

COMPARING STUDENT PERFORMANCE AND PERCEPTION OF COMPETENCE  
ON SUMMATIVE SCIENCE PERFORMANCE TASKS VERSUS WRITTEN  
SCIENCE SUMMATIVE TESTS AT THE SIXTH GRADE LEVEL

by

Carli Ruth Barnes

A professional paper submitted in partial fulfillment  
of the requirements for the degree

of

Master of Science

in

Science Education

MONTANA STATE UNIVERSITY  
Bozeman, Montana

July 2015

©COPYRIGHT

by

Carli Ruth Barnes

2015

All Rights Reserved

## DEDICATION

I dedicate this project to my very loyal and loving husband, Stanley. You had many lonely nights as I worked away on this project, my love. I look forward to more time for adventures with you. Thank you for supporting my love of science and learning.

“...and I want to just let you know that I adore you and I know for sure you’re the spark on the sun...” -- Eric D. Johnson

## ACKNOWLEDGEMENT

I want to acknowledge the extensive efforts of my colleagues James Davies, Pamela Enyeart, Jennifer Gwartney, and Ryan Heasley. They spent hours collaborating with me and collecting much of the data I used for this project and this research would not be complete without their phenomenal assistance. My thanks also go to my colleague Christine Jones for recommending the MSSE program. What a quality recommendation! I have thoroughly enjoyed my MSSE education. Thanks a million to all of my fabulous MSSE instructors, especially Dr. Graves. I have grown so much as a writer and science teacher under your guidance. Thank you.

## TABLE OF CONTENTS

1. INTRODUCTION AND BACKGROUND .....	1
2. CONCEPTUAL FRAMEWORK .....	3
3. METHODOLOGY .....	8
4. DATA AND ANALYSIS .....	18
5. INTERPRETATION AND CONCLUSION .....	33
6. VALUE .....	37
REFERENCES CITED .....	40
APPENDICES .....	43
APPENDIX A IRB Exemption .....	44
APPENDIX B Sixth Grade Student Focused List Questionnaire .....	46
APPENDIX C Sixth Grade Teacher Survey .....	49
APPENDIX D Cell Unit Pre-Test .....	51
APPENDIX E Cell Analogy Model Performance Task .....	54
APPENDIX F Cell Unit Written Summative Assessment .....	57
APPENDIX G Earth's Seasons Unit Pre-Test .....	61
APPENDIX H Earth's Seasons Performance Task .....	64
APPENDIX I Earth's Seasons Written Summative Assessment .....	67
APPENDIX J Moon Phases, Eclipses, and Tides Pre-Test .....	71
APPENDIX K Moon Phases, Eclipses, and Tides Performance Task .....	74
APPENDIX L Moon Phases, Eclipses, and Tides Written Summative Assessment .....	81
APPENDIX M Graphing Speed Pre-Test .....	86
APPENDIX N Graphing Speed Performance Task .....	88
APPENDIX O Graphing Speed Written Summative Assessment .....	90
APPENDIX P After Test Student Questionnaire .....	92
APPENDIX Q Sixth Grade Student Post Treatment Interview .....	94
APPENDIX R Sixth Grade Teacher Post Treatment Interview .....	96

## LIST OF TABLES

1. Data Triangulation Matrix .....	17
2. Contingency Table Chi Square Analysis Results for Sample A vs. Sample B's Raw Gains .....	21
3. Chi Square Goodness of Fit Test Results for Written Summative Assessments .....	22
4. Chi Square Goodness of Fit Test Results for Next Generation Science Standard Performance Tasks .....	23
5. Correlation Coefficients for Predicted vs. Actual Posttest Scores for Both Samples .....	24

## LIST OF FIGURES

1. Sixth Grade Focused List Questionnaire .....	18
2. Assessments Sixth Grade Colleagues Use .....	20
3. After Test Student Questionnaire Likert Score of Written Summative Assessment Performance vs. Actual Score, Sample A.....	25
4. After Test Student Questionnaire Likert Score of Written Summative Assessment Performance vs. Actual Score, Sample B .....	26
5. After Test Student Questionnaire Likert Score of Next Generation Science Standard Performance Task Performance vs. Actual Score, Sample A.....	27
6. After Test Student Questionnaire Likert Score of Next Generation Science Standard Performance Task Performance vs. Actual Score, Sample B.....	28
7. After Test Student Questionnaire Questions 4 and 5 Compared to Actual Passing Rate for Written Summative Assessments .....	29
8. After Test Student Questionnaire Questions 4 and 5 Compared to Actual Passing Rate for Performance Tasks .....	29

## ABSTRACT

This study examines the effects on student performance of designing performance tasks based on the Next Generation Science Standards. It also assesses sixth grade students' accuracy in determining their final test scores on both traditional written unit final tests and final performance tasks. Eighty-four sixth graders in four different classes were studied across four different science units. Two classes formed Sample A and two classes formed Sample B. They alternated taking traditional written end of unit tests and taking performance tasks instead. They recorded their predicted scores and their feelings about each test on a survey immediately after testing. The study's results are inconclusive. While the teachers felt less confident preparing students for the performance tasks, students appeared to prefer them, and did moderately better on them. How much better was not found to be statistically significant, though. Students in Sample B consistently outperformed Sample A on all assessments, but Sample A students reported having studied for final tests more often. In all, this study shows how middle school students can be unpredictable and are not always self aware, but do prefer to be engaged in hands-on learning. There may be something to be said for student enjoyment of an assessment, even if the final scores students earn on non traditional tests are only moderately better.

## INTRODUCTION AND BACKGROUND

Educating the whole child is one of my passions as an educator. In my time as an undergraduate ten years ago, the research of Howard Gardner and the idea that all humans have multiple ways of demonstrating their intelligence seemed very logical, and has continued to reflect my experience in working with students over the years (Gardner, 1983).

One of the things I noticed as a science educator, in particular, is how students that may not “shine” in some of the more verbal linguistic subjects, like the language arts, sometimes rise to the top of the class when asked to demonstrate scientific learning. Teaching in a homeroom model regularly allowed me to see the often dramatic variation in individual student’s performance throughout their school day. In fact, students who had difficulty writing an essay seemed to perk up academically and emotionally as soon as the class transitioned to a science lesson. I began to call into question my assessment practices: was I allowing for multiple representations of learning in my classroom?

As this question began to form for me, the Next Generation Science Standards (NGSS) came into prevalence. A colleague in a Walden University masters program and I began seriously discussing the benefits of using performance tasks. She had read about how performance tasks can often more accurately capture students’ true abilities, compared to traditional paper/pencil testing, and I had just begun trying them out during math. When I reviewed the NGSS for middle school, I discovered students are required to demonstrate more of their science knowledge through models and modeling. Since models often allow for multiple representations of student thinking, I decided the NGSS

were pointing me toward a thorough exploration of the benefits of learning how to assess students in differing formats, particularly via performance tasks.

In pursuing the research on performance tasks versus written summative assessments, I also began to wonder if self-awareness of testing preferences played a factor in student success in different testing situations. I had frequently used student surveys after tests or projects to promote students' voice in my class, but I had never intentionally given students two opposite styles of tests and asked them to try and predict their grade before seeing my final grade on their work. Despite the NGSS push for more hands-on testing, I considered it likely some students might prefer traditional written tests to the creative challenge of a performance task. If I became intentional in surveying students' testing experience and their assessment preferences for that particular unit, I might unravel which science topics best lent themselves to a certain style of test.

The backdrop of where I teach is Battle Ground, WA, a rural primarily Caucasian community approximately fifteen miles north of Portland, OR. As of 2014, the population is 18,680 ([cityofbg.org](http://cityofbg.org), 2014). Battle Ground Public Schools' service area is 273 square miles and includes nineteen schools and over 13,000 students. The northern district boundary line is only ten miles south of Mt. St. Helens ([battlegroundps.org](http://battlegroundps.org), 2014). According to L. Ladd, Tukes Valley Middle School (TVM) attendance secretary, TVM's enrollment as of February 2015 is 531, grades five through eight (personal communication, February 9, 2015). In May 2014, the Office of Superintendent of Public Instruction reported 86.5% percent of students identified as white, 7.2% identified as Hispanic, 0.8% identified as Black, and 0.4% identified as Asian or Pacific Islander (Washington State Report Card, 2014). As of January 2015, TVM's free and reduced

lunch rate is 32% (D. Katkansky, personal communication, February 12, 2015). Assistant Principal R. Cowl reported 1.3% percent of TVM students have section 504 plans (personal communication, February 9, 2015), and 3.8% of the study body are identified as transitionally bilingual (M. Gilberts, personal communication, February 9, 2015).

I teach sixth grade to 25 students at TVM with four other sixth grade teachers. All of our classes are self-contained and include reading, writing, math, science, and ancient history instruction. Each class has between 25 and 28 students. There are 136 sixth graders total. One-hundred and eleven participated in the initial survey, and 84 in the study as a whole. Twelve have individualized education plans, and one has a federal 504 accommodation plan.

Given the diversity of learning needs in my class, collegial interest, and my own personal curiosity on best practices, I created the focus statement, *What are the effects on student performance of designing performance tasks based on the NGSS middle school modeling tasks?* Second, *What are the effects of asking students to predict their scores on performance tasks and written summative assessment?*

### CONCEPTUAL FRAMEWORK

The year 2014 is a time of significant instructional change for American middle school science educators. The National Research Council's (NRC) Framework for K-12 Science Education (FFSE) emphasizes scientific inquiry and hopes students will be able to use inquiry to apply scientific skill and knowledge together. The NRC also designed eight K-12 science and engineering practices, including *Developing and Using Models*. It requires science teachers to help students use theories and models in the manner of professional scientists and engineers: to shape hypothesis, find solutions and create

explanations based on observations from the natural world. Students must combine their imagination and reasoning skills by extending existing models or suggesting new models for explaining scientific phenomena. *Constructing Explanations and Designing Solutions*, another NRC science and engineering practice, connects students' imagination to their logic by requiring them to use models to represent the physical world (NRC, 2012). The NGSS are the performance expectations that emerged from the FFSE. Modeling is required 16 times in the middle school Next Generation Science Standards (NGSS) (NGSS Lead States, 2013).

The NRC (2012) defines models as sketches or diagrams that can be physical, simulated, or mathematical. The FFSE differentiates models into two kinds: mental and conceptual. Mental models assist young scientists in building simple constructions of scientific ideas in their minds (Gilbert, 2011, p. IX), while conceptual models can improve students' mental modeling by leading to visualization, understanding, and solutions to problems by "explicitly representing scientific phenomena" (NRC, 2012, p. 56). In light of the NRC's emphasis on models and modeling, middle school science teachers must become proficient at defining models and modeling tasks and implementing effective, best-practice lessons to meet these new performance expectations. Testing practices must be reshaped to include model-based performance tasks in order to prove students are truly meeting NRC expectations.

The work of reshaping tests and preparing students for model-based performance tasks may seem daunting, but neurological research demonstrates middle school science teachers have years of student brain development to draw on to support model-based teaching and testing. For example, one neurobiology research report suggests conceptual

models can be tools for revising mental models. The report also discusses how image schemata are the brain's intuitive knowledge of an image or action and can be aroused via language (Rohrer 2005). Conversely, images or actions can arouse language, as well. Rohrer cites the research of Arterberry and Bornstein (2001) and Bertenthal (1993) as revealing that certain image schemata are present as young as three months of life. Rohrer also shares evidence for image schemata beginning in utero when babies begin to hear their mother's speech patterns. Given the language-rich environment of the classroom, teachers may be able to use class discussion to arouse students' intuitive image schemata in relation to science concepts at hand. Piaget's schema concept, as revisited by Bodner (1986), is also a reminder that students naturally build knowledge in their minds. Teachers can help students adapt pre-existing cognitive structures to the new science information at hand by providing comprehensible "keys" reusable for students' mental "locks" (Bodner, 1986, p. 873-874).

Other researchers inspired by Piaget's idea that knowledge is constructed in the mind of the learner, known as constructivism, include Niebert, Marsch, and Treagust (2012). They suggest a philosophical extension of constructivism is experientialism, which is the idea that human's everyday embodied experiences create basic schemata. Schemata from these embodied experiences could be something science teachers can use as building blocks when exploring new science concepts students. One building block is their definition of analogies, which are something science teachers might find themselves using to help create mental models for students. They state that analogies are direct experiences referenced to help understand something abstract, they indicate quality analogies require science teachers to be knowledgeable of their students' prior

foundational embodied experiences, such as up versus down, or contained versus not contained (Niebert, et al., 2012). Building on students' formative experiences through this type of analogical thinking might be another way science teachers can design quality model-based lessons.

If students' mental schemata have been accounted for during instruction, experientialism's embodied experience theory can also be referenced to point towards hands-on assessments as a way to allow students to live out the models in their minds. Implementing a greater number of performance tasks, which can often be very hands-on, could be a worthwhile pursuit for science teachers. Hibbard defines performance tasks as assessments that help students pull together content knowledge, process skills, and work habits. Performance tasks can range from taking only a few minutes to end-of-unit culminating projects (Hibbard, 1996). The research of Accongio and Doran (1993) was released 20 years before the NGSS but still speaks to the modern science teacher. They argue that while paper and pencil assessments are useful for assessing students' factual knowledge and vocabulary, they poorly assess the application of activities real scientists actually perform. It is essential teachers help students move away from an obsession with one right answer and into solving complex problems. They devote an entire chapter to recommendations for designing effective performance tasks, and state "It is profoundly important for science teachers to take as much care with test development and administration, and interpretation of collected data as they do in the preparation and delivery of their classroom instruction" (p. 2). They also state many different types of measurements must be taken regularly to examine the efficacy of lessons, to seek the rate of student learning, and to provide "proof of learning" (p. 4).

