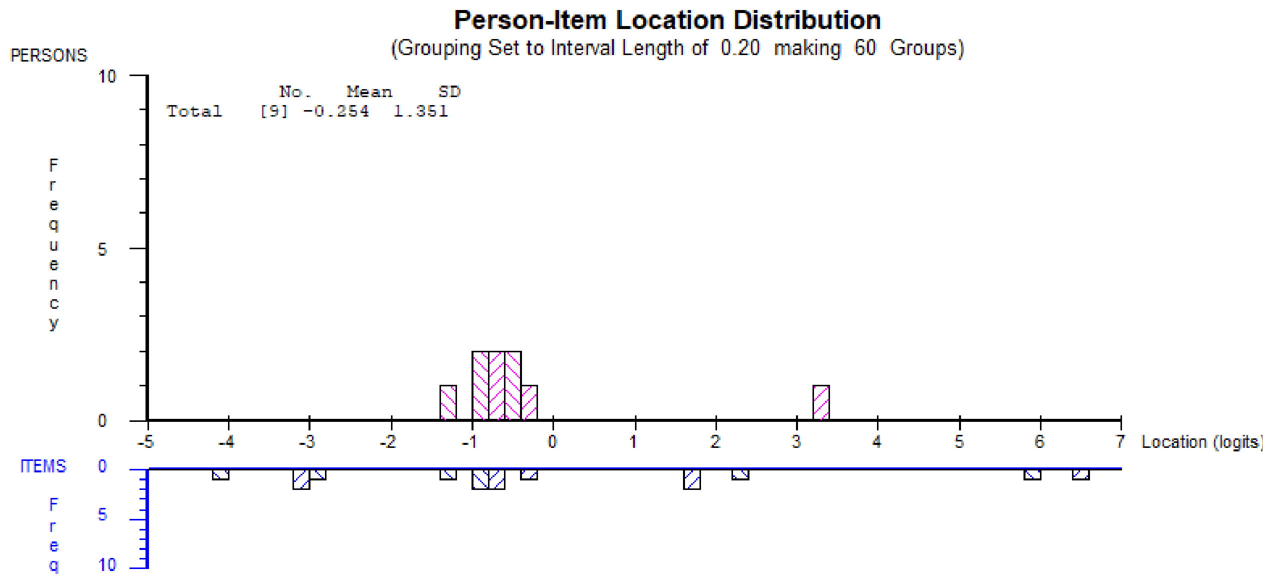
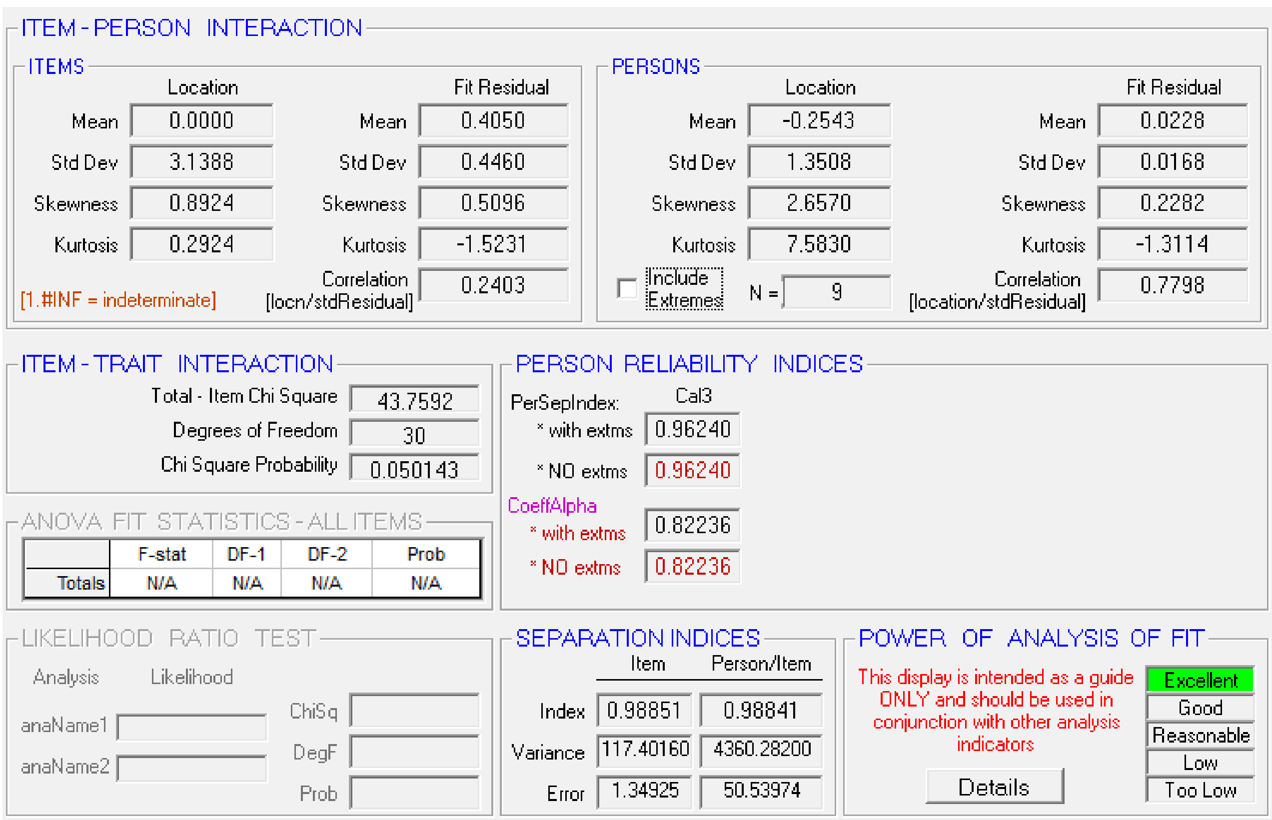
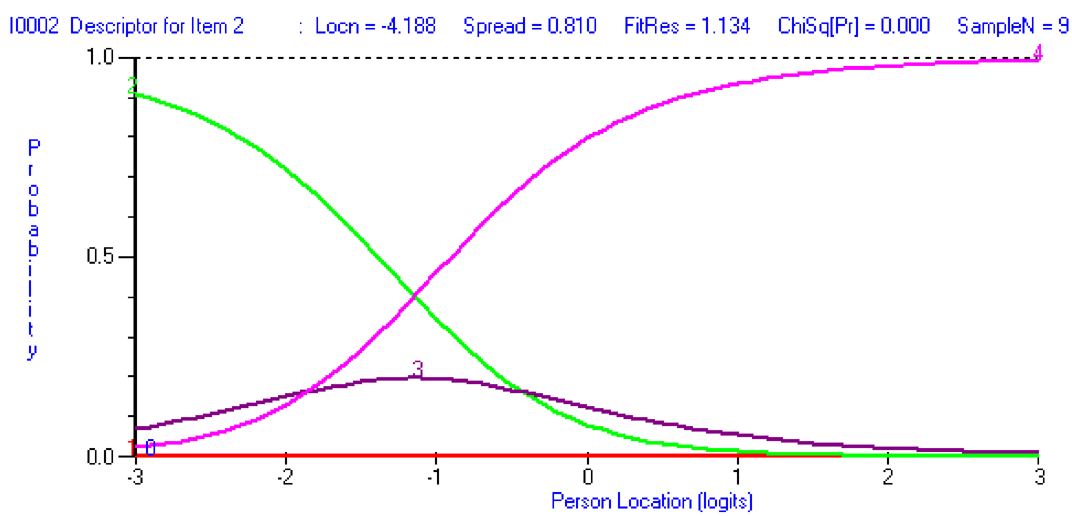


