Mathematical Reasoning and Assessment Scores

Christen VanNewkirk
TE 855: Fall 2010
Kristen Bieda
Michigan State University

Introduction

In an education system that is driven by standardized assessment scores, Annual Yearly Progress, and No Child Left Behind, teachers need to fully prepare students on the demanding curriculum at hand. In the Nation's eyes, the success or failure of a school is determined from the results of one test at the end of the year. Now more than ever, teachers are being held accountable, on how well his or her students are prepared for these assessments that are based solely on right or wrong answers.

But this concept raises a deep sense of concern. Since these assessments are based on right or wrong answers, students are not given the opportunity to prove their reasoning skills. With an assessment-driven educational system, educators want to ensure that their emphasis on mathematical reasoning will have a positive affect on a child's ability to pass standardized assessments. Therefore, since standardized testing is focused on right and wrong answers instead of conjecturing and inventing, do tasks with mathematical reasoning positively affect assessment scores?

Literature Review

Three articles assisted in the design of the tasks for this research project. In order to test the affects that mathematical reasoning has on assessment scores, it is necessary to develop tasks that are rich in mathematical reasoning. Ball and Bass (2003) express that "the notion of *mathematical understanding* is meaningless without a serious emphasis on reasoning" (p. 28). In order to attempt to make math reasonable to my students, I must be able to design tasks that are rich in mathematical reasoning. With these three articles, it was found that the task needs to have multiple representations, have multiple ways of solving it, and/or cross-curricular. However, these alone do not make a task successful. Rather, the teacher needs to monitor students constantly, encouraging them to develop higher-levels of thinking.

Hutchinson (2010) suggests a few different components that will help make mathematical reasoning tasks successful. The first component is that tasks need to allow students to solve the problem in multiple ways. Another example is that tasks can have multiple correct answers. Finally, it is acceptable for tasks to be outside of the discipline of study; "One or two well-designed problems will provide a wealth of material to discuss" (p. 46). When designing a task that allows students to solve it in different ways, it enables students to choose the way they would like to solve it. It provides the opportunity for students to be able to explain their process to students who solved it differently. Also, when tasks have only one correct answer, the educator is limiting the class to learning procedures and processes. Again, these mathematical reasoning tasks, force students to take responsibility for their own learning and work within their ability level. The final way that creates successful discussions is to create tasks that are cross-curricular. "Multidisciplinary problems have increased relevance for students, by the nature of their real-world applications. In addition, students who are not as interested in math can be engaged by the connections to other disciplines" (p. 47).

Henningsen and Stein (1997) explore "the classroom factors that hinder or support students' engagement in high-level mathematical thinking and reasoning for doing mathematics" (p. 527). They found that it is not merely the design of the task that creates a classroom full of students who are proficient in mathematical reasoning. Rather there are three aspects that had the greatest impact on the data; task design teacher communication/monitoring, and time management.

The task, as Hutchinson noted, needs to be open-ended and students need a choice as to how to solve the problem. However, this article added that setting students free on an open-ended question alone will not obtain the desired results. Rather teachers must constantly monitor the conversations between students, and push them to the highest comprehension levels possible. Finally, the time management must be flexible. Henningsen and Stein note that "In agreement with research on students' engagement with academic tasks, these findings suggest that planning for appropriate amounts of time and flexibility with timing decisions may play an important role in avoiding decreases in the level of cognitive activity engaged in by students as tasks unfold in the classroom." (Henningsen, 1997). Too much time can be just as detrimental as too little time, so teachers must balance options at all times.

The third article that assisted in forming a task that would prove to be rich in mathematical reasoning was Mary Kay Stein and Margaret Smith's way of classifying tasks. This article proved to be very useful in comparing the types of tasks used in each class that would be researched. There are two different levels of demands for a task: lower-level and higher-level. Lower-level demands can be either "memorization" or "procedures without connections." Memorization tasks "have no connection to the concepts or meaning that underlie the facts, rules, formulas, or definitions being learned or reproduced." (Stein & Smith) Rather they require students to simply reproduce facts, formulas or definitions, without showing any levels of comprehension. On the other hand, "procedures without connections" require the "use of the procedure either is specifically called for or is evident from prior instruction, experience, or placement of the task." (Stein & Smith) Notice that tasks with low-level demands would have only one right answer. Students would not have to make any connections to prior knowledge. On the other hand, tasks that have higher-level demands require students to have a deeper level of understanding. There are two different types of tasks that have higher level of demands: “procedures with connection” and “doing mathematics.” “Procedures with connections” are “usually represented in multiple ways, such as visual diagrams, manipulates, symbols, and problem situations. Making connections among multiple representations helps develop meaning.” (Stein & Smith) This level would prove to be the most common level in my “experimental class.” Finally, “doing mathematics” takes students reasoning ability to another level. This task would be developed so that students would not use “a predictable, well-rehearsed approach or pathway.” This would require the teacher to give students an open-ended task, and allow students to decide how they would solve it, using any prior knowledge.

In conclusion, these articles helped formulate the tasks that would be useful in creating lessons to increase student’s mathematical reasoning abilities. Mary Jo Hutchinson and Henningsen & Stein agreed that a task that is open-ended, and has multiple ways of solving, provides an opportunity for students to have conversations with each other on possible solutions. Henningsen & Stein added that constant monitoring and intervention is necessary for students to push their reasoning skills at all times. Students will not always push themselves to their academic ability, so with teacher feedback, the students can strive to meet the necessary expectations. Finally, Stein & Smith enabled me to have a clear indication of what task would require high-levels of thinking, and what tasks would expect little student thinking. Their article will promote consistency and direction when comparing the tasks in this project.