However, performance tasks may not be a perfect panacea for modern assessment needs. Shymansky et al.'s 1997 study compared ninth grade science students' performance on a series of active tasks to their performance on related items of the Iowa Tests of Educational Development, a paper and pencil multiple-choice test. The researchers found the students did not perform better on performance tasks than on the traditional test. Their study suggests prior experience with performance tasks may be a contributing factor to student performance. Noting that their study might have "produced more questions than answers" (p. 182), they acknowledge that "the business of developing meaningful and doable performance assessments is no simple task" (p. 182), and cite Champagne (1994) as suggesting there is potential in pursuing the creation of non-traditional assessments.

With the NGSS being a fairly new construct, and performance tasks having a mixed history of implementation, one area that may help science teachers improve performance task design could be seeking students' perception of their performance on performance tasks, particularly in comparison to how they perceive their performance on traditional written assessments. Simpson, Licht, Wagner, and Stader (1996) examined fifth grade students' perceived and actual abilities in math, reading, and overall general education. They found students' perception of performance differed according to how they were asked to define it. Students' perception of ability appeared to depend on the academic area. While this study did not include perception about science instruction in particular, the idea that students can differentiate performance between subjects might be useful feedback in performance task design. To refer back to Accongio and Doran (1993), many types of data must be collected to represent what is occurring in all aspects

of student learning. Making students aware of their life-long embodied experiences may support implementing the NGSS. Performance tasks, in combination with traditional assessments, may provide a window into how well students are learning the new standards. Student perception may be the final piece of the puzzle of understanding student growth in math and science.

### METHODOLOGY

The primary focus of this study was to analyze how students perform on two types of assessments whose design was based on the Next Generation Science Standards (NGSS). The NGSS include many modeling tasks that are best assessed as performance tasks. However, middle school students may have limited experience with this assessment modality. I was interested in seeing if students would perform better on traditional paper-and-pencil assessments. I also studied whether students could accurately predict their performance on both types of final exams and/or could state a preference for how they would have wanted to be assessed on a unit of study. The research methodology for this project received an exemption by Montana State University's Institutional Review Board and compliance for working with human subjects was maintained.

(Appendix A).

The initial student sample size was 111 students. TVM has five sixth grade classes that are each self-contained and study math, science, reading, writing, and history. Each class has between 25 and 28 students. The treatment portion of this study included my class and three of the other sixth grade classes, although it was intended to include all five classes at the outset. The fifth class's teacher was not able to implement all treatment procedures, so her class's data was only represented in the pre-treatment survey. For the

treatment, the sample size was 84 students. While there were more than 84 students in the four sixth grade classes who experienced the treatment, just 84 were present for all four units.

The pre-treatment period began the second week of school. All five sixth grade classes took the Sixth Grade Student Focused List Questionnaire (SGSFLQ), on which students recorded different types of tests they were aware of in all aspects of their lives and their personal testing preferences (Appendix B). It was an open-ended, qualitative measure of students' knowledge of tests, testing preferences, and testing procedures. Students were given three minutes to list different ideas related to how they had been tested in school, sports, or other settings over the course of their lives. They were advised to list the names of tests or to state ways they had been tested. Students were encouraged to record anything they might consider a test, even if not certain. They were given 20 boxes on the instrument to write their ideas in, with additional space for extra ideas. After three minutes, students were asked to circle in their list of test ideas all the ways they enjoyed or preferred to be tested. On the back of the SGSFLQ, students were asked, "In a perfect school, how would you like teachers to test you?" with a reminder that tests are just ways of communicating to a teacher what has been learned and a request to avoid writing, "No tests." They were given six large response boxes to record their ideas. Students' responses to all three portions of the SGSFLQ were grouped thematically according to either the type of test they stated (a specific content area, a standardized test, etc.), the format and grading style of the test, or according to ways they desired to be tested. An additional miscellaneous category for non-sequitur responses was found to be

necessary. The numbers of student responses for each thematic category were organized into a bar graph.

During the pre-treatment time period, the three sixth grade teachers whose classes participated in the treatment took the Sixth Grade Teacher Survey (Appendix C). It surveyed their prior knowledge and preferences in assessing students. It was a qualitative measure. Results of this survey were collected and used as points of comparison when analyzing students' responses to the SGSFLQ. Teachers' responses were also used in comparison to teachers' answers in a post treatment interview.

All treatment data was collected during each class's science class period. The treatment period began the seventh week of the school year. Classes needed time to learn proper microscope use and engage in a series of beginner microscope labs in order for science Unit One, Cells and Cell Systems, to launch successfully. Two of the classes became Sample A, while the other two became Sample B. Sample A used one post assessment strategy and Sample B used another, both of which are explained below in the context of the teaching units. The treatment period was four units long so that students in Sample A and Sample B could each have two turns taking each of the post assessment strategies. The samples alternated turns at each strategy so that one group always functioned as a treatment group while the other acted as the non-treatment group.

For Unit One, Sample B took the performance task (PT) post assessment strategy at the close of the unit, while Sample A used the written summative assessment (WSA) post assessment strategy. Unit One was a 30-day unit of study. To establish baseline prior knowledge, classes began Unit One with the Cell Unit Pre-Test (Appendix D). Students'

scores on this instrument were recorded as a percentage and used to calculate raw gain in comparison to their scores on the PT or WSA at the end the unit.

Sample A and Sample B alternated taking a WSA and a PT throughout this treatment. For Unit One and Unit Three, Sample A took a WSA, while Sample B took a PT. For Unit Two and Unit Four, Sample A took a PT while Sample B took a WSA.

Sample B's PT preparation was included in their Unit One instruction. While Sample A used traditional study methods such as reading the Prentice Hall science text to learn about cells, Sample B began Unit One by making cell models. After all students in Sample B presented their models to their classes, Sample B's teachers introduced the posttest Cell Analogy Model PT (Appendix E). Sample B students had two weeks to work individually to create an analogous system for all functions of either a plant or animal cell. Sample B's teachers were allowed to offer an initial analogy, such as "The cell is like a city," and were able to caution students against analogous models that were not systems, such as the impossible system analogy "The cell is like a cookie." When Sample B's teachers ensured each student's analogy was indeed based on a system, students created their model entirely on their own. When the Cell Analogy Model PT was complete, students in Sample B were assessed using its grading rubric. Students' scores were recorded as percentages and were compared pre/post using raw gain.

Sample A's WSA took less time to administer than the Cell Analogy Model PT. Reading and studying the animal and plant cells systems using various strategies took 14 days, compared to the full 30 days needed by Sample B. At the end of 14 days, Sample A took the posttest Cell Unit WSA (Appendix F), then proceeded to either a non-science unit or extended study of the cell. Data from Sample A's extended study of the cell was

not used in my research. Students' scores from this posttest were recorded as percentages and used to calculate raw gain pre/post.

Sample A's raw gains were compared to Sample B's average raw gains using a Contingency Table Chi Square Test that compared the number of students in Sample A who had a raw gain of at least 1% or greater to the number of students in Sample B who had a raw gain of at least 1% or greater. Chi Square tests are appropriate measures for smaller study populations, such as the one in this study ( $N = 84$ ). The null hypothesis tested for Unit One was that the students in Sample B performed equal to or worse than the students in Sample A. The alternate hypothesis tested was that the students in Sample B performed better than the students in Sample A. The null hypothesis was assumed, and a confidence level of  $\alpha = 0.05$  was used to determine if the likelihood of getting data like this was very low. If it was determined to be this low, then I was able to reject the null hypothesis and assume there was some significance to the difference in test scores between Sample A and Sample B.

A Contingency Table Chi Square Test was used to analyze the significance of difference between WSA scores and PT scores for the three additional science units. The same null hypothesis and confidence level were also used.

Unit Two of the treatment period focused on the study of the seasons. It began the day students returned from winter break, 15 weeks into the school year. Unit Two was a 10-day unit. To begin the treatment, all classes took Earth's Seasons Unit Pre-Test (Appendix G). Students' scores from this pre-test were recorded as percentages and used to calculate average normalized gain pre/post after the PT and WSA were given at unit's close. For this unit, Sample A and Sample B received similar instruction on the reasons

for Earth's seasons. Activities included graphing one year of temperature data from different parts of the world, graphing on year of monthly sunrise and sunset times from different parts of the world, and other reading or data based study activities. At the close of Unit Two, students in Sample A took the posttest Earth's Seasons PT (Appendix H). Sample B took the posttest Earth's Seasons WSA (Appendix I). Students' percentage scores on the Earth's Seasons PT and WSA were used to calculate each students' raw gain from their Earth's Seasons Unit Pre-Test percentage score.

Unit Three immediately followed Unit Two. It was a 25-day unit. Samples A and B began the unit with a pre-test called Moon Phases, Eclipses, and Tides Pre-Test (Appendix J). Students' scores from this pre-test were recorded as percentages and used to calculate raw gain pre/post after the PT and WSA were given at unit's close. Both samples received similar instruction, including analyzing moon data students collected via moon logs, Prentice Hall science readings, YouTube videos, live model-based simulations in a dark room with a simulated light bulb "sun", and graphing marigrams of tidal data.

At the close of Unit Three, Sample B took the posttest Moon Phases, Eclipses, and Tides PT (Appendix K), which was a group project that included a specific grading rubric. Sample B's rubric score percentages were used to calculate raw gain from their initial unit pre-test score. Sample A took the posttest Moon Phases, Eclipses, and Tides WSA (Appendix L). Their percentage scores on this WSA were also used to calculate their raw gain in pre/post.

Unit Four immediately followed Unit Three and was the final phase of treatment. It was a 10-day unit. To begin Unit 4, both samples took the Graphing Speed Pre-Test

(Appendix M). Students' scores from this pre-test were recorded as percentages and used to calculate average normalized gain pre/post after the PT and WSA were given at unit's close. Both samples received similar instruction based on the Prentice Hall science text, a motion lab that required them to graph different speeds they traveled, and practice problems interpreting speed graphs or graphing speed scenarios. At the close of Unit Four, Sample A took the posttest Graphing Speed PT (Appendix N). Sample B took the posttest Graphing Speed WSA (Appendix O). Students' percentage scores on this unit's PT and WSA were used to calculate each students' raw gain from their Graphing Speed Pre-Test percentage score.

An additional component to this research was measuring student perception of final test performance at the close of each of the four units. To measure this, the After Test Student Questionnaire (ATSQ) (Appendix P) was given to students in both samples to complete after each unit's posttest. Students turned it in with their test. This instrument consisted of six questions.

Question one regarded the students' predicted score on the assessment, ranging from the letter grade F (defined as 0%--59%) to the letter grade A (defined as 90%--100%). A Chi Square Test for Goodness of Fit was used to assess the significance of how many students expected to pass PTs compared to WSAs, with the terminology "expected to pass" including students who underestimated their score. "Passing" was defined as a score of a C (70%) or better. Four Goodness of Fit tests were run. One tested the significance of the percentage of students in Sample A that expected to pass the Unit One and Unit Three WSAs compared to the percentages that actually did. A second tested the significance of the percentage of students in Sample A that expected to pass the Unit Two

and Unit Four PTs compared to the percentages that actually did. A third tested the significance of the percentage of students in Sample B that expected to pass the Unit One and Unit Three PTs compared to the percentages that actually did. A fourth testing the significance of the percentage of students in Sample B that expected to pass the Unit Two and Unit Four WSAs compared to the percentage that actually did. A significance level of  $\alpha=0.05$  was used to analyze the results to all four Goodness of Fit test results, with the null hypothesis that the students in Samples A and B correctly determined their ability to pass their final unit tests, and the alternate hypothesis that students did not correctly determine their testing abilities.

To further analyze these findings, I created two correlation graphs: one to track the relationship between PT predicted scores and PT actual scores for both samples, and one to track the relationship between WSA predicted scores and actual scores for both samples. A  $R^2$  statistic for each graph was calculated to assess how well students' predicted PT and WSA scores matched their actual PT and WSA scores.

Question 2 of the ATSQ was a Likert scale question that asked students to circle whether they felt they did very badly, badly, were unsure, did okay, well, or very well on the assessment. I assigned responses a score on a scale of six, with the response "very badly" receiving a one, and the response "very well" receiving a six. If two ideas were circled, such as "badly/unsure," an average of the two was given. To enhance a visual comparison between students' perceptions and the reality of their scores, I assigned their actual posttest score a value ranging from one to six, as well. Scores 59% and lower were rated a one, scores 60%-69% were rated a two, scores 70%-75% were rated a three, scores 76%-85% were rated a four, scores 86%-94% were rated a five, and scores 95%-

100% were rated a six. Then I organized students' responses and their actual scores for display in two comparative box and whisker plot graphs, one to display Sample A and B's actual and perceived WSA performance, and one for their PT performance.

Question three of the ATSQ asked students to explain their answers to questions one and two. Students' feelings regarding their assessments were analyzed for common trends between students who actually passed the posttest with 70% or better and students who did not. Their responses were used as a commentary to enhance the analysis of students' predicted versus actual posttest outcomes.

Questions four and five of the ATSQ were analyzed in unison. On question four, students were asked to circle yes or no regarding whether they studied for the test. On question five, students were asked to circle whether the posttest they had just taken was a silent independent written test or a project, activity, or other test. To analyze whether studying and awareness of type of posttest had any relationship to passing the posttest, I created two bar graphs. One bar graph compared the percentage of Sample A who studied, passed, and correctly identified the PTs to Sample B, and the other compared the percentage of students in Sample A who studied, passed, and correctly identified the WSAs to Sample B. I again used 70% as the definition of passing.

Question six of the ATSQ asked students for ideas on other ways they might have wanted to have been tested on the unit's topic. Much like question three, answers from question six were used as commentary on the results of question one, two, four, and five's analysis.