Figure 3. Summary Statistics for Person-Item Interaction (Calibration) (*RUMM 2030*)



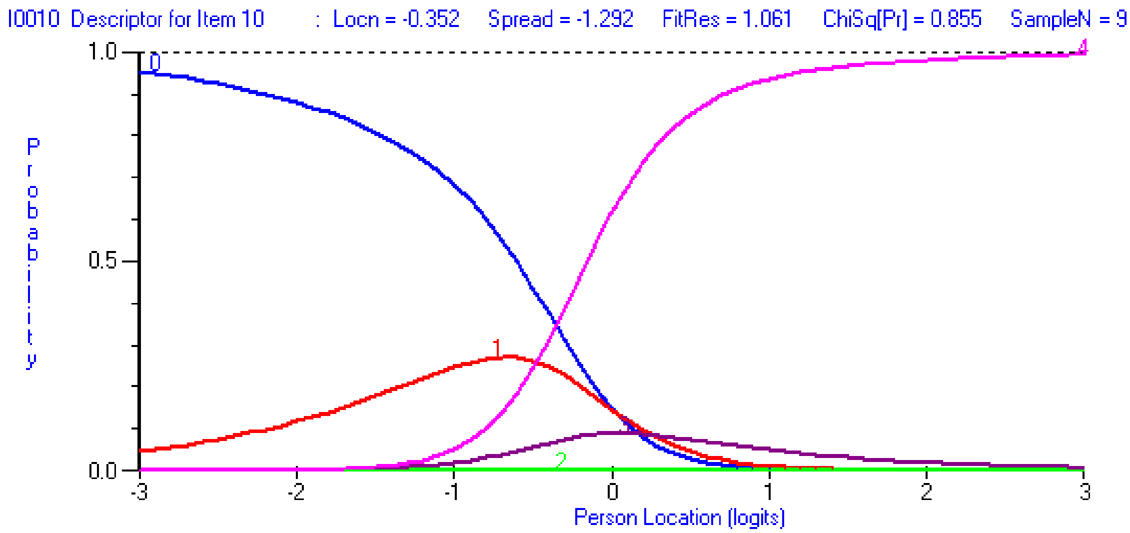
The Rasch model generates probability analyses for each item, which describe the person-item interaction. Figure 4 shows the Item Characteristic Curve for Problem 2 from the Pre-assessment. The *x*-axis represents the location of a person attempting the problem, and the *y*-axis is probability. The ogee curves represent the scores 2 and 4; the low bell-shaped curve represents a score of 3. (Scores of 0 and 1 on this problem were not represented in the data set.) As a person's problem-solving ability increases, the likelihood of a higher score increases. The overall location of the item is -4.188, indicating that this is an easy problem. The ICC shows that a person of location 0 (average, relative to the items analyzed) would have about an 80% chance of scoring a 4 on this problem, around a 10% chance of scoring a 3, and a slightly lower chance of scoring a 2.

Figure 4. ICC for Problem 2 (*RUMM 2030*)



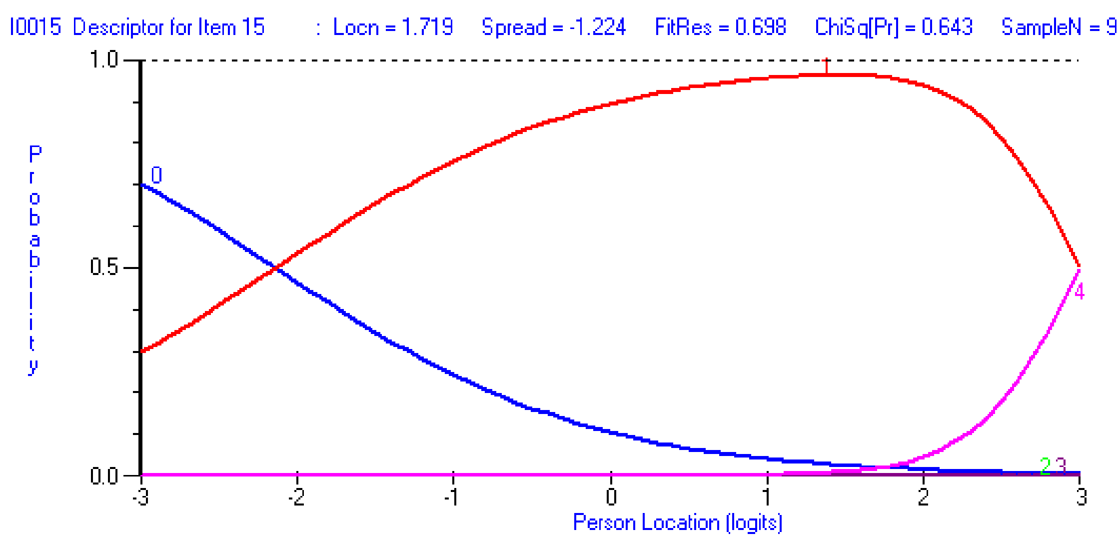
For comparison Figure 5 shows an ICC for problem 10, which was more difficult with a location of -0.352. Here a person with location 0 would have around a 60% chance of scoring a 4, and about a 15% chance of scoring a 0 and a 15% chance of scoring a 1.