However, all three of these articles were lacking a clear definition of mathematical reasoning, and what it sounds/looks like in the math classroom. Obviously desired results are to have students develop their reasoning skills in order to increase assessment scores. However, there must be a strong understanding of what mathematical reasoning is and is not. In this research project, mathematical reasoning will be defined as an argument made to justify one's process, procedure, or conjecture, to create strong conceptual foundations, in order for students to be able to process new information in math and other content areas. Since the definition uses mathematical reasoning to justify an answer, students should explore why their process makes sense. For example, a student noting that inverse operations are used to solve linear equations is not enough. Students need to understand that the reason why you add, subtract, multiply, or divide the same quantity on both sides is to ensure that the equation remains balanced at all times. Mathematical reasoning also includes students discussing open-ended questions among each other. When students are put into a position when they need to "defend" their opinion, it is necessary to have a strong understanding on why math works and using the appropriate vocabulary to express those reasons.

Methodology

Participants

This research is focused on comparing the assessment scores of two different Algebra I classes - a control group, and an experimental group. There are 15 females and 8 males in the control group. There are 14 females, and 9 males in the experimental group.

Tasks

During one chapter, one particular strategy was chosen for the tasks for each class. For the control group, the tasks had lower-level demands and focused on teaching strictly the required content. These tasks required students to apply different techniques, procedures and steps, to solve the problems. All lessons were focused on the "how-to's" of math. For example, how do you find the slope of a table, how do you find the slope of a line, is the slope of the line positive, negative or zero, etc… For the experimental group, the tasks were designed to help build the students mathematical reasoning skills. The tasks gave students the opportunity to make connections between concepts and be exposed to multiple representations of linear equations. For example, in a three-day lesson, students were ability grouped with two to three other students. The task was to determine which of the three plants would grow at a faster rate. However, the growth of each plant represented differently. The question can be found below.

Which plant is growing at a faster rate? How do you know?

Plant A: The growth of plant A, is shown by the following table:

Day	*0*	*1*	*2*	*3*	*4*
Height (cm)	*5.08*	*7.62*	*10.16*	*12.7*	*15.24*

Plant B: The height of Plant B is represented by the equation:

h = ¾ d + 5 (where h is the height in inches and d is the number of days)

Plant C: A plant with a height of ½ foot, grows ½ inch each day.

With this plan in mind, I hypothesized that the tasks for the experimental class will require them to use more mathematical reasoning and therefore, raise assessment scores, even though assessments tend to be driven by right or wrong answers. I also hypothesized that the conversations in the experimental class would be rich in mathematical reasoning and they would begin to look to each other for information.

Data Collection

To support any conclusions that are made, both quantitative and qualitative data was collected. Quantitative data would show evidence of possible changes in the assessments scores. A pre-assessment was a chapter quiz. For the mid-assessment, students were asked to solve simple, low-level questions based solely on finding slope of different linear representations.

Students were also required to solve the following, low-level questions:

Find the slope/rate of change of the following:

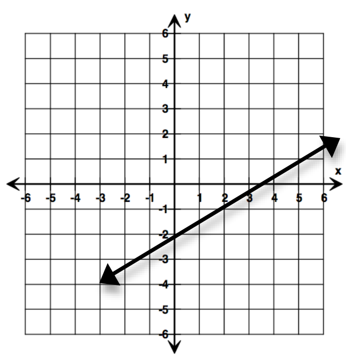
1.

x	*-5*	*0*	*5*	*10*
y	*12*	*7*	*2*	*-3*

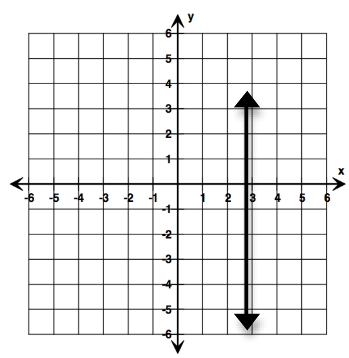
2.

x	*-1*	*0*	*2*	*5*
y	*2*	*0*	*-4*	*-10*

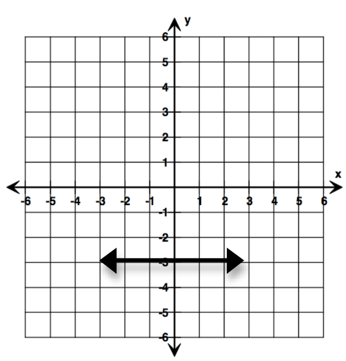
3.



4.



5.



6. $y = 10x - 3$

7. $f(x) = 7x$

8. Gas costs 1.89 per gallon. Find the rate of change.

9. For 16 hours of work, a paycheck is \$124. Find the rate of change.

Finally, the post-assessment was a chapter test.

The second form of quantitative data collected was ClassScape pre- and post-assessments. ClassScape is a computerized system that prepares students for questions that will be similar to standardized tests. The growth or decline made among student's assessment scores between both classes will be analyzed, and differences in class averages for both the pre- and post-assessment will be reviewed. The pre-assessment was given prior to any strategies, and the post-assessment was given two-weeks into the research project. The following tables summarize the data collected.

Quantitative Data	
Pre-, Mid- & Post-Assessment	ClassScape Pre- & Post-Assessment
Students were given a pre-assessment, a mid-assessment and a post-assessment. I reviewed the class averages of all assessments.	Students were given a pre-assessment and a post-assessment. I reviewed the class mean for each assessment.