To close this study, two students, one boy and one girl, were selected at random from each of the four treatment classes to be interviewed regarding their experiences in

Units One through Four using the Sixth Grade Post Treatment Interview (Appendix Q). Students' responses were used as commentary on the results of the posttest analysis and the ATSQ analysis.

My three participating colleagues were also interviewed regarding their experiences with the treatment using the Sixth Grade Teacher Post Treatment Interview (SGTPTOF) (Appendix R). Their answers were compared to their initial answers on the Sixth Grade Teacher Survey (Appendix D) and were used as commentary on the results of the posttest and ATSQ analysis.

A summary of the instruments is found in the Data Triangulation Matrix (Table 1).

Table 1  
*Data Triangulation Matrix*

Focus Question	Data Source 1	Data Source 2	Data Source 3
1. <i>Primary Question</i> How do students perform on performance tasks related to NGSS concepts compared to standard written summative assessments testing the same concepts?	Pre/Post Science Unit Tests (Appendices D-O)	After Test Student Questionnaire (Appendix P)	Sixth Grade Student Focused List Questionnaire & Sixth Grade Student Post Treatment Interview (Appendices B & Q)  Sixth Grade Teacher Survey & Sixth Grade Teacher Post Treatment Interview (Appendices C & R)
2. <i>Sub-question</i> Do students accurately perceive & predict their performance on performance tasks versus written summative assessments?	Pre/Post Summative Science Unit Test (Appendices D-O)	After Test Student Questionnaire (Appendix P)	Sixth Grade Student Focused List Questionnaire & Sixth Grade Student Post Treatment Interview (Appendices B & Q)

## DATA AND ANALYSIS

The results of the Sixth Grade Student Focused List Questionnaire indicated a wide variety in student awareness of what they considered a test, what types of tests students found preferable, and how students wished teachers would test them ( $N=111$ ). Figure 1 shows the variation in student response. Students were allowed multiple responses to each of the three components of this questionnaire, so each bar represents the total number of times the idea was stated. Students appeared to have the most experience with tests in sports and physical education, with many students naming multiple sports on a single questionnaire. In all, sports and physical education tests were named as having been experienced 222 times. Tests in this category were also most preferred by students, with 71 sports and physical education tests circled. Students also seemed to have a strong memory for subject-specific tests, with many students simply stating, “math” or “writing” on their questionnaire as a form of a test.

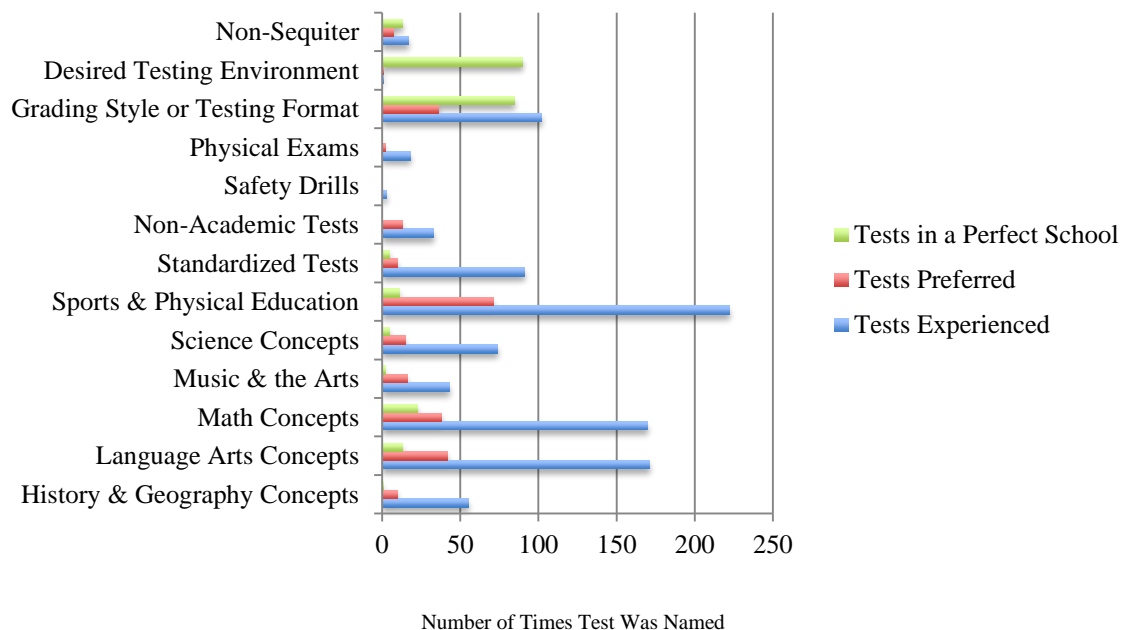


Figure 1. Sixth grade student focused list questionnaire, ( $N=111$ ).

Students had many responses that I categorized as “desired testing environment” that were provided for the third part of the questionnaire. Requests for how students wished their teachers would test them in a perfect school included automatic A’s, the ability to eat during tests, tests that allowed students to talk to their friends, limiting the number of tests, and simple requests, like games, using a computer, or worksheets. One student requested, “unlimited time writing tests,” while another wanted, “one big test at the beginning of the year and that’s it.”

Outliers in the responses included the physical exams, which contained tests like hearing or vision tests. None of these tests were listed as being of use in a perfect school. Safety drills were also listed just three times as tests experienced, but were never noted as preferred or existing in a perfect school. The nonsequiter category represents nonsensical answers provided to the questionnaire. There were 17 responses to tests experienced that were either illegible or were responses like, “I hate tests.” There were 13 responses to what kinds of tests should exist in a perfect school that fell into a similar vein. Some students wrote multiple nonsensical answers, so this does not represent individual surveys. One example of a test for a perfect school that I recorded as nonsensical was, “A test that how long can I stay outside with friends.” Also notable is that standardized tests were named 91 times as tests experienced, but only listed 5 times as tests for a perfect school.

Teacher responses to the Sixth Grade Teacher Survey (Appendix C) were concise ( $N=3$ ). Teachers listed quizzes, drawing name sticks, thumbs up/down, whiteboards, presentations, skits, making art, projects, informal observations, and written tests as types of assessments they give in their classrooms. When asked what types of assessments best

show them what students have learned, one teacher stated, “I feel I get a clearer picture when students are able to write, explain, and apply what they’ve learned,” while another shared, “Being able to observe students answer a question or see how they share information says more about how they understand a concept than just filling a bubble.” All three teacher participants said they were comfortable designing and/or giving written summative assessments, while two of them said the same for their comfort designing and/or giving students performance tasks. One stated they were unsure about designing and/or giving performance tasks.

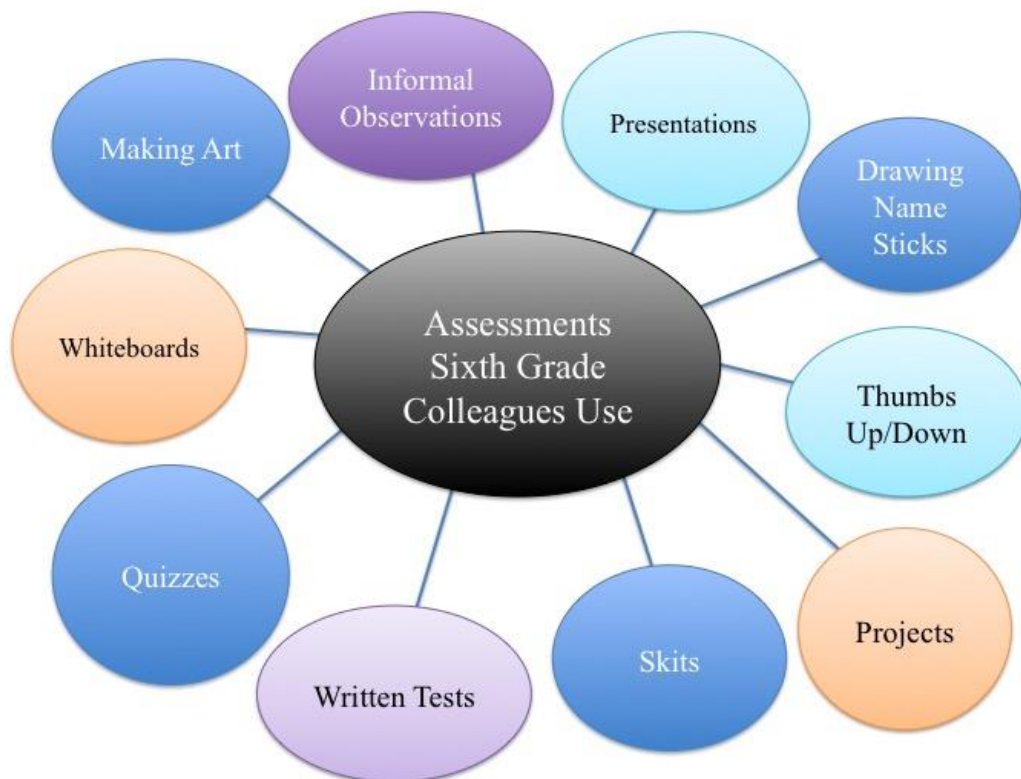


Figure 2. Assessments Sixth Grade colleagues use, (N=3).

The results of the Contingency Table Chi-Square analysis for students’ raw gain pre/post in Units One through Four indicated an inability to disprove the null (N=84).

One degree of freedom was used for this analysis. A chi square statistic greater than  $\chi^2 = 3.84$  was needed to disprove the null. Table 2 shows how Sample A's raw gains in Unit One, Three, and Four did not have a statistically significant difference between Sample B's.

Table 2

*Contingency Table Chi-Square Analysis Results for Sample A vs. Sample B's Raw Gains*

	Unit One	Unit Two	Unit Three	Unit Four
Sample A's Posttest	WSA	PT	WSA	PT
Sample B's Posttest	PT	WSA	PT	WSA
$\chi^2$	2.756	42.000	0.000	1.415
Greater than 3.84?	No	Yes	No	No
Null disproved?	No	Yes	No	No

*Note.* WSA=Written Summative Assessment, PT=Performance Task. Sample A, ( $n=38$ ). Sample B, ( $n=46$ ).

These results indicate that Unit Two was able to meet the criteria for the alternative hypothesis, meaning that there was a significant difference between Sample A's raw pre-to-post gain compared to Sample B's. In this case, 71% of Sample A had a pre/post raw gain of 1% or greater, while 35% of Sample B had the same pre/post raw gain.

However, the pre/post raw gains had no consistent pattern that would indicate a Performance Task (PT) written from the NGSS standards and used as a posttest made a difference compared to the raw gains made when students took the Written Summative Assessment (WSA) posttest. In Unit One, for example, 95% of Sample A had a pre/post raw gain of 1% or greater after taking a WSA posttest. Sample B took a PT posttest and had 100% of students make the same gain. This 5% difference is supported as statistically

insignificant by the chi-square analysis. A similarly insignificant 3% difference between samples was found for Unit Four. In Unit Four, 97% of Sample A had a pre/post raw gain of 1% or greater after taking a PT, while 100% of Sample B had the same gain. In Unit 3, there was no difference between Sample A and Sample B: in both samples, 100% of students had raw gains pre/post of 1% or greater. In all, the chi-square analysis helped assert that taking a NGSS-based PT posttest versus a WSA posttest appeared to have no significant effect on pre/post raw gain.

The results of the Chi Square analysis for Goodness of Fit between students' predicted posttest scores on the After Test Student Questionnaire (Appendix P) and their actual posttest score indicated mixed results. One degree of freedom was used, and to demonstrate significance, the test statistic found had to be greater than  $\chi^2=3.84$  in order to disprove the null hypothesis that students were correctly able to determine their testing abilities. Tables 3 and 4 show the results of the Goodness of Fit analysis.

Table 3  
*Chi Square Goodness of Fit Test Results for Written Summative Assessments*

Sample A (n = 38)			Sample B (n=46)		
	Unit One	Unit Three		Unit Two	Unit Four
% who expected to pass	89%	97%	% who expected to pass	96%	100%
# who actually passed	1	17	# who actually passed	15	31
$\chi^2 = 42.5$ , Null Rejected			$\chi^2 = 24$ , Null Rejected		

Note. (N=84).

Table 4  
*Chi Square Goodness of Fit Test Results for Performance Tasks*

Sample A (n = 38)			Sample B (n=46)		
	Unit Two	Unit Four		Unit One	Unit Three
% who expected to pass	97%	89%	% who expected to pass	93%	100%
# who actually passed	32	24	# who actually passed	38	44
$\chi^2 = 3.49$ , Null Cannot Be Rejected			$\chi^2 = .667$ , Null Cannot Be Rejected		

*Note.* (N=84).

The results are polar. Both Samples A and B were reasonably able to predict passing their posttests when they took the NGSS-based PT posttests, but were also both unable to predict their WSA posttest scores. Question three on the ATSQ asked students to offer insight into their predicted outcome. For the WSAs, there was no definitive pattern in passing versus failing students' concepts of what their actual score would be. In both samples, students who failed reported a wide range of explanations for their predicted scores. Some said, "it was kind of easy," or made comments like, "I feel really good." They mentioned having studied. Others stated, "I got stuck on a couple problems," or, "cause science is not my strong suit." One student who failed stated, "This test was not very good because if we don't do a lot of fun activity things don't 'click'."

Students who passed the WSAs also expressed both confidence and ease with the testing, with one pointing out, "because I could remember everything." However, there were a few passing students who doubted their final scores, with comments like, "it was confusing!" and, from the only student in Sample A who passed the Unit One WSA, "worried about sloppy handwriting." The wide array of comments from both passing and

failing students in both Samples A and B support the chi square analysis that students could not predict their outcomes on the WSA with much success.