Figure 5: ICC for Problem 10 (*RUMM 2030*)



The Item Characteristic Curve for problem 15 is shown in Figure 6. With a location of 1.719, this problem was quite a bit more difficult. The ICC indicates that a person with location 0 would have a greater than 80% chance of scoring a 1.

Figure 6: ICC for Problem 15 (*RUMM 2030*)



Using the individual item locations, each student's overall score on the pre- or post-assessment can be converted into a measurement of problem-solving ability ("person location"). This method of test-equating works because total raw score is a sufficient statistic, as described by R.A. Fisher, for person location (Linacre, 1992; see also Andrich, 1988, p. 35-39). Figures 7 and 8 show the total score to location conversion scales for the pre- and post-assessments. Each Group B student response was scored 0-4 using the rubric in Appendix B2. Since one Extreme item was excluded, the total possible scores were 28 on the pre-assessment and 32 on the post-assessment. Note that, importantly, the Rasch analysis takes problem difficulty into account when assigning person locations. By chance, the post-assessment (Items 9-16) contained the most difficult problems and was overall much harder (see Figure 2 above, and further discussion below). As a result, as shown in Figures 7 and 8, an overall score of 17 on the pre-assessment would indicate a person location of -0.633, but on the post-assessment it would indicate a location of 2.429.

Figure 7. Score to Location Conversion for Pre-assessment (*RUMM 2030***)**

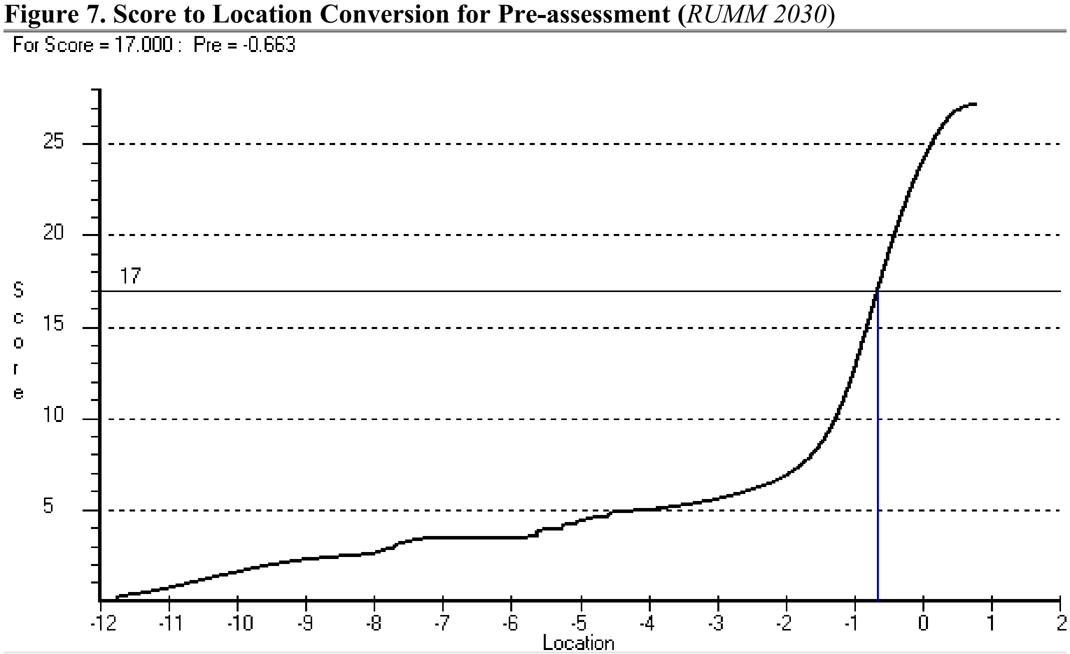
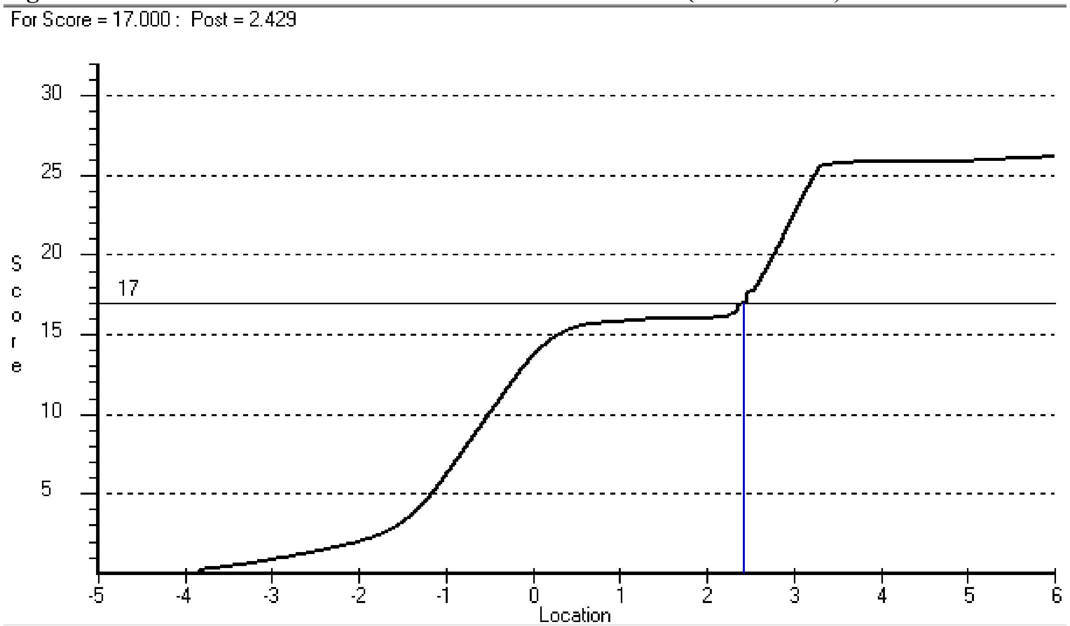