Qualitative data is equally important as well, since it will show any students using mathematical reasoning to explain their answer. By videotaping my class, the types of conversations that occur in both classes will be compared and contrasted. These transcripts and conversations will show whether or not students in either classes are using their reasoning skills to defend their response, or connect with the math, at a deeper level. When analyzing the data, I reviewed the types of conversations they were having and whether or not students were using higher-levels of thinking. I also analyzed how the task affects how reacted to the students questions.

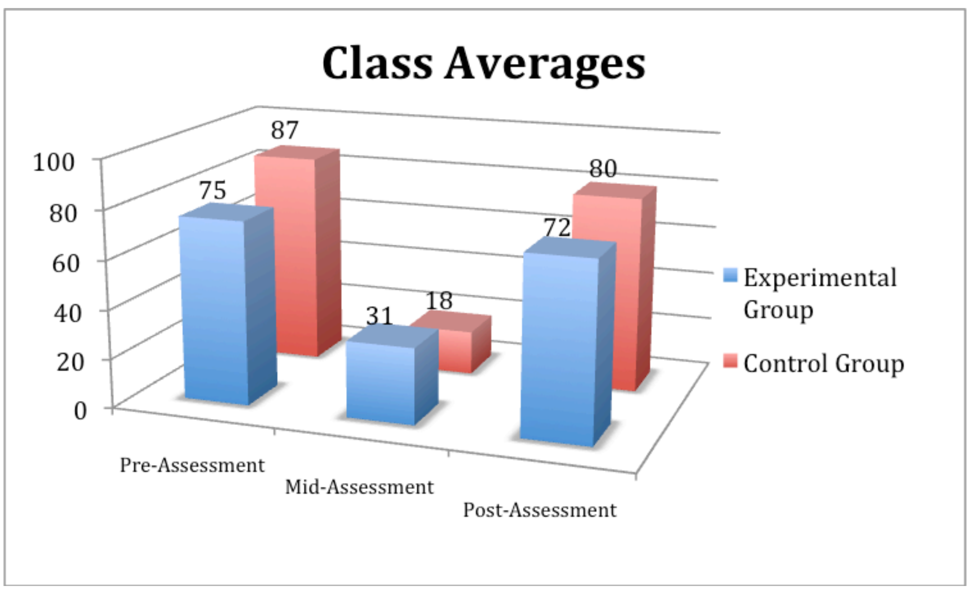
Results

Quantitative Data

After analyzing the qualitative data, the results were inconclusive in answering the research question of how does mathematical reasoning tasks affect assesement scores. The experimental and control group showed similar trends in pre-, mid- and post-assessment scores, along with the pre- and post-assessments with ClassScape. The only difference that appeared to occur was when students were asked to find the slope given a real-world situation. The experimental group appeared to perform better than the control group in this area.

Pre-, Mid- and Post-Assessments

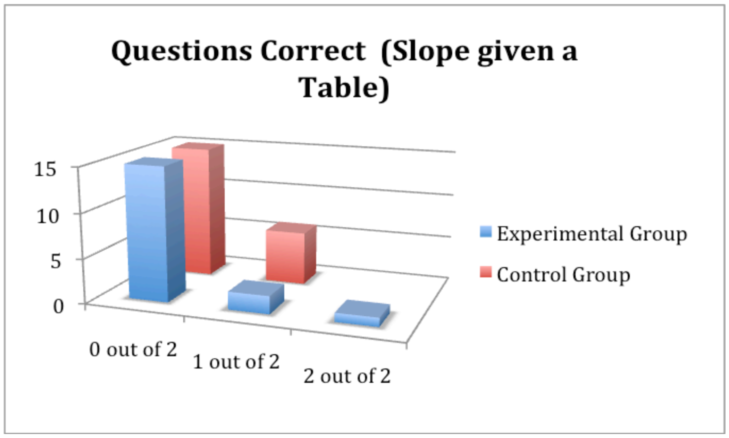
The class averages of the pre-, mid- and post-assessment class averages were 87%, 18% and 80% respectively, for the control group. The class averages for the pre-, mid- and post-assessment class averages were 75%, 31% and 72% respectively, for the experimental group. The results are shown in the graph below.



Since there were dramatic and unexpected results from the mid-assessment, further analysis of the data was necessary, to show where students had difficulty in each class. To do this, the mid-assessment was broken down, according to task. The mid-assessment was comprised of four "tasks." Students had to find the slope when given a (1) table, (2) graph, (3) equation and (4) a real-world situation. To determine what tasks posed the greatest challenge for students, the data was graphed for each task separately. The results are as shown.

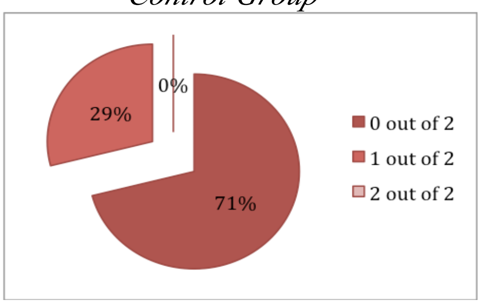
Mid-Assessment: Task 1

The table shows the quantity of students, who received various scores, in both the experimental group and the control group for the first task (how to find the slope when given a table).