The chi square analysis found that for the NGSS-based PTs, both samples were better able to predict their final scores. Student comments on ATSQ question three did indicate more congruency between passing and comments that realistically explained how students in both samples passed. Less students overall failed this type of assessment, and those who did not pass wrote comments like, “Because I don’t understand it,” and “I didn’t really know what to do.” A large amount of confidence was exuded by the passing students, with some providing very specific evidence in statements like, “We got all the data we need, so I feel great about the test,” and “I work very hard and double checked many times.”

In addition to chi square analysis, I also analyzed students’ predicted and actual scores for any significant correlation by finding the  $R^2$  statistic for each sample in each of the units. An  $R^2$  statistic of zero indicates no correlation, whereas a statistic of one indicates perfect correlation. Some correlation between predicted and actual outcome was found, but none of large significance. Table 5 is a record of each sample’s correlation coefficient.

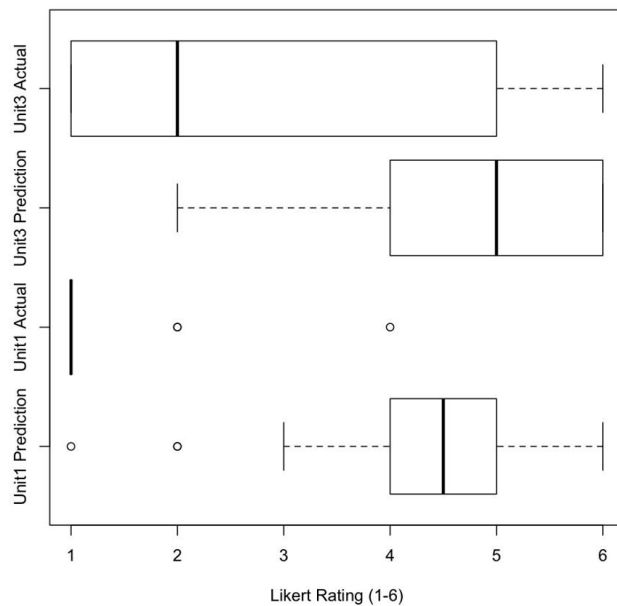
Table 5  
*Correlation Coefficients for Predicted vs. Actual Posttest Scores for Both Samples*

Unit One			Unit Two			Unit Three			Unit Four		
	A, WSA	B, PT		A, PT	B, WSA		A, WSA	B, PT		A, PT	B, WSA
$R^2$	.09	.28	$R^2$	.35	.41	$R^2$	.006	.36	$R^2$	.10	.15

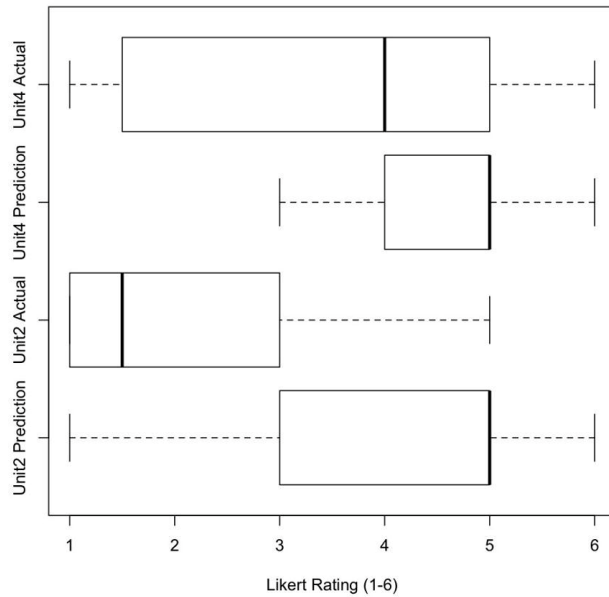
*Note.* A=Sample A, B=Sample B. (N=84).

For question two of the ATSQ, where I asked students to circle how they felt the test went for them on a six-point scale ranging from very badly to very well, I compared

the students' perception of their posttest performance in Units One through Four using box and whisker plots. Figures 3 and 4 show the disparity between actual and perceived performance on WSAs. All four WSAs given were paper and pencil tests with vocabulary, fill-in-the-blank, short answer, and short essay questions often taken right from the students' textbooks. Students in both samples tended to overrate their performance, with most guessing they would earn a C or higher, when the reality was that many received failing final scores.

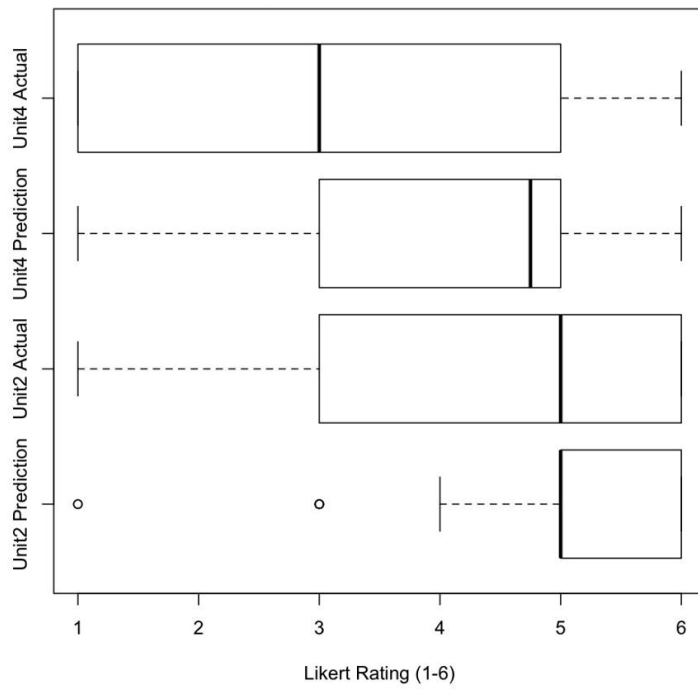


*Figure 3.* After Test Student Questionnaire Likert score of written summative assessment performance vs. actual score, sample A, ( $n=38$ ). Likert Rating Key: 1= very badly, 2= badly, 3= unsure, 4= okay, 5= well, 6= very well.

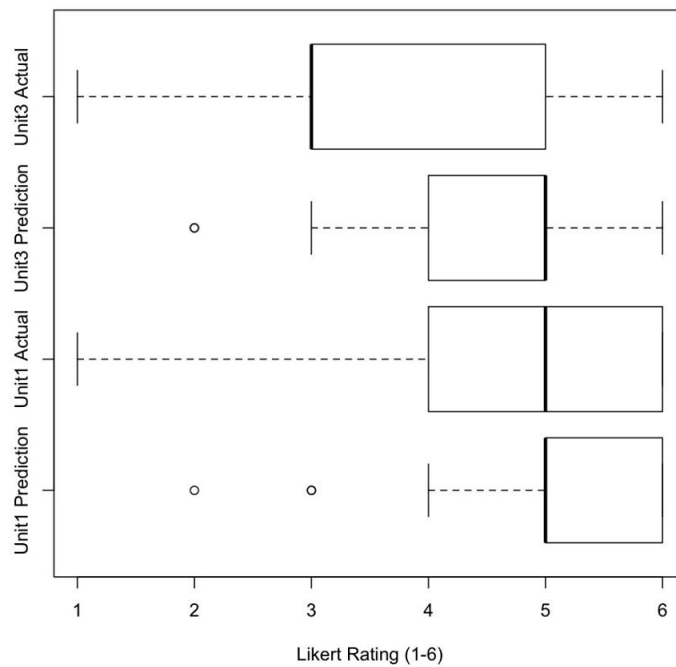


*Figure 4.* After Test Student Questionnaire Likert score of written summative assessment performance vs. actual score, sample B, ( $n=46$ ). Likert Rating Key: 1= very badly, 2= badly, 3= unsure, 4= okay, 5= well, 6= very well.

Figures 5 and 6 reveal that students were moderately better at correctly predicting their NGSS PT posttest scores, but that their predicted scores were not as diverse as their actual scores, with many still over predicting their final scores. The NGSS PTs given ranged from independent to group activities, but all had a strong creative basis. Presenting findings, drawing, graphing, or building were all components of the PTs.

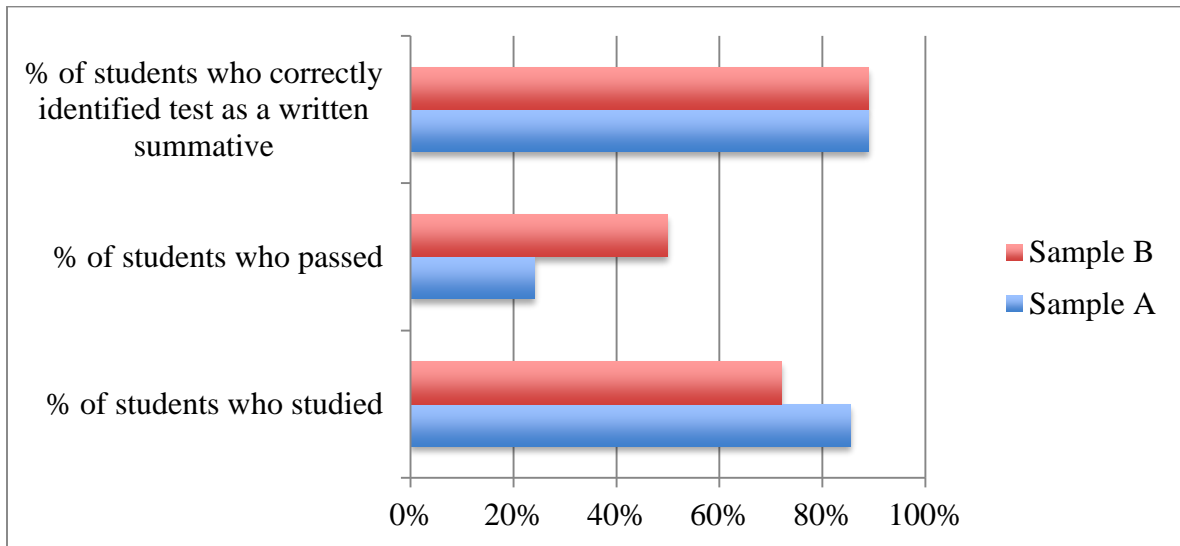


*Figure 5.* After Test Student Questionnaire Likert score of Next Generation Science Standard performance task performance vs. actual score, sample A, ( $n=38$ ). Likert Rating Key: 1= very badly, 2= badly, 3= unsure, 4= okay, 5= well, 6= very well.

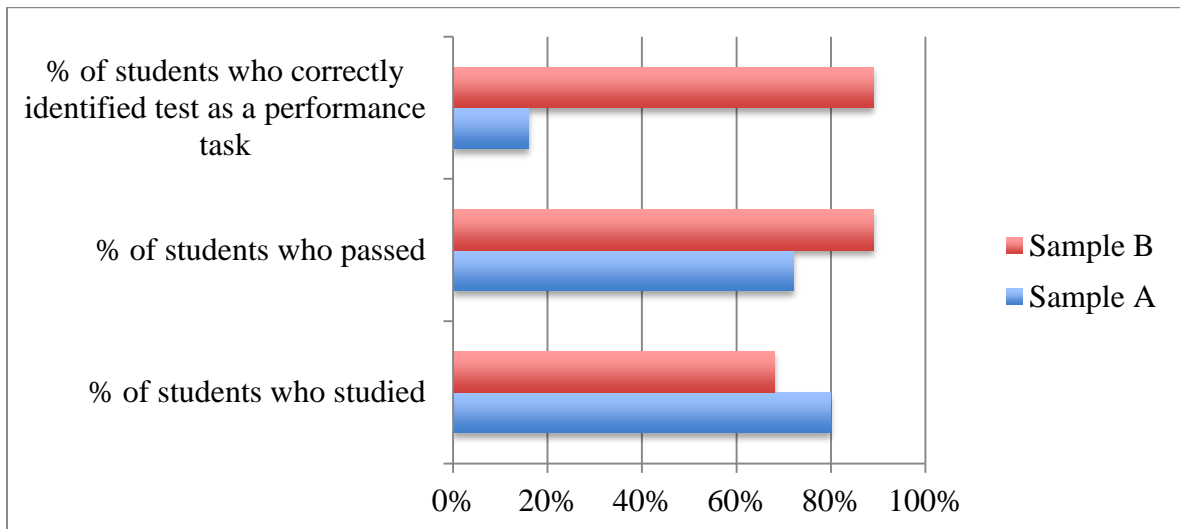


*Figure 6.* After Test Student Questionnaire Likert score of Next Generation Science Standard performance task performance vs. actual score, sample B, ( $n=46$ ). Likert Rating Key: 1= very badly, 2= badly, 3= unsure, 4= okay, 5= well, 6= very well.

The ATSQ also asked students to state whether they had studied for the posttest and to circle whether they had just taken a silent independent written test or completed a project, activity, or other test. I graphed the outcomes of this question in comparison with the percentage of each sample who actually passed with the posttest with a 70% or greater. Figure 7 is the comparisons between studying, passing, and correctly identifying the two WSAs, and Figure 8 offers the same comparisons for the two PTs.



*Figure 7.* After Test Student Questionnaire questions 4 and 5 compared to actual passing rate for written summative assessments, ( $N=84$ ).



*Figure 8.* After Test Student Questionnaire questions 4 and 5 compared to actual passing rate for Next Generation Science Standard performance tasks, ( $N=84$ ).

Both samples had much greater success identifying WSAs correctly than PTs.

Eighty-nine percent of both samples correctly identified their WSA posttests, but while 89% of Sample B correctly identified their PT posttests, just 16% of Sample A did so.

The percent of students who stated they had studied did not vary greatly between PTs and WSAs, but the percent that passed did. At least 39% more of each sample passed their

PTs. Despite being larger, Sample B also consistently outperformed Sample A on both types of assessments. However, about 12% more students in Sample A stated they studied for each type of assessment than in Sample B.