Figure 8. Score to Location Conversion for Post-assessment (*RUMM 2030*)



Summary of Results

The pre- and post-assessment scores, along with the corresponding measurements of problem-solving ability ("location"), appear in Figure 9. The location of every student was higher after the instructional intervention. The mean location of the Group B students at the time of the pre-assessment was -0.546, somewhat below Group A (-0.254), with a standard deviation of 0.414. The mean location of the Group B students at the time of the post-assessment was 2.284, with a standard deviation of 2.675. The mean location for Group B was 2.830 logits greater after the intervention, than before. The least increase in location by an individual student was 0.18, and the greatest increase by an individual student was 8.903. The mean of the differences of student locations from before and after the instructional intervention – in other words, the average amount of improvement by each student – was 2.830 logits.

Figure 9: Student Pre- and Post-assessment Data

Student ID	Pre-assessment score (0-28)	Pre-assessment location	Post-assessment score (0-32)	Post-assessment location
11	23	-0.125	28	8.778
12	18	-0.588	18	2.559
13	17	-0.633	16	1.358
14	23	-0.125	28	0.055
15	23	-0.125	18	2.559
16	18	-0.588	22	2.951
17	13	-0.977	24	3.134
18	21	-0.33	14	0.055
19	9	-1.424	7	-0.897

A histogram showing the distribution of Group B student locations before and after the intervention appears in Figure 10. Box and whisker plots of Group B student locations, measured by the pre- and post-assessments, shown in Figure 11, depict increases in the summary statistics. These charts show both a universal increase for students in the group (min, max, mean and median all increased), and a widening of the spread of students' locations (range, interquartile range, standard deviation increased), from before the intervention to after.

Figure 10: Distribution of Group B Locations

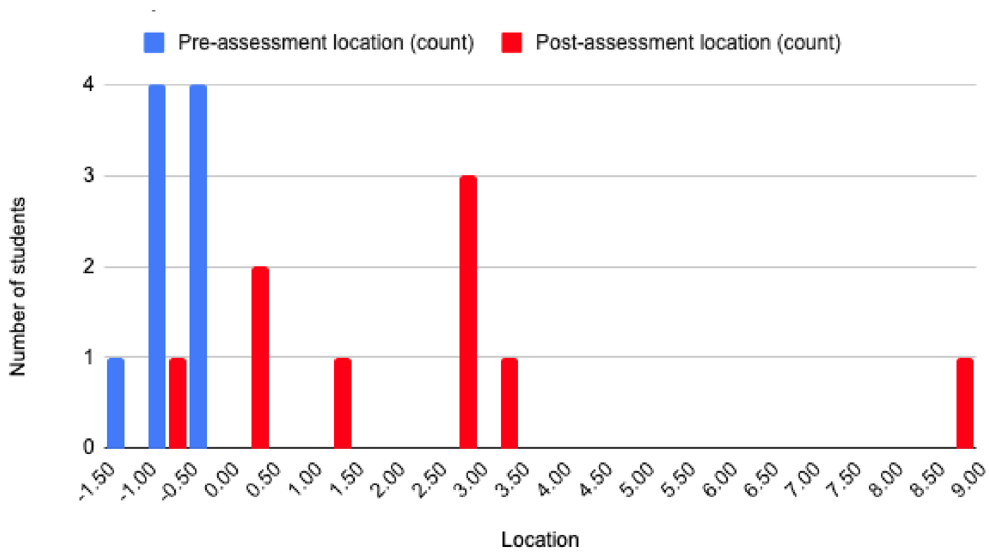
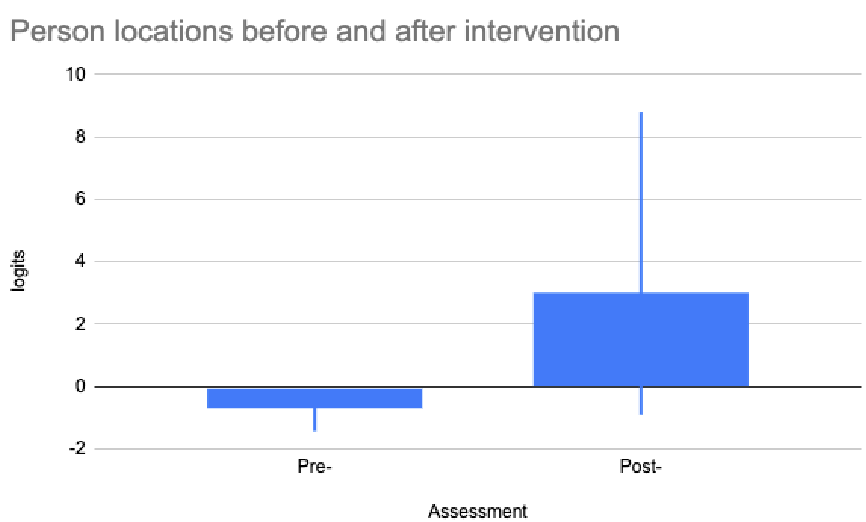


Figure 11. Box plots of Group B locations



Details of Implementation

Before the 2021-22 school year began, I used a 20-sided die to randomly divide the 18 students enrolled in my AP Calculus class into Group A and Group B. I selected 16 non-routine problems and labeled them with the letters A through P; I used Scrabble tiles to randomly select 8 of these to comprise the pre-assessment, leaving 8 for the post-assessment. Appendix B1 contains the original list (with problems denoted by letters A-P), and the two assessments. I also created an instructional plan, and a list of rich tasks for use during Thinking Classroom instruction; these appear in Appendix A.

On the second class meeting of the school year, I administered the pre-assessment to all students, and the post-assessment to Group A students and a precalculus Algebra review to Group B students. I secured the pre- and post-assessments in a locked cabinet in my classroom.