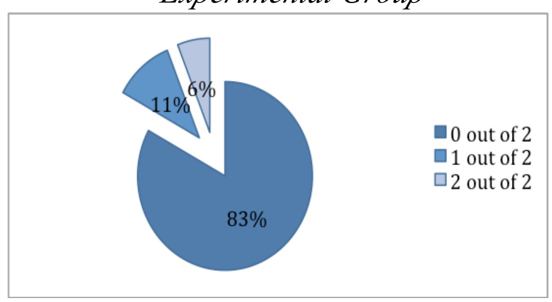


Percent of Students, with specific scores

Control Group

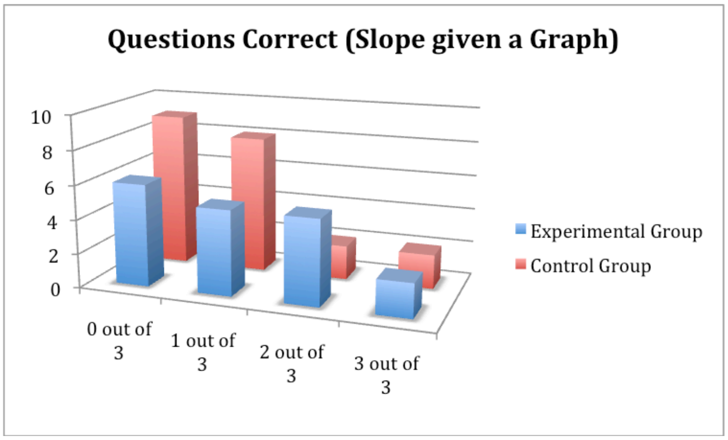


Experimental Group



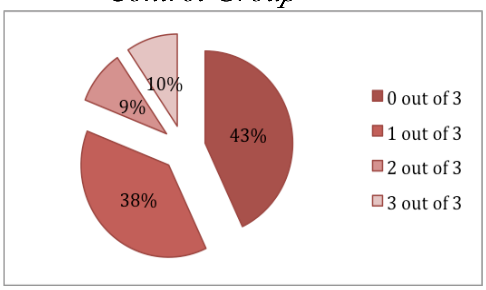
Both the control group, and the experimental group seemed to have similar results. In general, students were not successful in finding the slope, when given a table.

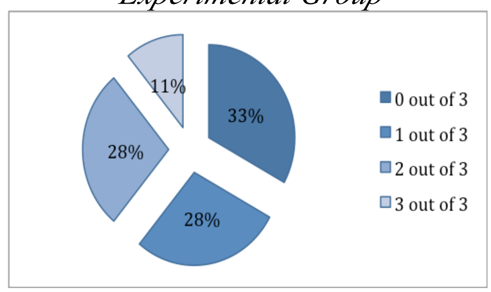
Mid-Assessment: Task 2



Percent of Students, with specific scores

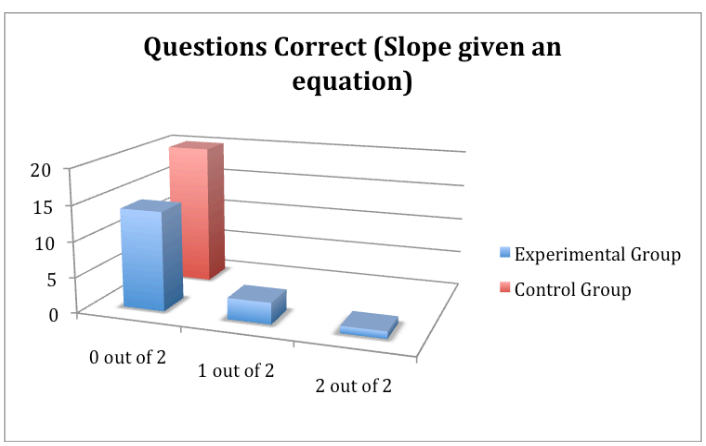
Control Group *Experimental Group*





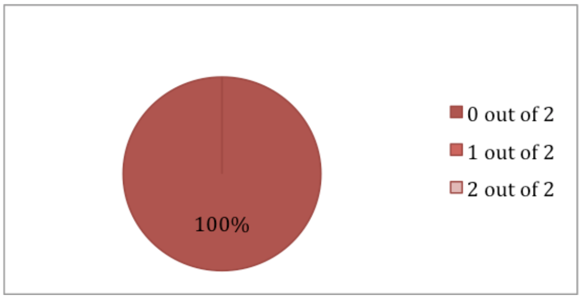
The experimental group had a higher percentage of students receive at least one problem correct, but in general, the results were fairly consistent between groups for task 2.