The final question of the ATSQ asked students to suggest how, in a perfect school, they would like to have been tested on the unit of study. A common theme among responses after both type of posttest was “IDK” and “I dunno” type responses, or stating they didn’t think the test should be any different. One student stated after their test in Unit Four, “There’s no such thing as a perfect school, but I don’t really know. All tests are boring.” However, many students suggested improvements similar to their ideas on the Sixth Grade Focused List Questionnaire. They requested technology, someone to work with, for things to be hands on, or for the test to have more activities or games. Some students also requested models or projects as their final test. There were also comments about having testing be optional.

The six students interviewed for the Sixth Grade Student Post Test Interview (Appendix Q) had similar comments about what types of science assessments they enjoyed this year. The participants tended to mention components of the units that were not actually the posttests. One stated they liked a portion of the physics unit, “...because we got to go outside,” in reference to a lab where students recorded the time it took their peers to walk, hop, walk backwards, and speed walk a short distance. Two students mentioned enjoying the Cell Analogy Model PT, but only one was in Sample B and actually graded on it. The other student, from Sample A, had done the PT for no score as a supplemental activity after taking the Cell Unit WSA. Her comment was, “I liked the cell project the most because you got to make a model – we got to choose to make a

model and I chose a cake because I like to bake and I liked that the most because I got to do what I like to do.” Students from both samples reported material from the Moon Phases, Eclipses, and Tides unit as “boring,” with one student in Sample A reporting, “I didn’t like the moon one this year because Mr. Davies does not give us HW and that was kind of like HW.” They were referencing the moon log, a daily observation packet all students completed at home before the beginning of Unit Two.

The interview participants continued to report common themes from the Sixth Grade Focused List Questionnaire and the ATSQ. They exhibited strong knowledge of what makes a traditional test a traditional test, by suggesting they are “by yourself, with testing folders,” are “tedious,” can be multiple choice, and that you generally write the answers down on a piece of paper. When asked to define a PT, one laughed as he said, “A task that you perform?” They agreed it meant having to do something and it possibly included working with other people. Sample B experienced the option to work with others in both of their PTs, due to the nature of the tasks, but Sample A did not. When asked how next years’ sixth graders should be assessed in science, there was some sense that things like the lab in the speed unit and the Cell Analogy Performance Task should stay the same, with the request for “more outside activities.” Another student said, “I like doing hands on projects.” Zero students mentioned repeating any of the WSAs given. In closing the interview, one student said, “Science doesn’t seem like it doesn’t help me as much as other subjects – it teaches me stuff but not like math. It doesn’t really stay in my head.” An opposite opinion was expressed by another student from the same classroom, who said, “I liked the science because most of the stuff – the hands on projects – are one

of my favorite things. Science isn't my best subject, but it's the most fun out of the subjects we do."

The three sixth grade teachers who participated in the study had many thoughts on giving the PTs and WSAs. They noted that the Cell Analogy PT went beyond science to include the language arts concept of an analogy, which they explicitly taught in a language arts lesson before beginning the PT. However, they observed that there were student difficulties with the Moon Phases, Eclipses, and Tides PT because they had received no prior instruction on cause/effect or mathematical reasoning, two skills that appeared integral to success on that PT. But, they enjoyed the amount of critical thinking the PTs required of students. Their observations about preparing students for the WSAs and PTs are in contrast to how students actually performed in Units One through Four. When asked how they felt about giving the WSAs, one teacher said, "It was a lot easier to prepare the kids because they're used to it and because there's one right answer. The practice journals they did were all written, so it's easy to go from science journal to WSA." Another teacher said regarding the PTs, "It seemed like it was out of my control to really prepare for the PT because the students had to use all of their knowledge to complete the task and it wasn't as easy to have students study for it." While teachers felt more in control of preparing students for the WSAs, students performed better on the PTs. As one teacher said, "Maybe straightforward thinking is detrimental to kids because it does not get them thinking – we need to get kids to create their own solutions, versus book work."

## INTERPRETATION AND CONCLUSION

The analysis of the results of this study finds that the effects of the Next Generation Science Standard (NGSS) Performance Tasks (PTs), as they were designed, are inconclusive in comparison to the Written Summative Assessments (WSAs). The raw gains students exhibited on the NGSS PTs reflect a small difference between the raw gains when students took the WSAs. It is important to note that the raw gain pre/post was usually more than the 1% I needed to perform chi square analysis for both types of tests. Also, if student satisfaction is to be valued higher than statistical significance, then the small difference in gain may matter more. This minor difference in raw gain also illustrates the research of Shymanksy et al. (1997), who found that performance tasks produced results very similar to mine: more questions than answers. In fact, Shymanksy et al. and one of my colleagues both reference prior student experience with PTs as a factor in student performance. "If you did a PT every three to six weeks...they would get used to it," one colleague theorized, while another suggested, "Doing more PTs will naturally make instructors better at giving them and the kids more comfortable in doing them."

While students were not gifted in describing their desired tests as PTs, or even at defining a PT when asked, they did appear to be comfortable with this testing format, given all of their suggestions for different types of hands on learning as their preferred unit outcomes. Perhaps as teachers grow in their comfort with performance tasks the raw gain pre/post for students who take a PT will reach statistical significance in comparison to growth on a WSA.

Also, the confusion that Sample A had in comparison to Sample B in correctly identifying their PTs as PTs might stem from what Accongio and Doran (1998) suggest about how carefully teachers must design PTs. Based on the nature of the units of study, Sample A was given two independent PTs, while Sample B was given two project based PTs that included working with a partner or group. Sample B's perception of a PT was more creative than Sample A's, who continued to see their PTs in a written, silent form when it came time to complete them. While it does not seem necessary for students to know that they are taking a PT if they can describe how they prefer to be tested in plain language, I do think it was an unfortunate choice on my part to create one type of experience for Sample A and another type of experience for Sample B. PTs do not have to be any one kind of task, they just need to be well designed and provide students with a rubric. Because this study also examined student perception, explaining to students that independent written performance tasks can be equally creative might be useful if the instructor wants to encourage a more open mode of thinking for students.

Accessing students' brain schemata appeared to be another factor in PT design that might explain the lackluster differences between WSA and PT performance. It also connects to the component of my first research question, since the PTs were based off of the Next Generation Science Standards. Specifically, the PTs I created were based on the NGSS modeling tasks, and according to the research I did on modeling, accessing students' prior embodied experiences might provide greater success for them when they are asked to build a model. Unit One's study of cells asked students to think about themselves as humans, and the PT offered them free reign of how they wanted to represent what they had learned. With so many students choosing "comfort" projects that

included favorite activities such as baking, their perception of their learning was more in line with what their focused list questionnaires said about how they liked to be tested: with friends, with food, and with options. The Unit Four PT, on the other hand, asked students to build a mathematical model to represent speed. Students' experience with considering a graph a model might be more limited due to their age, and it is not necessarily a method that accesses their direct living experience, like the option to bake a cake. In all, my own inconsistency in considering what types of models I was choosing for each PT might have affected the study results. While I do not think there is anything improper about having sixth graders model using a graph, more instruction as to why a graph is a useful type of model might increase student confidence and help them perceive a graph as truly a "hands on" method of solving in some scientific circumstances.

A last point to make in terms of variation in the performance of Sample A and Sample B is that this study represents four classrooms, each run in their own unique way. Differences are likely in some way attributable simply to the natural style one teacher has versus another. Since the results of this study are ultimately inconclusive statistically, it also supports the idea that students did not necessarily have incomparable experiences, either. Variation between class demographics and teaching style will always be one of the factors that makes education research variable by nature.

Another theme from this study is that sixth graders do not have a well-developed concept of failing, as the traditional grading system defines it. Perhaps twelve year olds are just optimistic, or perhaps this is a reflection of how counterintuitive the letter grade system can be for students. In terms of raw gain, I was very pleased as an instructor with how much pre/post growth was represented for all students across all four units and both

styles of test. But in terms of “failing” being a D or F and a score of less than 70%, there did not appear to be strong student understanding of this. Even students who acknowledged how poorly they had done still tended to rate themselves as earning a C, with some students using an entirely different logic of “I studied and tried hard so I deserve an A.” There is some merit to teaching students that one must “earn” an A, but students did not seem to have a good barometer for how hard one must work to hit such a high goal.

In addition, the results of the focused list questionnaire make me wonder about how students’ experience growing up in a culture of standardized testing has affected them. It was clear based on the wide array of testing genres and formats they offered that they are aware teachers are always watching them to see if they are learning, but they couldn’t determine exactly what counted towards being graded versus being observed. That burden falls on the instructor to make the expectations clear, especially for the students who simply wished not to be tested. I wonder, and am not able to determine, if this is because they are 12, and the reality of why teachers test has not set in for them, or if this implies that their years of No Child Left Behind standardized testing have darkened their minds towards testing. No Child Left Behind has been in place their entire lives, so comparing their dislike for testing to a sample of sixth graders from 15 to 20 years ago would be one of the only ways to answer this question.

While some students’ suggestions for testing on the focused list questionnaire and on their after test questionnaires were not realistic, a huge percentage of their ideas reflect what teachers know to be true about the effects of student choice in the classroom. When students have a voice in how they demonstrate their learning, their minds and hearts are

engaged, and they are happier to be in school. Despite a lack of statistical significance, the moderate gains in PT performance over WSA performance might reflect that PTs are worth pursuing in the name of student satisfaction. Happy students, happy teachers.

#### VALUE

The study of Next Generation Science Standard (NGSS) performance task (PT) implementation in the middle school science class was extremely valuable, not only for my small cohort of science teachers, but for the greater experience with the new culture of the NGSS. Just this year, the NGSS have released some of the sample classroom assessment tasks, for which I am so thankful someone else has taken the time to design. My study has taught me multitudes about excellent PT design, with one of the most important learnings being that a quality task can take hours to find and alter or to write from scratch. Also significant is ensuring all instructors grade students' finished products the same way. NGSS is gearing up to provide rubrics for grading their tasks, which will add some clarity to the hard work I had in determining what elements of a task should "count" towards the final score. Consistency of rigor is already hard to achieve between classes, but having a federal plan will hopefully unify students' experience state-to-state should they move.

In terms of best practice, I now, more than ever, am going to continue pursuing PTs as one of the primary ways I assess science students. The thought processes I was privy to while watching students work, and the excellent collaborative questions raised by my colleagues when discussing how to implement the task, provided a wealth of information about how my students think. It also made collaboration periods with my colleagues much more rich discussions of the heart of what we needed students to learn.

In addition, the grading of a PT was far less labor intensive at the end of the day because, in general, a rubric was built in. In fact, for my colleagues who represented Sample B, and who had students present their PTs in class, the feedback loop with students was nearly immediate. They could grade students while they watched them present.

I am merely joining a rising tide, as researching the effects of PTs is a very wide movement amongst science educators at the moment. But, I am excited that my research gives me more reasons to enjoy fully implementing the NGSS next year. On a national level, if colleagues everywhere can be having rich conversations about how their students are demonstrating their learning, collaborations both vertical and horizontal will be more enriched.

The question this study raises in my mind is how to appropriately determine what level of critical thinking a student of a particular age level should reach. It seems like teacher opinion in this matter varies greatly. Teaching critical thinking is difficult, but appears to be one of the larger themes of shifting to performance based grading. While I have found the NGSS to be helpful education policy, the overall federal policy that students must always be ready to demonstrate their understanding in a testing format seems counterintuitive. Does this place the burden of advocating for classroom based assessments or other alternate methods of “proving” science students have learned the necessary material at each level back on science instructors? If so, that is a movement I will readily join.

The second question this study raises is the value of listening to student preferences in testing format and conditions. Many schools, my own included, have at points arbitrary rules about what students can and cannot do. While structure and order

teaches children about adult life, is there value in considering their requests to have food, or to choose their work partners? So often teachers, schools, and districts obsess over giving children realistic experiences, without first considering whose reality they are trying to create. If the reality of student concerns is answered, especially in regards to anxiety provoking situations like testing, perhaps more than the marginal raw gain improvements I saw in students' PT performance could be had on a grander scale. If the aim of a business is to create happy, fulfilled customers, there may be something to be said for the idea of reasonable "customer service" in middle school education.

As a result of this study, I have deepened my resolve to listen to student voice in the classroom. I already enjoyed giving pre and posttest surveys, but the students in my study showed me they had some useful instincts on how they prefer to be tested, even if others were a little fanciful. If nothing else, I noticed in my study that asking students for their thoughts and opinions made them more open minded to whatever test I was giving because they could be honest with me. This study has also changed my view of PT design. I will eagerly look to the NGSS pre-designed rubrics when they arrive on the NGSS website – now knowing for myself how difficult accurate, quality PT design really is, I have never felt more appreciative of such a resource coming into existence and will happily try all of them out. This study also changed my view of assessment as a teacher. As the research warned me at the outset, there is *no* "magic bullet" in assessment design, only thoughtful consideration of what students are expected to know and how accurate an instrument the assessment, be it written or performed, really is in allowing students to demonstrate competence.