During the next two class meetings, I used the Thinking Classroom model to engage students with problem-solving tasks, and to establish collaborative work at vertical whiteboards as a classroom routine. In those two 80-minute blocks, students worked on the square peg problem (problem i in Appendix A), the goat and barn (problem xii), the half-size cylinder (viii), and reviewed unit circle trigonometry with problem xi. Following these introductory classes, we moved into the first unit of Calculus content. From this point on, students were responsible outside of class for watching videos introducing new content, and answering a few relevant questions as formative assessment, each week. This "flipped classroom" model allowed the majority of class time to be spent on collaborative problem-solving, rather than lecture for content delivery.

During the next seven class meetings, the Thinking Classroom model was used to introduce students to the key ideas that appear at the beginning of a Calculus course, and to review algebra content, while emphasizing problem-solving skills. In addition to problems vi, vii, x, and xiv from Appendix A, students worked in the Thinking Classroom style on a variety of problems involving limits expressed graphically, continuity, and the Intermediate Value Theorem. There was a total of ten hours of instruction delivered via the Thinking Classroom model.

Following the intervention, I administered the post-assessment to students in Group B, and the precalculus algebra review to students in Group A. Using the rubric in Appendix B2, I scored all of the students' pre- and post-assessment, evaluating one problem at a time (all responses to problem 1, then all responses to problem 2, an so on).

Answer to the Research Question

The universal increase in location described above in Results (Figures 6 and 7) provides strong evidence that implementing the Thinking Classroom model in grade 12 AP Calculus resulted in a significant increase in students' problem-solving ability. The location of every student in the study, as measured by the Rasch model, increased by an average of 2.830 logits. Per the chosen observational framework (see Andrich, 1988), we can conclude that students were better problem-solvers after the instructional intervention than they were before.

The increase in the spread of the data from before the intervention to after indicates that, while all student locations increased, there was variation in the amount of change. In other words, all students improved at non-routine problem solving, but some improved much more than others.

Chapter 5: Conclusions

Overview

In seeking to measure an instructional intervention's effect on student problem-solving ability, the study adopted David Andrich's framing of a constructed variable and an observational framework. The resulting transparency may provide a window into the researcher's personal biases, and an opportunity for other researchers to consider a defineition of problem-solving for their own students, aligned with their own preferences. While the size and composition of the cohort of students who participated in the study, the timing of the research, and the teacher's experience level may limit how broadly we interpret the results, the methodology was useful at the classroom level. The design and implementation of the study provided insight for the researcher into how and why we measure our students' abilities, and yielded an opportunity for reflection on how measurement interacts with beliefs about student growth.

Strengths and Weaknesses of Methodology

Paul Lockhart's assertion that "a good problem is something you don't know *how* to solve" highlights the central challenge of this research. Measurement of the constructed variable (non-routine problem-solving ability) requires an observational framework – in this case, the original set of 16 problems. The Rasch analysis, and test-equating, allow comparable measurements of student performance on unfamiliar problems, on both the pre- and post-assessments.

One strength of this research methodology is that the framework must be explicitly stated at the outset. In other words, in seeking to quantitatively measure the constructed variable, the researcher must reveal what he or she means by non-routine problem-solving ability. As a result his or her own personal preferences and biases (which inevitably exist) are made transparent.

To illustrate this point, note that I did not include in my observational framework any problems that were combinatoric in nature, although these appear frequently in puzzle books and websites. (For example: "How many different 10-digit numbers, such as 9,734,289,294, can be written by using all 10 digits? Numbers starting with zero are excluded" (Vita, 2019).) This exclusion may reflect a bias against these types of problems, or a distaste for assigning them to my students, or a belief that counting problems are less important than other types of problems, which I did include. A researcher who disagrees with this omission, has a different understanding of problem-solving, or who values different component skills, could provide a different observational framework.

In conjunction with this strength, one weakness of the methodology is its potential lack of transferability. Selecting the observational framework required an understanding of my students' prior math learning and reading ability, along with some amount of instinct and guessing. The goal was to curate a selection of problems that were accessible and appropriately challenging for the participating students. To perform this research at the same scale with a different group of participating students would likely require a different observational framework, informed by similar knowledge of the student subject group. And, the potential exists for the generation of an unusable data set, if the researcher fails to select an appropriate set of problems – for example, if several are too easy are too hard for the students. The best method for avoiding this potential problem would be to administer a much larger set of problems, with the premise that several would be eliminated as Extreme, but this would result in more testing time, and the associated discomfort, for students.

Another weakness of the study is the relatively small number of participating students, and the homogeneity of the group. All 18 participants were grade 12 students enrolled in AP Calculus, indicating prior success in several high school math classes. This was an adequate size population to perform the Rasch analysis of person-item interactions, which revealed a several-logit range of ability even within this cohort. Still, one would have more confidence that the results were more broadly transferable, if the study were conducted on a larger, more diverse population of students.

Influential Factors

The very strong result detected by this study may have been influenced by two positive factors, and/or one negative factor.

The pre-assessment was administered very early in the 2021 school year, when students were returning to class after significant disruptions due to the Covid pandemic. The post-assessment was administered three weeks later. It is likely that students' quantitative reasoning and problem-solving skills were somewhat dormant following the summer, and much more active following three weeks of full-time school. Since summer learning loss has been widely documented in a variety of subject areas and grades, particularly for students with disadvantaged backgrounds (Kuhfeld, 2019), it would be unsurprising if students' non-routine problem-solving skills also diminished over the summer. If so, the pre-assessment scores may have been lower than if the study had been conducted later in the school year.

Second, the random division of the 16 problems into two sets yielded a post-assessment that was much more difficult than the pre-assessment. In fact, all five of the highest-location (i.e., hardest) problems ended up in the post-assessment. Because the item locations were not determined until the end of the study, when the Rasch analysis was complete, I could not foresee this inequitable distribution of difficult problems. The difference in difficulty of the two assessments resulted in a test-equating situation where students could be assigned a much higher location from the post-assessment than the pre-assessment, for a similar score, as can be seen in Figures 7 and 8. While these conversion scales indicate the potential for over-estimating the problem-solving skill of students at the time of the post-assessment, it should be noted that the purpose of test-equating within the Rasch model is to measure and account for problem difficulty in the calculation of person-item interactions. Setting aside the sophisticated Rasch analysis for a moment, the data show that Group B students performed much better on the post-assessment after the instructional intervention than Group A did beforehand (average raw score of 19.44 vs 10.56, out of 28) while on the pre-assessments the groups did about the same (average raw score of 18.33 vs 18, out of 32). Still, the unequal distribution of item locations may have caused the Rasch model to exaggerate this result, in seeking to quantify it.