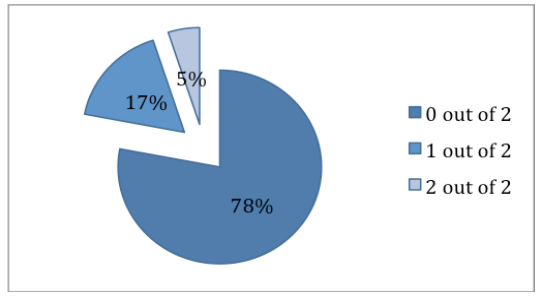
Mid-Assessment: Task 3



Percent of Students, with specifc scores

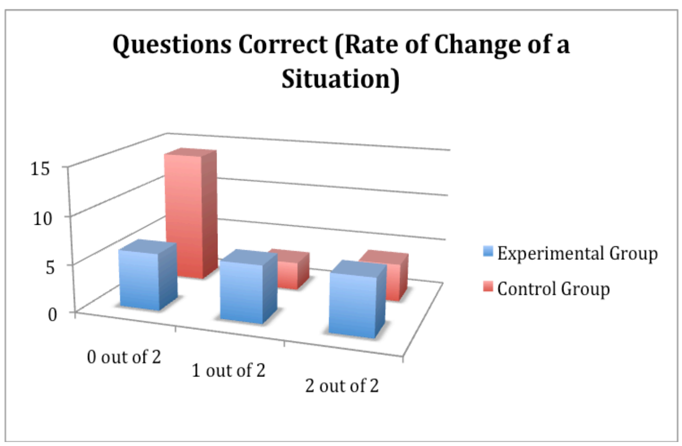
Control Group *Experimental Group*





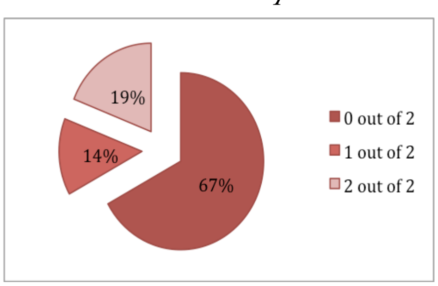
In task 3, one can conclude that the experimental group is beginning to have different results from the control group. Not one person could find the slope of an equation, but in the experimental group, 23% could find the slope of at least one equation. It is possible that students were able to determine the connections between graphs, tables, and equations, from the higher-level of thinking tasks given.

Mid-Assessment: Task 4

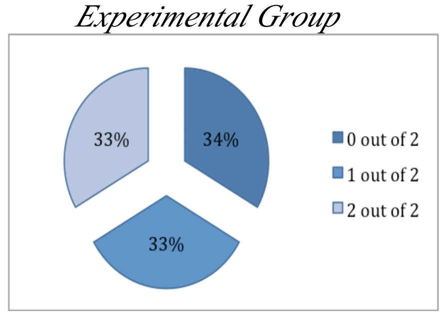


Percent of Students, with specific scores

Control Group



Experimental Group



Finally in task 4, we see equal representation from the experimental group. However, in comparison to the control group, only 33% of students could find the rate of change when given a situation.

In the bar graphs, and pie charts above, one can conclude that the experimental group had a higher percentage of students would were able to find the slope, given specific situations, and when given an equation. I hypothesize that this is because the three-day lesson on the growth of plants was driven from real-world situations. Therefore, when students practice reasoning through real-world situations, they are more capable of solving real-world problems.

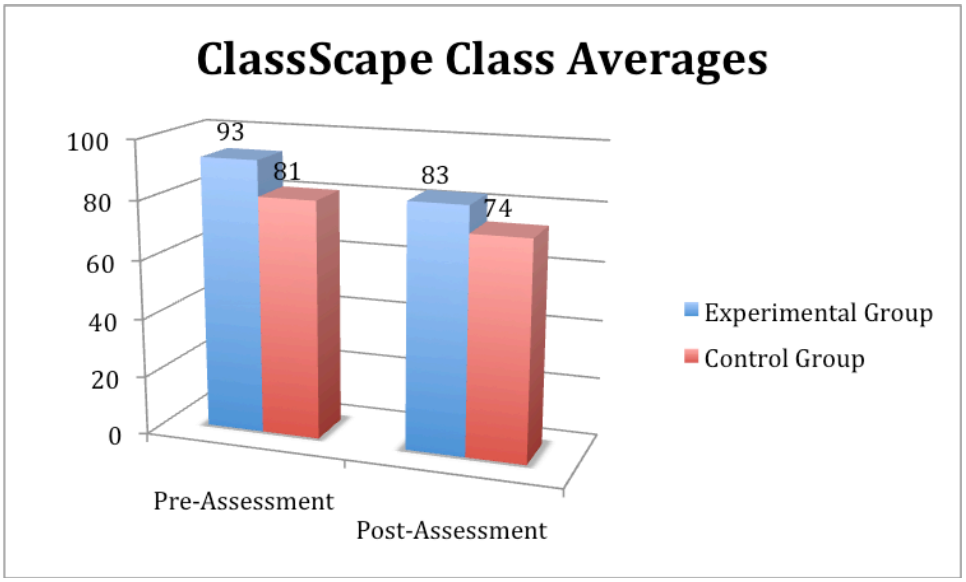
Again, these findings were not expected. I expected students in the experimental group, to show a general increase in their assessment scores, since they were spending the majority of the unit, completing reasoning tasks. Both the control group, and the experimental group decreased in class averages from the pre-assessment and the post-assessment. However, it should be noticed that the control group decreased their class average by 7%, while the experimental group decreased by only 3%. This could show that mathematical reasoning tasks proved to be slightly more successful with students making connections to real-world situations.

ClassScape Assessment

The pre-assessment was comprised of 37.5% “easy” questions, 50% “medium” questions and 12.5% “hard” questions. The assessment tested students on the North Carolina Standard Course of Study (NCSCOS), objective number 4.01a. According to this objective, students should be able to “Use linear functions or inequalities to model and solve problems; justify results. a. Solve using tables, graphs, and algebraic properties.”

The post-assessment was comprised of 43% “easy” questions, 43% “medium” questions and 14% “hard” questions. The assessment tested students on the NCSCOS objective number 4.01a and 4.01b. Objective 4.01b notes that students will be able to “interpret constants and coefficients in the context of the problem."

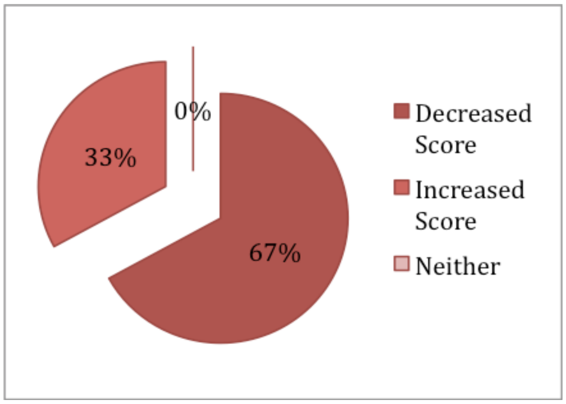
The class averages for the pre- and post-assessment for the control class were 81% and 74% respectively. The class averages for the pre- and post-assessment for the experimental class were 93% and 83% respectively. Therefore, both classes performed lower on the post-assessment. The graph below shows the results.