REFERENCES CITED

- Accongio, J.L., & Doran, R.L. (1993). *Classroom assessment: Key to reform in secondary science education*. Columbus, OH: ERIC Clearinghouse for Science, Mathematics, and Environmental Education.
- Aubusson, P., & Fogwill, S. (2006). Role play as analogical modeling in science. In Aubusson, P, Harrison, A, & Ritchie, S. (Eds.). *Metaphor and analogy in science education* (pp. 93-104). Dordrecht, Netherlands: Springer.
- Battle Ground Public Schools (n.d.), *About Us Overview*. Retrieved from <http://battlegroundps.org/node/370>
- Bodner, G. (1986). Constructivism: A theory of knowledge. *Journal of Chemical Education*, 63(10), 873-874.
- City of Battleground, (n.d.) *Population*. Retrieved from <http://www.cityofbg.org/index.aspx?nid=89>
- Gardner, H. (1983) *Frames of mind: the theory of multiple intelligences*. New York: Basic Books
- Gilbert, S.W. (2011). *Models based science teaching*. Arlington, VA: NSTA Press.
- Hibbard, K.M., Wagenen, L., Lewbel, S., Waterbury-Wyatt, S., Shaw, S., Pelletier, K...Wislocki, J.A.(1996) *Teacher's Guide to Performance-Based Learning and Assessment*. Virginia: Association for Supervision & Curriculum Development
- National Research Council. (2012). Scientific and engineering practices. *A framework for k-12 science education: Practices, crosscutting concepts, and core ideas* (pp. 41- 59). Washington, D.C: The National Academies Press.
- NGSS Lead States. (2013). *Next generation science standards: For States, by States*. Washington, D.C: The National Academies Press.
- Niebert, K., Marsch, S., Treagust, D. (2012). Understanding needs embodiment: A theory-guided reanalysis of the role of metaphors and analogies in understanding science. *Science Education*, 96(5), 849-877.
- Risinger, T., (2009). *Earth Sun Moon Phenomena: Seasons, Lunar Cycle, Tides*. Retrieved from [http://digitalcommons.trinity.edu/educ\\_understandings/115](http://digitalcommons.trinity.edu/educ_understandings/115)
- Rohrer, T. (2005). Image schemata in the brain. In B. Hampe & J. Grady (Eds.), *From perception to meaning: Images schemas in cognitive linguistics* (pp. 165–196). Berlin: Mouton de Gruyter.

Shymansky, J.A., Enger, S., Chidsey, J.L., Yore, L.D., Jorgensen, M., Henrique, S.L., & Wolfe, E.W. (1997). Performance Assessment in Science as a Tool to Enhance the Picture of Student Learning. *School Science and Mathematics*, 97:172-183. doi:10.1111/j.1949-8594.1997.tb17364.x

Simpson, S.M., Licht, B.G., Wagner, R.K., & Stader, S.R. (1996). Organization of children's academic ability-related self-perceptions. *Journal of Educational Psychology*, 88 (3), 387-396. doi:http://dx.doi.org/10.1037/0022-0663.88.3.387

Washington State Report Card. (2014). Retrieved from <http://reportcard.ospi.k12.wa.us/SideBySide.aspx?schoolId=>

APPENDICES

APPENDIX A  
IRB EXEMPTION



**INSTITUTIONAL REVIEW BOARD**  
**For the Protection of Human Subjects**  
**FWA 00000165**

960 Technology Blvd. Room 127  
 c/o Immunology & Infectious Diseases  
 Montana State University  
 Bozeman, MT 59718  
 Telephone: 406-994-6783  
 FAX: 406-994-4303  
 E-mail: cherylj@montana.edu

*Chair:* Mark Quinn  
 406-994-5721  
 mquinn@montana.edu  
*Administrator:*  
 Cheryl Johnson  
 406-994-6783  
 cherylj@montana.edu

**MEMORANDUM**

**TO:** Carli Barnes and John Graves  
**FROM:** Mark Quinn, Chair *Mark Quinn CJ'*  
**DATE:** August 29, 2014  
**RE:** "Comparing Student Performance and Perception of Competency on Summative Science Performance Tasks Versus Written Science Summative Tests at the Sixth Grade Level" [CB082914-EX]

The above research, described in your submission of August 29, 2014, is exempt from the requirement of review by the Institutional Review Board in accordance with the Code of Federal regulations, Part 46, section 101. The specific paragraph which applies to your research is:

- (b) (1) Research conducted in established or commonly accepted educational settings, involving normal educational practices such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.
- (b) (2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the subjects' financial standing, employability, or reputation.
- (b) (3) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior that is not exempt under paragraph (b)(2) of this section, if: (i) the human subjects are elected or appointed public officials or candidates for public office; or (ii) federal statute(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter.
- (b) (4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available, or if the information is recorded by the investigator in such a manner that the subjects cannot be identified, directly or through identifiers linked to the subjects.
- (b) (5) Research and demonstration projects, which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine: (i) public benefit or service programs; (ii) procedures for obtaining benefits or services under those programs; (iii) possible changes in or alternatives to those programs or procedures; or (iv) possible changes in methods or levels of payment for benefits or services under those programs.
- (b) (6) Taste and food quality evaluation and consumer acceptance studies, (i) if wholesome foods without additives are consumed, or (ii) if a food is consumed that contains a food ingredient at or below the level and for a use found to be safe, or agricultural chemical or environmental contaminant at or below the level found to be safe, by the FDA, or approved by the EPA, or the Food Safety and Inspection Service of the USDA.

Although review by the Institutional Review Board is not required for the above research, the Committee will be glad to review it. If you wish a review and committee approval, please submit 3 copies of the usual application form and it will be processed by expedited review.

APPENDIX B

SIXTH GRADE STUDENT FOCUSED LIST QUESTIONNAIRE

### Sixth Grade Student Focused List Questionnaire

#### Tests

Your teacher will set a timer for three minutes. When the timer begins, please use the space below to list as many ways as you can remember that you have been tested in school, sports, or in other settings over the years. You can list names of tests and/or ways you have been tested. Remember, tests are given in many ways, so include anything you think might have been a test, even if you are not 100% sure. No wrong answers.

Use the numbered boxes below:

1.)	11.)
2.)	12.)
3.)	13.)
4.)	14.)
5.)	15.)
6.)	16.)
7.)	17.)
8.)	18.)
9.)	19.)
10.)	20.)

You can add extra ideas here:

*When time is up, please **CIRCLE** all the ways you ENJOY or PREFER to be tested. (Think about what ways make you feel the best about yourself as a student.)*

In a perfect school, how would you like teachers to test you? Remember, a test is just supposed to show teachers what you have learned.

(DO NOT say, "No tests." You know better )

Please list your ideas in the boxes below:


Participation in this research is voluntary and participation or non-participation will not affect your grades, class, or collegial standing in any way.

APPENDIX C  
SIXTH GRADE TEACHER SURVEY

**Sixth Grade Teacher Survey**

- 1.) What types of assessments do you give in your classroom?
  
  
  
  
  
  
  
  
  
  
- 2.) What type or types of assessments do you believe help students best show you what they have learned?
  
  
  
  
  
  
  
  
  
  
- 3.) How comfortable are you designing or giving students written summative assessments? Please select the one best answer.

**Very uncomfortable      Uncomfortable      Unsure      Comfortable      Very Comfortable**

- 4.) How comfortable are you designing or giving students performance assessments? Please select the one best answer.

**Very uncomfortable      Uncomfortable      Unsure      Comfortable      Very Comfortable**

APPENDIX D  
CELL UNIT PRE-TEST

Name: \_\_\_\_\_ Date: \_\_\_\_\_ Class: \_\_\_\_\_

### Cell Unit Pre-Test

#### Part I:

1. In your own words, use a complete sentence to define what a "cell" is: \_\_\_\_\_
2. In your own words, use a complete sentence to define what an organism is: \_\_\_\_\_
3. In your own words, use a complete sentence to state how a cell and an organism are related: \_\_\_\_\_  
\_\_\_\_\_
4. What is the key difference between a plant and an animal cell? \_\_\_\_\_

#### Part II:

Word Bank							
cytoplasm	endoplasmic reticulum	nucleus	cell membrane	mitochondria	chloroplasts	vacuole	organelles

1. Use the Word Bank to fill in the blank to complete each statement.  
(Hint: There are more words than there are blanks.)
2. Draw a Line from each completed statement to the number on the figure of a simplified animal cell.

In a cell without cell walls, the

\_\_\_\_\_ forms the outside boundary that separates the cell from its environment.

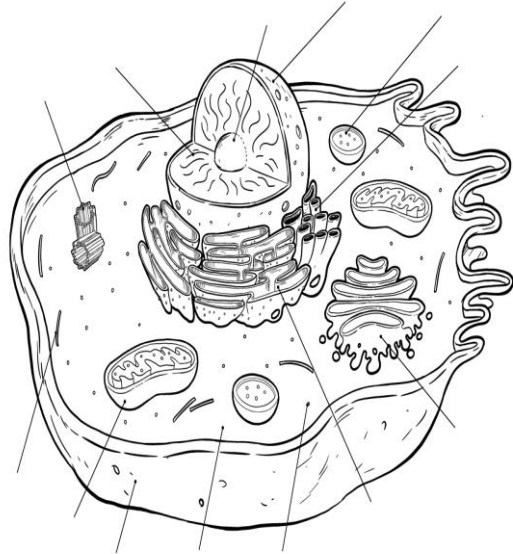
The

\_\_\_\_\_ is a large, oval structure that directs all of the cell's activities.

\_\_\_\_\_ produce most of the energy the cell needs to carry out its functions.

A maze of passageways called the

\_\_\_\_\_ carries proteins and other materials from one part of the cell to another.



APPENDIX E

CELL ANALOGY MODEL PERFORMANCE TASK

Student Name: \_\_\_\_\_ Homeroom Teacher: \_\_\_\_\_

*\*Please make sure to use 1 assessment per student so grade is individual\**

### Cell Analogy Model Performance Task

**Task Objectives:**

- a.) Each student can offer reasonable, clear scientific reasons their cell is living (as opposed to non-living).
- b.) Each student can say precisely why their cell is either a plant or animal cell using the vocabulary of cell organelles.
- c.) The analogies used in the cell model correctly reflect what the organelles of a plant or animal cell do.

<b>PART I</b>		
<i>*For this portion of the assessment, please have a brief, private conversation with the student*</i>		
	<b>1 pt.</b>	<b>0 pts.</b>
Student offers reasonable scientific explanation for why a cell is a living thing.	YES	NO
Student clearly explains which type of cell they elected to model, using key organelle vocabulary.	YES	NO
<b>POINT TOTAL:</b> _____ +/ 2+		

<b>PART II</b>			
<small>*indicates a part specific to plant cells only</small>			
ORGANELLE	PRESENT = 1 PT.	REASONABLE ANALOGY = 1 PT.	-1 PT. (for organelles that do not belong in cell student intended to model)
*Cell Wall			
Cell Membrane			
Cytoplasm			
Nucleus			
<b>Bonus:</b> Nuclear Envelope or Nucleolis			
Endoplasmic Reticulum			
Ribosomes (many)			
Golgi Body (at least one)			
Mitochondria (at least one)			
Lysosome (at least one)			
Vacuole (at least one w/ specific purpose)			

<b>*Chloroplast (at least one)</b>			
<b>Bonus:</b> any additional cell parts (1 pt. max for all)			
<b>POINT TOTALS:</b>			

**TASK POINT TOTAL:** \_\_\_\_\_ +/13+ = \_\_\_\_\_ %

APPENDIX F

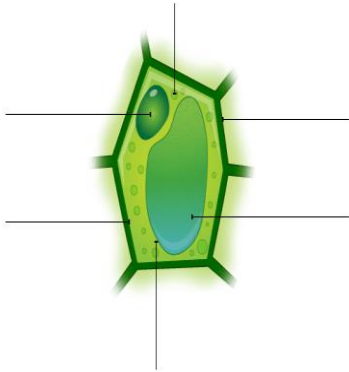
CELL UNIT WRITTEN SUMMATIVE ASSESSMENT

Name: \_\_\_\_\_ #: \_\_\_\_\_ Date: \_\_\_\_\_ Homeroom Teacher: \_\_\_\_\_

## Cell Unit Final Test

### Interpreting Diagrams

Cell A



Cell B



1. Would *photosynthesis* occur in Cell A, Cell B, or both? Please circle your answer.
  - a. Cell A
  - b. Cell B
  - c. both Cell A and Cell B

Why did you select the letter you selected? (Provide CLEAR evidence -- not just "Because it is a \_\_\_ cell" or "Because I think so")

---



---

2. Which cell is an animal cell? Please circle your answer.
    - a. Cell A
    - b. Cell B
    - c. both Cell A and Cell B
- Why did you select the letter you selected? (Provide CLEAR evidence)

---



---

### Multiple Choice & Short Answer

<p>3.) In the box below, name something that is <b>*living*</b> and state <b>WHY</b> it is living. Provide clear evidence.</p>	<p>3.) In the box below, name something that is <b>NOT living</b> and state <b>WHY</b> it is not living. Provide clear evidence.</p>
<p>Living thing: _____</p>	<p>Non-living thing: _____</p>

Why is it living? _____ _____ _____	Why is it non-living? _____ _____ _____
---	---

4. "All living things are made of cells, and all things that are not living are not made of cells." Is this statement TRUE or FALSE?