A third influential factor may have affected the study's result in the opposite direction. Enacting the Thinking Classroom model requires a significant amount of skill, creativity, flexibility, and experience. I must admit here that I am an enthusiastic adopter, but not an expert practitioner, of this teaching style. I have been striving to create a Thinking Classroom for three school years; while I continue to improve, I still have much to learn. The results of this study are undoubtedly positive, but I wonder if a teacher with more expertise could have achieved an even stronger effect. In particular, I am curious if a more faithful implementation of the Thinking Classroom model could result in gains being more evenly-distributed among the students – that is, without the increase in the spread of the data.

Recommendations for Further Investigation

Measuring students' improvement at non-routine problem-solving is more difficult than measuring their acquisition of content knowledge. This may be one reason that content knowledge forms the basis of most, if not all, of the assessment data collected in math classrooms. And yet, as math educators who deeply value problem-solving, we should be intentional in our efforts at both cultivating and measuring this skill. The small study described here shows that quantifying the growth of our students' problem-solving ability can be accomplished at the classroom scale. I would encourage other math educators to strive to measure what they truly value, not just what is convenient.

It would also be enlightening to see larger-scale studies connecting instructional style with student problem-solving skill. How do lecture-style classes, drill-based curricula, or online math classes impact students' growth as problem solvers? Can we find evidence that constructivist-inspired teaching interventions cause students to improve at non-routine problems? By quantifying students' problem-solving ability at multiple points in time, we may be better able to determine what strategies are most effective at fostering growth in this critical area.

Limitations of Implementation

This study was limited to my current teaching assignment, and constrained by the time frame prescribed by the capstone assignment. Given the strength of the result, I am confident in concluding that students in my grade 12 AP Calculus class were better problem-solvers after 10 hours of Thinking Classroom instruction than they were at the beginning of the school year, within the observational framework created by my selection of the 16 non-routine problems. It is unknown whether or to what degree this result could be generalized to other populations of students. Because I believe that non-routine problem-solving can be cultivated as a transferable skill, I suspect that the measurements taken in this study may correlate with other observational frameworks of problem-solving – for example, AP Calculus exam questions. However, a test of that hypothesis is beyond the scope of this study.

Implications of Research on Educational Practice

Designing and implementing this action research study was beneficial to me as a classroom teacher in several ways. The process required me to reflect on what aspects of math education I value the most, to investigate how these may be measured, and to consider the constraints on such measurements. Utilizing the Rasch model also provided me with some new insight into the problems I selected. Finally, my adoption of a mathematical technique from the field of psychometrics, and a consideration of the history of that field, provided an opportunity to think about the beliefs and attitudes educators hold regarding student ability.

In planning this study, I was motivated to address a problem of practice which I believe is common in many math education settings: how do we help our students become more independent and capable problem-solvers? As I considered possible experimental designs, I gained an understanding of the role and value of non-routine problems in the math classroom. This insight aligned well with Liljedahl's Thinking Classroom model, where rich problems are central. My desire to measure students' problem-solving ability required me to research and curate a set of non-routine problems that I considered appropriately challenging for a specific cohort of students. Through research into the work of Rasch and others, I came to understand this process as one of mathematical modeling: as explained by David Andrich in *Rasch Models for Measurement*, I constructed a unidimensional variable and chose an observational framework.

I now see that as educators we engage in this kind of modeling regularly, although we might not always be aware of it. In many cases, our goal – what we truly want students to be able to do, beyond the acquisition of procedural skills – only becomes manageable when we model it as a variable that can be measured with a discrete task or set of problems to which a some kind of scale (such as a rubric) can be applied. It is important, then, to be honest and transparent about the observational framework we have chosen, and to acknowledge the modeling process. Since, as statistician George Box famously said, "All models are wrong, but some are useful," we should strive for the best frameworks we can, and also acknowledge their inevitable shortcomings.

The quantitative sophistication of the Rasch model, as harnessed by Dr. Andrich's powerful RUMM 2030 software, provided me with a much deeper insight into the interaction of my students and the chosen problems than would have otherwise been available. By assigning a location, and an Item Characteristic Curve, to each problem, the analysis showed in detail how easy or difficult each problem was for the students. I regret that I did not write down predictions of relative problem difficulty before administering the assessments, as this would have been an interesting test of my instincts as a math teacher. I am sure that I would not have foreseen that problem A was so easy (determined Extreme by the analysis), nor that problem 14 would prove to be so difficult for the students (location 6.405). In the end, I was gratified that, with the exception of the single too-easy item, the chosen problems did provide an appropriate level of challenge for the class.

I decided to use Rasch analysis in this study because I believed it was the appropriate tool for this application, but also with an awareness that the history of psychometrics contains both motives and applications which conflict with my own values. The Victorian scientist Francis

Galton invented both psychometrics and eugenics (and coined the two terms); his desire to measure people was inextricably linked to now-discredited beliefs in fixed mental abilities and in a hierarchy of races. Many early psychometricians sought to quantify human intelligence, believing it to be a stable, inherent trait, and failing to understand the bias inherent in their measurement tools. The erroneous belief in biological determinism, the problems with IQ testing, and the history of racism in the social sciences (and psychometrics specifically) are well-documented by Stephen Jay Gould in *The Mismeasure of Man* (1996).

My construction of a unidimensional variable representing problem-solving ability does not reflect a belief that it is a fixed characteristic of my students; in fact, the opposite is true. Rather than approach problem-solving skill as a latent trait, I began this study with two central premises: that non-routine problem-solving ability can be measured, and that it can be increased. The Rasch analysis allowed me to measure students at two points in time, with different non-routine problems, in order to show improvement. As a result, I hope that I have applied a mathematical method from the field of psychometrics in a way that demonstrates my belief that measurement can support the growth of all students.