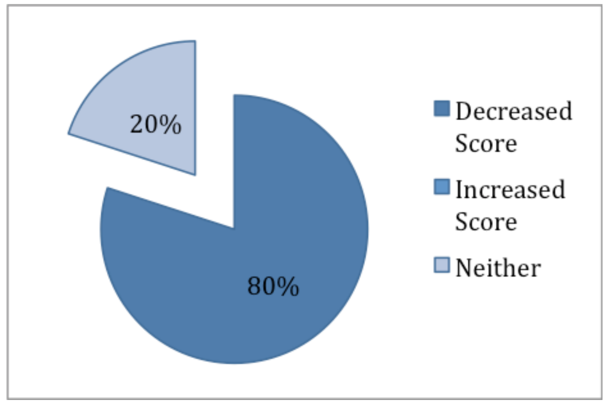


In the control group, 67% of students decreased their score, from the pre- and post-assessment. The remaining 33% of students raised their score. In the experimental group, 80% of students received a lower score on the post-assessment, while the remaining 20% received the same score. The pie chart below shows the results.

This pie chart shows the percentage of students who received a higher score, a lower score, or received the same score between the pre- and post-assessment.

Control Group *Experimental Group*





Obviously these are unexpected results as well. To attempt to hypothesize why there was such a decrease in scores, further analysis required the review of the difficultly level, of the test. These tests are very similar in this category, however there was one difference between these two assessments. The pre-assessment was based solely on North Carolinas Standard Course of Study (NCSOS), objective number 4.01b, which notes that students should be able to "interpret constants and coefficients in the context of the problem." In the post-assessment, only 56% of the assessment was from this objective. The remaining 44% was from NCSOS, objective number 4.01a. From this objective, students will be able to "Use linear functions or inequalities to model and solve problems." I hypothesize that even though the "level" of these questions were the same, being that they were from different objectives, there is still the potential that the test was more difficult, since there were more objectives being assessed. Objective 4.01a may have posed more of a challenge for them.

This hypothesis is consistent with the results from the mid-assessment as well, because students had greater difficulty with objective number 4.01a. In the experimental group, when the focus was directed on finding the "rate of change" of a plant, which is objective 4.01a. The three-day lesson, focused students comprehension on making connections between graphs, tables, and equations. However, students were still were not able to complete the task of 4.01a in the mid-assessment, nor the post-assessments. Therefore, it can be concluded that students, in both the control group, and the experimental group, had more difficulty in "solving using graphs, tables, equations, and algebraic expressions."

Qualitative Data

When the videotapes from each class were reviewed, immediate reflection on the ability levels of each class took place. In the experimental group, the task allowed a greater amount of flexibility to begin with where students are, and build them from that point. For example, in the experimental group, when students were given the lesson on plant growth (found above), I was able to either direct student's attention on points that they are struggling with, or advance their thinking to other concepts that students may not have thought about.

For example, when working with a student, who has a strong sense of mathematics, and it's reasoning, I was able to ask more advanced questions that were not part of the task. When I asked Student 1 (S1), which plant was growing and the fastest rate, she responded in the following manner.

> S1: Plant A because it goes from 2 inches all the way up to 6 inches. And you can do 6 subtract 4, 6 subtract 2 is four, (while pointing to the graph from Plant A), 8 subtract 5, that's only 3 (while pointing to the graph of plant B), 8 subtract 6 is two (while pointing to the graph of plant C), so this one increased the most over the 4 days.

Since S1 had a strong understanding of "rate of change," her thinking needed to be challenged with other concepts that were not yet discussed.

> VanNewkirk: What do these y-intercepts represent?
>
> S1: The lines right here?
>
> VanNewkirk: Yeah, that right there (while pointing to the y-intercept of plant A)
>
> S1: On the, when you first got the plant, how tall the plant was. Like on the zero days,

when you first got the plant, the plant was 2 inches (while pointing to the y-intercept of Plant A), when you first got the plant it was 5 inches (while point to the y-intercept of Plant B), when you first got the plant it was 6 inches (while pointing to the y-intercept of Plant C).

Me: Good.

In comparison to another group who needed more support and guidance, another conversation occurred between two other students, S2 and S3 and myself.

VanNewkirk: Which one has the highest rate of increase for the plants?

S2: Those two (pointing to the graph from Plant B and Plant C)

VanNewkirk: Those two. Okay. How do you know?

S3: Because it comes to a point when they're both the same thing (pointing to the point that the graph of Plant B and Plant C intersect.)

VanNewkirk: What's the same thing about those two plants?

S2: They're the same height

S3: Yeah, in inches.

VanNewkirk: Okay, so when I say which one in growing at the greatest rate, you're thinking that… What determines rate?

S2: Which one's growing faster?

VanNewkirk: Which one's growing faster?

S3: Yeah.

VanNewkirk: So how does them both being 8 say that they're growing faster?

S3: I don't know but we went over it.

S2: Well, I would say that Plant C was growing faster.

S3: At first and then Plant B would be growing faster.

S2: At one point.

S3: Yeah.

VanNewkirk: Okay, well how fast is Plant A, B and C growing? (pause) How fast is Plant A growing?