\_\_\_\_\_

5. ATOM → MOLECULE → \_\_\_\_\_ → CELL → \_\_\_\_\_ → ORGANS → \_\_\_\_\_ → \*COMPLEX MULTICELLULAR ORGANISMS\*

6. One difference between plant and animal cells is \_\_\_\_\_. Another important difference is \_\_\_\_\_.

\_\_\_\_\_

7. "I am a pocket rocket that releases all the energy for your body. I am one of Cell's internal machines. I make a food to fuel Cell called ATP. I am a tiny power plant working inside every cell. What am I?" *Please circle ONE answer.*

- a. mitochondria
- b. ribosome
- c. endoplasmic reticulum

8. "I am Cell's delicate workshop and the only guy that can read his ancient genetic code. I am the craftworker that builds the proteins your cells need. There are a ton of me, zipping together nutrients called amino acids to make you muscles, hair, or other useful proteins. What am I?" *Please circle ONE answer.*

- a. Golgi body
- b. lysosome
- c. ribosome

**SCIENTIFIC MODELING**

In the boxes below, please sketch a plant and animal cell and all the organelles they contain. Use the word bank to label each correctly. Words in the word bank can be used more than once if needed. Then, include a few words to describe what each part does for the cell. \*1 bonus point per cell if you color it realistically\* \*SPELLING counts!!\*

<b>WORD BANK</b>					
mitochondria	endoplasmic reticulum	ribosome			
Golgi body	cytoplasm				
chloroplast	vacuole	lysosome	nucleus	cell membrane	cell wall

<b>PLANT CELL</b>	<b>ANIMAL CELL</b>

<p>MLA citation for questions 7 &amp; 8: Green, Dan. "Building Blocks." <i>Biology: Life as We Know It!</i> New York: Henry Holt and LLC, 2008. 10, 16. Print.</p>	

APPENDIX G

EARTH'S SEASONS UNIT PRE-TEST

Name: \_\_\_\_\_ #: \_\_\_\_\_ Homeroom Teacher: \_\_\_\_\_ Score: \_\_\_\_\_/7+

## THE REASON FOR EARTH'S SEASONS

### *Unit Pre-Test*

#### Part I

1. Circle the letter of each sentence that is TRUE about seasons on Earth.
  - a. In summer, Earth tilts towards the sun.
  - b. Earth's axis does not slant.
  - c. In winter, Earth's axis tilts away from the sun.
2. Which season are the days longer in? \_\_\_\_\_
3. Which season are the days shorter in? \_\_\_\_\_
4. "In summer, the day is longer than the night." TRUE FALSE
5. "In winter, the Northern Hemisphere tilts away from the sun." TRUE FALSE

#### Part II

Ethan and Ian are pondering the meaning of the universe. They pause to think about why Battle Ground has four seasons. Ethan decides to ask his friend Tristan. Tristan says, "Well, I've heard that we experience seasons because of the Earth's changing distance from the Sun. I've always thought the Earth is closer to the Sun in the summer, and farther in the winter." Ethan returns to Ian to report his findings.

Ian, being an excellent scientist, decides to do a little fact checking of his own. At recess, he asks Ella, "Why do the seasons occur?" Ella says, "My dad told me it's because the Earth is tilted about 23.5 degrees, so it's never straight up and down compared to the Sun. When it's summer in Battle Ground, the North Pole and the rest of the Northern Hemisphere are pointing toward the Sun. During the summer, the Sun is high in the sky. The Sun's rays hit us more directly during summer, making the sunlight more intense and warmer. In winter, the Sun is low in the sky, and the light rays are spread more thinly, and they do not warm us as much."

"Wow," Ian says, and he returns to report to Ethan. Ian and Ethan, BOTH being fabulous scientists, decide to let their findings stand up to one last report. They approach Laurel in P.E. and ask her why the seasons occur. "Well, my older brother told me that because of the tilt of the Earth, it's toasty warm summer in Battle Ground when the Earth is tilted directly towards the sun, and it's frosty cold winter when the Earth is tilted away." Suddenly, Tirna interjects to say, "My sister Amie also said we don't have a perfectly circular orbit around the Sun, so that helps create the seasons." Ethan and Ian thank Laurel and Tirna for their knowledge. They look at each other in terror, realizing their minds are about to explode with confusion. They've heard four different explanations of the seasons....which one is correct?

Who do you agree with? Circle the letter next to the person whose explanation you most agree with.

(Make sure to select just one.)

- A.) Tristan: "... I've always thought the Earth is closer to the Sun in the summer, and farther in the winter."
- B.) Ella's dad: "...the Earth [is] never straight up and down compared to the Sun...The Sun's rays hit us more directly [in] summer [and]...In winter, the light rays are spread more thinly..."

C.) Laurel's brother: *"...when the Earth is tilted directly towards the sun, it's...summer...and [in] winter...the Earth is tilted away...."*

D.) Tirna's sister: *"We don't have a perfectly circular orbit around the Sun."*

APPENDIX H

EARTH'S SEASONS PERFORMANCE TASK

Name: \_\_\_\_\_ # \_\_\_\_\_ Date: \_\_\_\_\_ Homeroom Teacher: \_\_\_\_\_

### The Four Seasons: A Scientific “Calendar” of the Northern Hemisphere

Battle Ground is in the Northern Hemisphere, so we experience summer in June, July, and August, and winter in December, January, and February. Use this knowledge and what you have learned in science the past two weeks to help you draw a small model for what the Earth and Sun are doing in space during each month of the year to create our sensation of these seasons.

In each box, please draw a diagram showing the relative location of Sun and Earth and how Earth would be tilted in relation to the Sun during that month. Add any labels or written descriptions of your picture as needed. Please also add logical colors.

In addition, please **label** the 2 months that have *solstices* and the 2 months that have *equinoxes*. Simply write “solstice” or “equinox” in the box of the proper month. Try and label *which* solstice or equinox, specifically.

Feel free to ask your teacher for a globe or ball to help you better visualize the lessons from class and assist you in correctly drawing.

*This calendar begins with winter.*

<b>WINTER</b>			
	<b>DECEMBER</b>	<b>JANUARY</b>	<b>FEBRUARY</b>
<b>SPRING</b>			

	<b>MARCH</b>	<b>APRIL</b>	<b>MAY</b>
<b>SUMMER</b>			
	<b>JUNE</b>	<b>JULY</b>	<b>AUGUST</b>
<b>FALL (AUTUMN)</b>			
	<b>SEPTEMBER</b>	<b>OCTOBER</b>	<b>NOVEMBER</b>

APPENDIX I

EARTH'S SEASONS WRITTEN SUMMATIVE ASSESSMENT

Name: \_\_\_\_\_ #: \_\_\_\_\_ Homeroom Teacher: \_\_\_\_\_ Score: \_\_\_\_\_/

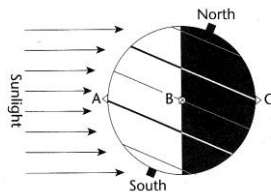
**THE REASON FOR EARTH'S SEASONS*****Final Test*****Part I: Vocabulary**

Match each term with its definition by writing the letter of the correct definition in the right column on the line beside the term in the left column. (There are more definitions than there are terms.)

TERM	DEFINITION
____ 1.) axis	a. The path of earth as it revolves around the sun.
____ 2.) rotation	b. Line passing through Earth's center and poles.
____ 3.) revolution	c. The sun is farthest north or south of the equator at this time.
____ 4.) orbit	d. Movement of Earth around sun.
____ 5.) equinox	e. Movement of Earth on its axis
____ 6.) solstice	f. The noon sun is directly overhead at the equator at this time.
____ 7.) hemisphere	g. System of organizing time that defines the beginning, length, and divisions of a year.
____ 8.) equator	h. The study of the moon, stars, and other objects in space.
	i. a half of the earth, usually as divided into northern and southern halves by the equator, or into western and eastern halves by an imaginary line passing through the poles.
	j. an imaginary line drawn around the earth equally distant from both poles, dividing the earth into northern and southern hemispheres.

**Part II: Interpreting a Diagram**

Use the following figure to answer question 9.



9.) In the diagram, what season is it in North America? \_\_\_\_\_

**Part III: Conquering Misconceptions**

Below is a list of possible supporting claims that might explain the following statement:

“In the United States, it’s hotter outside in June than in December.”

But...WHY is this true? Multiple explanations are listed below. They are either scientifically true (but not known by many adults), OR are common UNTRUE beliefs (held by many adults!) Please indicate whether each statement is either true or false, and then state some evidence to support your thinking.

TRUE	FALSE	The U.S. is hotter in June because the Sun itself gives off more heat & light energy in June (and less in December). <i>I think this claim is _____ because</i> _____ _____
------	-------	--

TRUE	FALSE	The U.S. is hotter in June because the Earth is closer to the Sun in June, and farther away from the Sun in December. <i>I think this claim is _____ because</i> _____ _____
------	-------	---

TRUE	FALSE	The U.S. is hotter in June because the U.S. is closer to the Sun in June, and farther away from the Sun in December. <i>I think this claim is _____ because</i> _____ _____
------	-------	--

TRUE	FALSE	The U.S. is hotter in June because the U.S. is facing more toward the Sun in June and away from the Sun in December. <i>I think this claim is _____ because</i> _____ _____
------	-------	--

TRUE	FALSE	The U.S. is hotter in June because the Sun gets higher in the sky in June, so its rays are more concentrated on the ground. <i>I think this claim is _____ because</i> _____ _____
------	-------	---

TRUE	FALSE	The U.S. is hotter in June because in the U.S., there are more hours of daylight in June than in December. <i>I think this claim is _____ because</i> _____ _____
------	-------	--

**Part IV: The Seasons Worldwide & Special Phenomenon**

Use the front vocabulary to help you.

10.) Each of the two days of the year when the noon sun is farthest north or south of the equator is called a(n)

\_\_\_\_\_.

11.) Each of the two days of the year when neither hemisphere is tilted toward or away from the sun is called a(n)

\_\_\_\_\_.

12.) Circle the letters of the statements that are true about Earth's seasons in the Northern Hemisphere. Use a globe to help you if you need one.

a. When the Northern Hemisphere has summer, the Southern Hemisphere is tilted away from the sun.

b. In December, the shortest daytime on Earth is found in the Southern Hemisphere.

c. An *equinox* occurs on the same days at the same time in both hemispheres.

13.) \_\_\_\_\_ happens in June and marks the longest daytime of the year for the Northern Hemisphere.

14.) \_\_\_\_\_ happens in December and marks the shortest daytime of the year for the Northern Hemisphere.

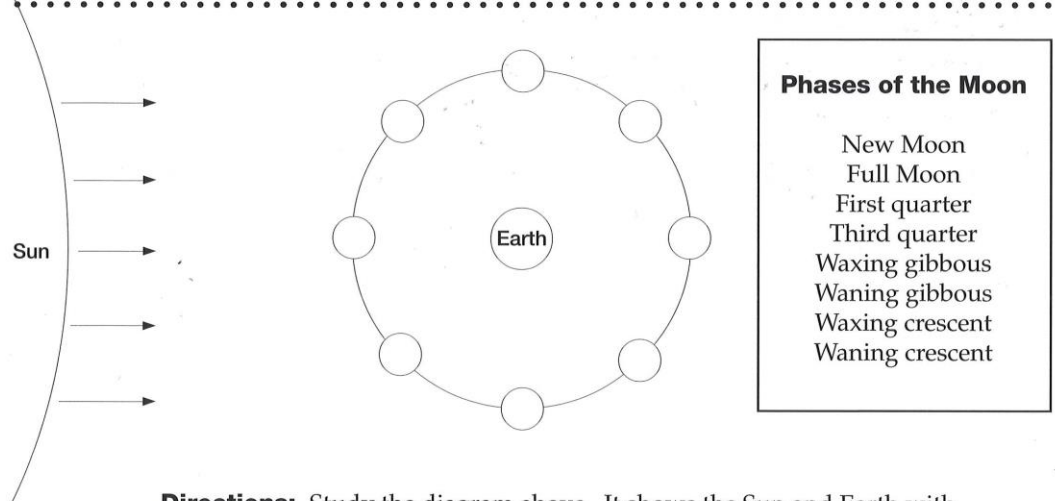
APPENDIX J

MOON PHASES, ECLIPSES, AND TIDES PRE-TEST

Name \_\_\_\_\_

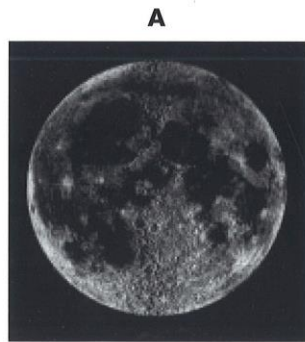
Date \_\_\_\_\_

**MID-SUMMATIVE EXAM 9**

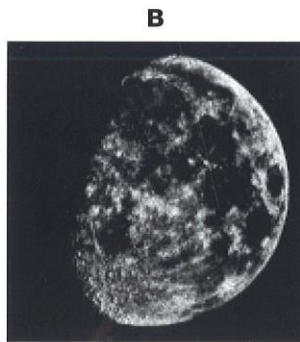


**Directions:** Study the diagram above. It shows the Sun and Earth with several possible positions for the Moon as it orbits Earth. You are looking down on the system from above Earth's North Pole.

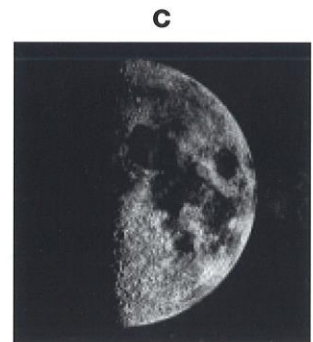
1. Draw an arrow on the diagram to show which way the Moon is revolving around Earth.
2. Look at the pictures below that show different phases of the Moon. Below each picture, write the name of the phase it shows.



Phase name \_\_\_\_\_



Phase name \_\_\_\_\_



Phase name \_\_\_\_\_

3. Label the diagram at the top of the page with the letters A, B, and C to show where the Moon would be in its orbit when it is showing the phase shown in photos A, B, and C.

**4. While sitting on the beach on a lazy summer day you accidentally fall asleep. When you awaken you remember that your mom wanted you home at 6:00 p.m. You have no time-piece but you see a waning crescent moon setting in the western sky. Are you late for your curfew? Why or why not?**

**5. When do high tides occur? Why?**

**6. When do low tides occur? Why?**

**7. You suddenly notice the moon turning red. Can you name what is happening?**

APPENDIX K

MOON PHASES, ECLIPSES, AND TIDES PERFORMANCE TASK

**Performance Assessment  
Scenario Cards, Presentation Outline, & Rubric**

**Scenario Overviews**

- I. Earth has been struck by a small Asteroid! After the collision there appears to be a change in its degree of tilt and rate of rotation. Please explain what life on earth will likely be like.
- II. Due to drastic "climate change" the South Pole is melting quickly. Earth is now unevenly weighted and appears to be slowly tipping over. Earth's tilt is slowly increasing and things are not the same. Please predict what life on earth will be like.
- III. The moon has been struck by an asteroid! We now have two moons, an old large moon, and a new baby one! Re-construct the Lunar Cycle and any changes that may be experience on earth.
- IV. The Sun as suddenly increased in size! The earth is now revolving around the sun much quicker and rotating much slower. We need a new calendar and clock.