References

Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43, 561–73.

Andrich, D. (1988). *Rasch models for measurement*. Sage.

Bandura, A. (1997). *Self-efficacy: The exercise of control.* W H Freeman/Times Books/ Henry Holt & Co.

Degol, J. L., Wang, M.-T., Zhang, Y., & Allerton, J. (2018). Do growth mindsets in math benefit females? Identifying pathways between gender, mindset, and motivation. *Journal of Youth & Adolescence*, 47(5), 976–990. https://doi.org/10.1007/s10964-017-0739-8

Dweck, C. S. (2007). *Mindset: The New Psychology of Success* (Illustrated ed.). Ballantine Books.

Faulkenberry, E. and Faulkenberry, T. (2006). Constructivism in mathematics education: A historical and personal perspective. *The Texas Science Teacher*, Vol. 35 p. 17-21.

Gould, S. J. (1996). *The Mismeasure of Man*. New York: Norton.

Kuhfeld, M. (2019). Surprising new evidence on summer learning loss. Phi Delta Kappan, 101(1), 25–29.

Liljedahl, P., Malaspina, M., and Bruder, R. (2016). Problem solving in mathematics education. Springer Open. https://library.oapen.org/bitstream/handle/20.500.12657/27743/1002262.pdf?sequence=1

Liljedahl, P. (2016). Building thinking classrooms: Conditions for problem solving. In P. Felmer, J. Kilpatrick, & E. Pekhonen (eds.) *Posing and Solving Mathematical Problems: Advances and New Perspectives*. New York, NY: Springer.

Linacre, J. (1992). Why fuss about statistical sufficiency? *Rasch Measurement Transactions*. 6:3, p. 230.

Linacre J.M. & Wright B.D. (1989). The "length" of a logit. *Rasch Measurement Transactions.* 3:2, p. 54-55. https://www.rasch.org/rmt/rmt32b.htm.

Lockhart, P. (2009). A mathematician's lament: How school cheats us out of our most fascinating and imaginative art form. New York, NY: Bellevue Literary Review.

Malone, J.A., Douglas, G.A., Kissane, B.V., and Mortlock, R.S. (2007). Measuring problem solving ability. *Stepping stones for the 21st century*, p. 187-200. Australasian Mathematics Education Research. Brill.

Masters, G.N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47, 149–174.

Mayer, R.E. (1992). *Thinking, problem solving, cognition: second ed*. New York: Freeman.

Mayer, R. E. (1998). Cognitive, metacognitive, and motivational aspects of problem solving. *Instructional science* 26: 49–63. Kluwer Academic Publications.

National Governors Association Center for Best Practices, Council of Chief State School Officers. "Common core standards for mathematical practice." (2010). Washington D.C. http://www.corestandards.org/Math/Content/HSF/IF/

"Ohio Digital Mathematics Project." (2021). Ohio State University Department of Education and Human Ecology. Web. 26 July 2021. https://u.osu.edu/odmp/

Polya, G. (2009). *How to solve it.* Princeton University Press.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: Danmarks Paedogogiske Institut.

RUMM Laboratory. (2021). http://www.rummlab.com.au/

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, Florida: Academic Press Inc.

Shoenfeld, A. (2003). "Math wars." August 5, 2003. http://pi.math.cornell.edu/~henderson/courses/EdMath-F04/MathWars.pdf

Steffe, L. and Kieren, T. (1994). Radical constructivism and mathematics education. *Journal for Research in Mathematics Education*, volume 25, issue 6, p. 711-733.

Thompson, P. W. (2013).: Constructivism in mathematics education. In Lerman, S. (Ed.) *Encyclopedia of Mathematics Education*. SpringerReference.

Vita, J. (2019). “5 combinatorics puzzles you need to try before 2020.” *Medium*. https://medium.com/@jakevita/5-combinatorics-puzzles-you-need-to-try-before-2020-7a947d0ce619

Appendix A: Instructional Plans

Two weeks of instruction (up to 10 hours) should be conducted using the Thinking Classroom model (Liljedahl, 2016). Each class period should carefully adhere to the nine elements described in the article on pages 381 and 382. Simultaneously, assign asynchronous instructional content for students to watch as homework. A typical class period during the intervention will be structured as follows:

Using a visibly random method, divide the class into groups of three students. Each group can select a vertical whiteboard as a workspace; these are located around the perimeter of the classroom. Each group receives one dry-erase marker. The norm is that only one student writes at a time, and the marker is passed around the group. Problem-solving tasks are given verbally to all groups at once. The teacher circulates and interacts with groups, answering keep-thinking questions only. When a group's work yields particular insight, the teacher may draw the whole class' attention to the work and discuss it. Provide hints and extensions to groups with the goal of creating and maintaining flow (Csíkszentmihályi, 1990). Bring tasks to completion as described by Liljedahl: "When every group has passed a minimum threshold, the teacher needs to engage in discussion about the experience and understanding the whole class now shares" (2016, p. 382). These discussions should touch on both details of math content, and problem-solving techniques and heuristics (see Bruder, 2016).

Problem-solving tasks: The following problem-solving tasks are intended to be assigned to groups in verbal-instruction form. Depending on the pace of group work, one or more of these may form the lesson for each class period during the instructional intervention:

i. What is a better fit, a square peg in a round hole or a round peg in a square hole?

ii. How many squares (of any size) are on a chessboard?

iii. Make the numbers from 1 to 30 using four 4's and any operations.

iv. Paint all the sides of a 3 x 3 x 3 cube. Once it is dry take it apart into 1 x 1 x 1 unit cubes. How many of these unit cubes have paint on three faces? Two faces? One face? No faces? What about for an *n* x *n* x *n* cube?

v. A spherical basketball is packed in a cubical box into which it fits exactly. What percent of the volume of the box is used up by the basketball?

vi. Find the slope of the secant line in the parabola y=x2 between (1,1) and (3,9). Keep the point (1,1) the same, and move the other point closer and closer to it. What can you determine about the slope of these new secant lines?