S3: It's growing by an inch.

VanNewkirk: An inch in how long?

S2: An inch in…

S3: Well right here it's growing by 2 inches (pointing to the table of plant A), and then it goes by an inch.

VanNewkirk: Okay, it grows by an inch. But it doesn't grow by two inches…"

S2: It's already 2 inches.

VanNewkirk: It's already two inches. Okay? It just starts out. When I plant it, it starts out to be two inches. Okay? Plant B. How much is Plant B growing each day?

S3: Three-quarters of an inch.

VanNewkirk: What's that?

S3: Three-quarters of an inch.

VanNewkirk: Three-quarters of an inch. What about Plant C?

S2 & S3: Half inch.

VanNewkirk: So which one is growing at the fastest rate?

(After reflecting on the process, S2 begins to understand the meaning of "growth rate")

S2: Wait, Plant A will be growing faster!

VanNewkirk: Why?

S2: Because it grows by an inch every day! These are only growing by three-quarters of an inch, and a half an inch.

To assist students to develop their connections between graphs, tables and equations, I then drew their attention to the graph, and attempted to push students to think about rate of change, when displayed in a graph.

VanNewkirk: Day. Okay, so this right here (while pointing to the graph), which one, according to this graph is growing at a faster rate?

S2: Plant A still, because it came from one inch, all the way up to 6 inches.

VanNewkirk: In how many days?

S2: In four days. These came from 5 inches to 8 inches in 4 days.

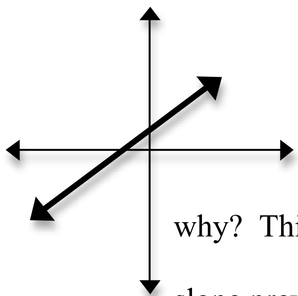
VanNewkirk: Okay, good. So now if I ask you, which one's the tallest plant, you would say…"

S2: Plant B, oh, Plant B and C.

VanNewkirk: OK, good. Now, so you have to be careful with what it means by what's growing at a faster rate."

The conversation continued with pushing these two students towards what the y-intercept represents as well.

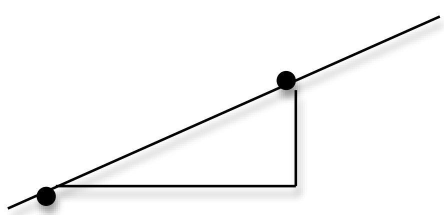
From these two transcripts, one can reflect on the large variances between groups, and how different a teacher must react to these differences. Some students do not require anything besides more concepts to draw connections to. While other students need to be asked a series of scaffolding questions, to help them think about the reasoning behind the mathematics. This dramatic ability for differentiation among students and groups is largely based on the task that was presented. This high-level task, allowed that flexibility for students to solve the problem in multiple ways. I was then able to ask students focusing questions, to direct students thinking. However, these were both great conversations that showed that students were using reasoning, to solve problems.



In the control class, the conversations were much different, and not differentiated according to ability levels. For this particular task, students were asked to determine whether the slope of the line was positive or negative and why? This lesson was the first lesson in this chapter, so students were not introduced to slope previously in this class. Students were introduced to slope in the seventh grade, however, this is the first time this year, this concept was reviewed. Much like the experimental group, all students were given the same task. For example, in a whole-class discussion on this topic, the following conversation occurred.

S1: What are you doing (S2)?

S2: (S2 had just finished drawing the following picture on the board)



S2: It's basically rise over run.

S3: Duh.

S2: But if you put two points on the line, you start from the point that's like nearest to you or whatever and you go to the next point and that will tell you if it's negative or positive.

VanNewkirk: How is that showing that that's positive slope?

S2: Because it's going over and up. It's going over and then it will go up.

S1: It's going UP, so it will be positive.

S4: Yeah but you just said it's starting from the negative and going up. So that means it's starting form negative and going up.

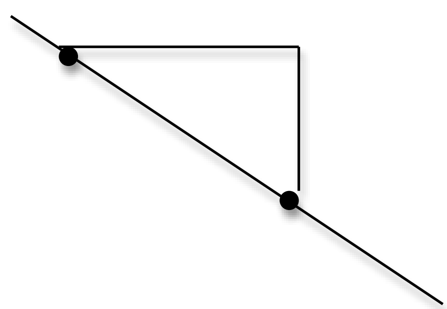
S2: Yeah.

S1: It's not where the line is, because the line goes on forever.

At this point in the conversation, students started to get off-task and began to talk about roller-coasters, so I quickly re-directed their attention.

S1: So you rise up, then run to the point.

VanNewkirk: I'm gonna draw one for you here. I'm gonna draw a negative one, according to your definition right.



S1: Uh huh. You rise up, then you run.

VanNewkirk: So how does that show that it's negative?

S1: But you're going up.

VanNewkirk: I'm going up in both.

S1: Yeah, so they're both positive?

VanNewkirk: They're both positive?

S2: Like really, I think it depends on where the line is if it's like down here (pointing to Quadrant III and Quadrant IV), way down there and the points are like there, then it would be negative.

VanNewkirk: So you don't think my black line is negative?

S4: I do.

VanNewkirk: Okay, hold on a sec real quick. Tell me how rise and run shows me that that line is positive.

S1: Because it's starting from right here, then it goes to the side then up.

VanNewkirk: But how is that rise over run?

S2: You have, okay, I'm not going to draw the boxes, but your line is going like this and you put 2 points, one point nearest you and one point wherever and you wanna get to that point your gonna go over.