vii. I drove from here to another town. Part 1. The first hour I drove 20 mph. The second hour I drove 30 mph, then I arrived. What was my average speed on the trip? Part 2. The first half of the way I drove 20 mph. The second half of the way I drove at 30 mph. What was my average speed on the trip?

viii. When the contents of a cylinder are poured into a second cylinder whose radius is 2 inches greater, the height reached in the second cylinder is one half of that reached in the first. Find the radius of the smaller cylinder.

ix. The short leg of a 30-60-90 triangle is on the *x*-axis. What are the slopes of the three sides of the triangle? What if the other leg, or the hypotenuse, is on the *x*-axis?

x. The short leg of a 3-4-5 triangle is on the *x*-axis. What are the slopes of the three sides of the triangle? What if the other leg, or the hypotenuse, is on the *x*-axis?

xi. What angle between 0 and 2 has the maximum value for the sum of its sine and cosine? Its sine, cosine, and tangent?

xii. A goat is tied to the corner of a barn. The barn is 20 x 40 feet, and the rope is 50 feet long. No trees or other obstructions are in the way. What is the available area of grass that the goat can eat?

xiii. Seven tangent circles of the same size are inscribed in a larger circle. The total area enclosed by the small circles is what fraction of the area enclosed by the larger circle?

xiv. A function has the following properties: limx3f(x)=1, limx2-f(x)=4, limx2+f(x)=-4. From this information, what do you know about f(3)and f(2)? Explain how you know.

<u>Sources of the problem-solving tasks</u>: Some of the problems above are original; the others are from one of the following sources.

https://u.osu.edu/odmp/
https://www.peterliljedahl.com/teachers/good-problem
http://www.cut-the-knot.org/content.shtml

Appendix B1: Pre- and Post-assessment Problems

Instructions: Please attempt each problem.

Show all your work, including anything you try that doesn't work out.

---

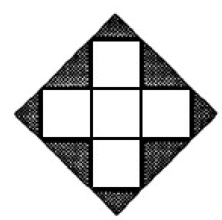
1. A father promised to pay his child 8¢ for every math problem solved correctly and to fine the child 5¢ for each incorrect solution. After 26 problems neither owed anything to the other. How many problems did the child solve correctly?

---

2. A fireman stood on the middle rung of a ladder, directing water into a burning building. As the smoke lessened, he stepped up three rungs. A sudden flare-up forced him to go down five rungs. Later he climbed up seven rungs and worked there until the fire was out. Then he climbed the remaining six rungs to the top of the ladder and entered the building. How many rungs did the whole ladder have?

---

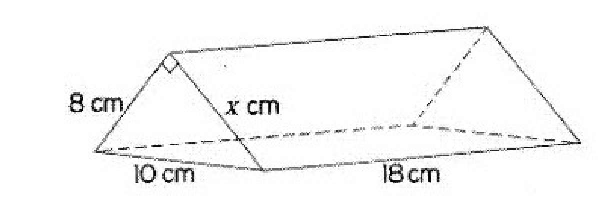
3. The figure at the right consists of squares and isosceles triangles. What percent of the entire figure is shaded?



---

4. A fishing boat sails 40 miles east, then 80 miles south, and finally 20 miles east again. How far is the boat from its starting point?

5. Find the total surface area of the right prism shown below.



6. Two miles of fence will enclose a square of 156.25 acres. How large a square pasture will 4 miles of fence enclose?

7. If a cylindrical jar of peanut butter that is 3 inches in diameter and 4 inches tall sells for $2.00, what is a fair price for a jar that is 6 inches in diameter and 6 inches tall?

8. Of two identical barrels, one is half full and one is two-thirds full. One quarter of the liquid in the second barrel is poured into the first. The first barrel now contains 25 more gallons of liquid than the second. Find the capacity in gallons of one of the barrels.

Name ________________________

Instructions: Please attempt each problem.

Show all your work, including anything you try that doesn't work out.

---

9. At noon, I started driving from city A to city B at a steady 40 mph. Fifteen minutes later, at 12:15, you started driving from city B to city A at a steady 50 mph. We passed each other at a point midway between the two cities. How far apart are city A and city B?

---

10. A 10 cm tall cylinder fits perfectly inside a box that is 4 cm x 4 cm x 10 cm. What is the surface area of the cylinder?

---

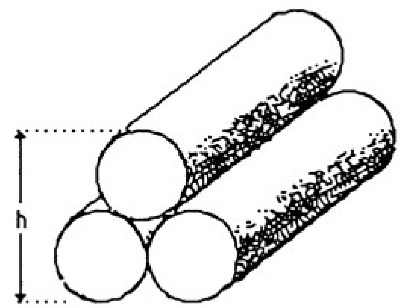
11. Each side of a triangle ABC is 12 units in length. Point D is the foot of the perpendicular drawn from A to side BC Point E is the mid-point of segment AD Find the length of segment BE.

---

12. The supermarket sends a bill for 24 dozen eggs, but leaves off the first and last digit of the cost: $_2.4_. If eggs cost less than one dollar a dozen, how much should the bill be for?

13. Zach and his family were driving into the city for the ball game. Zach fell asleep when they were halfway there. When he woke up the distance they still had to go was half as far as they went while he was asleep. For what fraction of the way did Zach catch his z's?

14. Three pipes are stacked as shown. The outside diameter of each pipe is 12 inches. How tall is the stack (*h* in the picture)?



15. The lengths of the sides of a triangle measured in inches are three consecutive integers. The length of the shortest side is 30% of the perimeter. Find the lengths of the three sides.

16. A circle has a square inscribed inside it, and another square circumscribed about it. What is the ratio of the perimeter of the outer square to that of the inner square?

Appendix B2: Rubric for Evaluating Non-Routine Problem-Solving

Score	Solution Stage
0	*No Start* The student is unable to begin the problem or hands in work that is meaningless.
1	*Approach* The student approaches the problem with meaningful work, indicating some understanding of the problem, but reaches an early impasse or goes astray.
2	*Substance* Sufficient detail demonstrates that the student has proceeded toward a rational solution, but major errors or misinterpretations obstruct the correct solution process.
3	*Result* The student very nearly solves the problem; minor errors produce an invalid final solution.
4	*Completion* The student provides an appropriate method to yield a valid solution.