VanNewkirk: Okay, over means what? What part of my rise over run?

S2: Oh, wait wait wait. You would go up…

S1: And THEN over!

VanNewkirk: Up would be…

S2: Up would be rise and then run.

Since these two students were on the right track at this point in time, I allowed students to talk to their group members about what they thought about positive and negative lines. It was evident to me, that students would be able to continue without constant supervision because they were beginning to see the connection between the formula (rise/run) and how that related to the graph. The formula gives students a step-by-step process, to give them directions to move from one point to the next. But the question that still remained for these two students to discuss was "What makes the slope of a line positive, and what makes the slope of a line negative?" By the time that I came back to those students, they had came to the conclusion that even though up means positive, the sign of the slope also depends on if you are going "forward" or "backward" as well.

S2: You're going to start at the first point nearest to you and you're going to rise and then you're going to run to the next point.

VanNewkirk: And how does that make it positive?

S2: 'Cause you're running forward.

S1: And then if you do this one (pointing to the graph with a negative slope), you're line is like this. You have your two points here, you're rising up, then you run that way, so you're running backwards, so that would make it negative.

S2: So if you run backwards, it would be negative.

S1: If you're running forward, it's gonna be positive.

These compelling conversations that happen between the students and myself are unique in every aspect. When this research project was set out, the focus was strictly on how assessment scores would be impacted by doing tasks that were driven by mathematical reasoning. It was quickly noticed, however, that when the tasks are changed, it changes the limitations one has on a lesson. When tasks are designed to be flexible, teachers can differentiate the groups and stretch the thinking of all students. On the other hand, when tasks are based from lower-levels, there is little to no flexibility. The students will learn exactly what is taught, and usually, no more.

Conclusions

When conducting lessons that have the potential to be rich in mathematical reasoning, it is important to be prepared for unexpected results. This research question showed those unexpected results. Though the original focus was assessment scores, and how mathematical reasoning tasks affect the assessment scores, this quickly changed. With a large amount of variances in the assessment scores, it is not possible to draw conclusions on how the mathematical reasoning tasks affected or did not affect assessment scores. However, while conducting these lessons, the conclusion that can be made is that the task that is created will drive the types of questions that one can ask students.

In my research, I strictly focused on the tasks that students were given. The tasks of the experimental group had multiple methods to solve the problem, and required students to think through the process. From the well-developed tasks designed in the experimental group, I was able to have deeper conversations that were rich in mathematical reasoning, with students. Students in each group were not asked the same questions. Rather, I took where the students were at, and built them from that point.

Well-developed tasks truly changed the dynamics of the class, in ways that differentiate your lesson, without actually creating different tasks for different groups. Each group worked in the level that they were comfortable with. From this, I noticed that the task that you create determines what you can/cannot ask students. If you create a task that is broad, and could be taken in multiple directions, a teacher can ask questions that stretch even the highest-level of students. On the other hand, if you create tasks that are knowledge driven, the questions you ask students are driven by right/wrong answers. Therefore, students are driven towards only seeking the right answer, instead of reasoning through math, to help it make sense.

Limitations

The focus of this research question was designed to show how mathematical reasoning tasks affect assessment scores. But with such a large amount of variances between assessments, it's very difficult to determine whether or not the increase or decrease in assessment scores was based solely on the tasks. For example, one may review the difficulty of the assessment, or the question design to determine if those two components played a key role in the decrease in assessment scores. One may also hypothesize that there may have been a lack of motivation for the ClassScape assessment, since it was scheduled on a Friday. Therefore, to make the conclusion that mathematical reasoning tasks actually have a negative impact on assessment scores would be false. There is not a sufficient amount of data to support that conclusion. This research plan was conducted on one short chapter. If such conclusions were made, the sample size, and research plan would be tested on a much larger scale.

Next Steps

There were no conclusions made to the initial research question of "Do tasks with mathematical reasoning positively affect assessment scores?" However, there was one question that could be answered from this research question: "How does the task affect the types of questions students can be answered?" To answer the original question however, it would be necessary to conduct research on a much larger scale. For example, instead of completing the research on one chapter, it could be possible to conduct research for an entire year. It is also important to analyze the data for longer as well. How would the focus of mathematical reasoning tasks affect assessment scores for years to come? It would be interesting to track classes throughout high school, to determine any trends found between the assessment scores of students who were enrolled in the "experimental classes" versus the "control classes." Mathematical reasoning skills take a great amount of time to develop for students, so it is necessary to analyze the results on a much larger scale. Therefore, even though this particular research question was left unanswered, other conclusions were made, that are just as vital to the success of students and teachers alike. This research project was not a failure, rather an avenue for much more research to come.

Ball, D. & Bass, H. (2003). Making mathematics reasonable in school. From J. Kilpatrick, W.G. Martin, and D. Schifter (Eds.), *A research companion to the Principles and Standards for School Mathematics* (p. 27-44). Reston, VA: NCTM.

Henningsen, M., & Stein, M. (1997). Mathematical Tasks and Student Cognition: Classroom-Based Factors that Support and Inhibit High-Level Mathematical Thinking and Reasoning. *Journal for Research in Mathematics Education, 28(5),* 524-49. JSTOR.org

Hutchinson, M.J. (2010). When Talking in Class Is a Good Thing: Math Discussions Provide a Window to Student Understanding. *Education Digest: Essential Readings Condensed for Quick Review. 75(8),* 46-48. ERIC.

Smith, M.S. & Stein, M.K. (1998). Selecting and creating mathematical tasks: From research to practice. *Mathematics Teaching in the Middle School, 3,* 344-350.