For scenarios I & II students will also need to be assigned a city or country so their predictions can be tailored to that location (latitude is the major factor, further from the equator will experience more drastic changes)

**Questions that need answers in the presentation**

How will the length of a year change?

How will the length of a day change?

How will the length of 'length of day' and 'length of a night' change?

How will the seasons change?

What will our new climate be like?

What will tides be like?

How will these changes effect: food sources, energy use, travel, timing, our calendar, international trade?

**Requirements for Presentation**

Students must address the above questions.

Students must construct a visual illustration or model to explain their findings.

Students must present relevant information and be able to address their governments questions

**Requirements for Students acting as Government Officials.**

Students will evaluate the presentation for accuracy

Each official must ask a challenge question

## **Earth has been struck by a small Asteroid!**

Don't worry you survived. The asteroid was small but still it has caused significant changes to Earth!

### **WHAT YOU KNOW:**

You have discovered that Earth's tilt of 23.5 degrees is gone. Earth appears to be well no longer tilted on its axis.

Also, you believe that earth is not rotation as fast.

You know that this means significant changes to our seasonal cycle, days and nights, climate, and other aspects of life on Earth.

### **WHAT YOU NEED TO ACCOMPLISH:**

Your government is asking your team of Scientist to present your findings and help them come up with solutions to problems that will arise. You have 48 hours before you must present your findings to your government.

### **ADDRESS THE FOLLOWING QUESTIONS**

- How will the length of a year change?
- How will the length of a day change?
- How will the length of 'length of day' and 'length of a night' change?
- How will the seasons change?
- What will our new climate be like?
- What will tides be like?
- How will these changes effect: food sources, energy use, travel, timing, our calendar, international trade? Identify two major problems that will likely be experienced and possible solutions.

# The South Pole is Melting and Earth is Tipping Over!

## WHAT YOU KNOW:

Ocean levels have been on the slow and steady rise but due to drastic and rapid "climate change" the South Pole is melting even faster now, super fast, like an ice cube on hot pavement. You believe that because of this the earth is going to be unevenly weighted and it will start to tip over. You have discovered that Earth's tilt of 23.5 degrees is changing. In fact it is increasing and quickly. Currently it is at 28.5 degrees and changing.

You know that this means significant changes to our seasonal cycle, days and nights, climate, and other aspects of life on Earth.

## WHAT YOU NEED TO ACCOMPLISH:

Your government is asking your team of Scientist to present your findings and help them come up with solutions to problems that will arise. You have 48 hours before you must present your findings to your government.

## ADDRESS THE FOLLOWING QUESTIONS

- How will the length of a year change?
- How will the length of a day change?
- How will the length of 'length of day' and 'length of a night' change?
- How will the seasons change?
- What will our new climate be like?
- What will tides be like?
- How will these changes effect: food sources, energy use, travel, timing, our calendar, international trade? Identify two major problems that will likely be experienced and possible solutions.

## **Did you see the moon explode last night? NO... What, we have two Moons?**

The Earth's Moon has been struck by an asteroid. You did not see the collision because it was during the full moon and it happened during your day time. Too bad!

### **WHAT YOU KNOW:**

Our moon is not the same. In fact we have two Moons. The original moon is now smaller and missing a chunk off its side. Our second moon appears to be the missing chunk off the original moon. The new moon appears to be revolving around the old moon. You know that this means our old lunar cycle is obsolete. You have been working very hard to determine a new Lunar Cycle and what effects this new cycle will have on Earth.

### **WHAT YOU NEED TO ACCOMPLISH:**

You will present your new Lunar Cycle to your government. You have 48 hours before you must present your predictions to your government.

### **ADDRESS THE FOLLOWING QUESTIONS**

- How are the two objects now moving in space?
- How many Moon Phases will there be?
- What effect will two moons have on tides?
- Will this affect other aspects of life on earth?

## **The Sun is Gigantic! ...and Summer Was Really Cold This Year**

Over the past year the sun has been increasing in size. It looks really cool! However summer break was strange, it was really cold.

### **WHAT YOU KNOW:**

The sun has been increasing in size over the past year. But it has not gotten any hotter. You are certain that the Earth is revolving around the sun much faster... twice as fast in fact! You are also observing that days are really long and the sun is rising and setting strange times, you are pretty sure it is rotating 25% slower, meaning days are longer.

You know that this means significant changes to our seasonal cycle, days and nights, climate, and other aspects of life on Earth.

### **WHAT YOU NEED TO ACCOMPLISH:**

Your government is asking your team of Scientist to present your findings and develop a new calendar and clock to account for the changes. You have 48 hours before you must present your findings to your government.

### **ADDRESS THE FOLLOWING QUESTIONS**

- How will the length of a year change?
- How will the length of a day change?
- How will the length of 'length of day' and 'length of a night' change?
- How will the seasons change?
- What will our new climate be like?
- What will tides be like?
- Construct an new calendar that will match the new length in "year"
- Come up with a new clock that will work with our "new length of day"

NAMES: \_\_\_\_\_ DATE: \_\_\_\_ PERIOD: \_\_\_\_  
 GROUP YOU ARE EVALUATING \_\_\_\_\_

SECTION 1	Vague, Incorrect (0)	Unclear, Somewhat Accurate (1)	Correct & Accurate (2)	Very Clear, Accurate & Detailed (3)
<b>Presentation Content &amp; Quality</b>				
<ul style="list-style-type: none"> <li>• Changes in Year length</li> <li>• Changes in "Day"</li> <li>• Length of Day</li> <li>• Length of Night</li> <li>• Changes in Seasons</li> <li>• What will tides be like</li> </ul>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<b>Visual/Graphic /Model</b>				
<ul style="list-style-type: none"> <li>• Show what has happened</li> <li>• Show why things are happening</li> <li>• Corresponds to what was verbally explained</li> </ul>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<b>First Problem and Solutions</b>				
<ul style="list-style-type: none"> <li>• Is this a possible problem we may face</li> <li>• Is the problem clearly identified</li> <li>• Is the solution to the problem appropriate.</li> </ul>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
<b>Second Problem and Solutions</b>				
<ul style="list-style-type: none"> <li>• Is this a possible problem we may face</li> <li>• Is the problem clearly identified</li> <li>• Is the solution to the problem appropriate.</li> </ul>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>	<input type="checkbox"/> <input type="checkbox"/> <input type="checkbox"/>
Total Pints Earned: _____ (48 Possible)	0 Points	____ Points	____ Points	____ Points
Section 2	Points Possible	Student Evaluation	Teacher Evaluation	
<b>Presentation:</b> Clarity and Organization	12			
<b>Creativity:</b> How Original are your ideas and Solution	12			
<b>Effort:</b> Seen in Preparation and Presentation	12			
<b>Cooperation/Behavior:</b> working as a group, attention	12			
<b>Extra:</b> Above and beyond necessary requirements	4			
Total Points Possible Earned:	52			
	SECTION 1	SECTION 2	TOTAL (100)	
Final Student Evaluation				

This performance task comes from:

Risinger, T., (2009). Earth Sun Moon Phenomena: Seasons, Lunar Cycle, Tides.  
 Retrieved from [http://digitalcommons.trinity.edu/educ\\_understandings/115](http://digitalcommons.trinity.edu/educ_understandings/115)

APPENDIX L

MOON PHASES, ECLIPSES, AND TIDES WRITTEN SUMMATIVE ASSESSMENT

Name: \_\_\_\_\_ #: \_\_\_\_\_ Date: \_\_\_\_\_ Homeroom Teacher: \_\_\_\_\_

## Moon Phases, Eclipses, & Tides Unit Test

### Part I: Moon Phase Identification

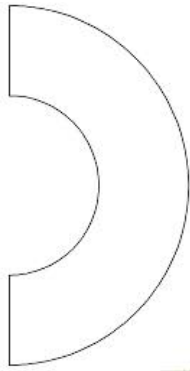
Look at the pictures below that show different phases of the Moon. Below each picture, write the name of the phase it shows. (Hint: They are *not* in the order Earth sees Moon's phases.)

a.	b.	c.	d.
e.	f.	g.	h.

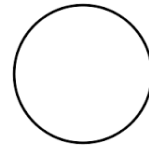
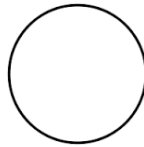
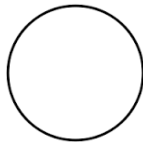
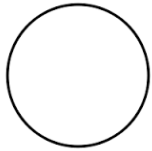
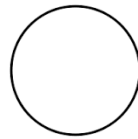
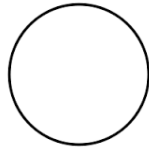
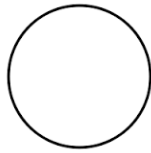
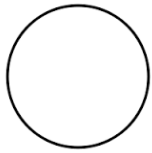
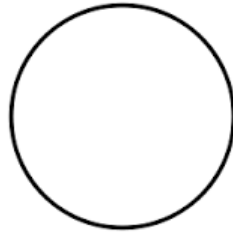
2.) 2 Steps:

Step 1: Label the diagram below with the letters **a, b, c, d, e, f, g, & h** to show where the Moon would be in its orbit when it is the phase shown in each corresponding photo.

Step 2: Draw an arrow on the diagram to show which way the Moon is revolving around Earth.



[www.kidslearninggames.com](http://www.kidslearninggames.com)



(Note: I cut each of these moons out placed them around the circle that represents the earth before photocopying this test)

**Part II: Lunar Reasoning & Applications**

3.) You have been kidnapped and knocked out for some time. When you wake up cold and frightened, you think to yourself, "When will the sun come up?!?" You look up and see a full moon directly overhead.

a.) In about how many hours will the sun rise? (*Assume the sun rises at 6 a.m.*)

b.) How do you know this?

5.) Can you see a Waxing Crescent Moon at midnight?

How do you know this?

6.) What time will a First Quarter Moon rise?

7.) While sitting on the beach at Lincoln City on a lazy summer evening, you accidentally fall asleep. When you wake, you remember that your mom wanted you home by 10:00 p.m. Your cell phone is dead and you have no watch, but you look towards the ocean and see a First Quarter Moon setting in the western sky.

a.) Are you late?

b.) Why or why not?

8.) How does the Moon create the tides on Earth?

9.) a.) How often do high tides occur?

b.) How often do low tides occur?

10.) 2 pts bonus: Explain what "neap" and "spring" tides are...

11.) Combining Tidal Cycles & Lunar Phases

Use ALL of your knowledge of the Moon to help you answer this question.

Arlin K. Postma, expert fisherman and Mrs. Barnes' favorite grandpa, tells her that it is best to go ocean fishing when it's **high tide**. You want to embark on a journey to follow his advice, but your cell phone is dead, and you have no watch. However...you know you saw a **full moon** last night, and when you read a local tidal chart, you see that there was a high tide *during* that full moon. What time TODAY can you launch your boat at Depot Bay into a **high tide**?

- a.) Time:  
b.) Why:

**Part III: Eclipses**

In the boxes below, please draw and label each eclipse with as much detail as possible. Please remember to include:

- the Moon
- Sun
- Earth
- the *penumbra*
- the *umbra*

lunar eclipse	solar eclipse

APPENDIX M  
GRAPHING SPEED PRE-TEST

Name: \_\_\_\_\_ #: \_\_\_\_\_ Homeroom Teacher: \_\_\_\_\_

### Graphing Speed Pretest

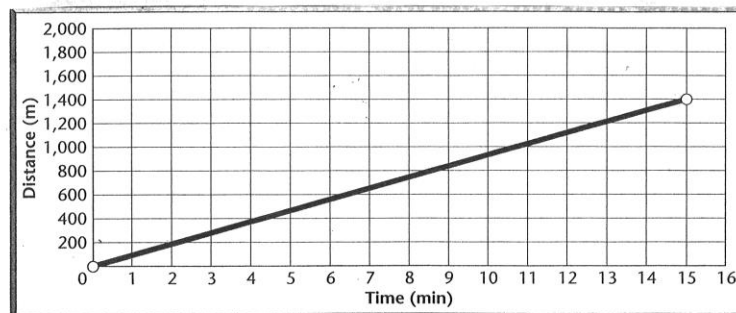
1.) What is the formula used to calculate the speed of an object?

\_\_\_\_\_

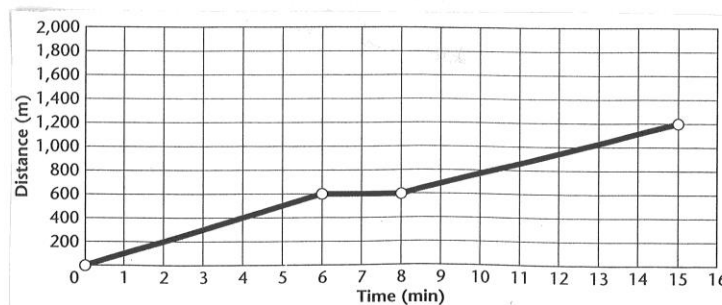
2.) Mr. Barnes likes to ride his bike super fast. He asks Mrs. Barnes to calculate his average speed so he can impress her with how fast he is. What data does he ask Mrs. Barnes to collect so she can make the calculation?

\_\_\_\_\_

3.) The distance-versus-time graph below shows the motion of Mr. Davies on his most recent run. How far did Mr. Davies run in 15 minutes?



4.) The distance-versus-time graph below shows the motion of Mrs. Gwartney on while walking her dogs in her neighborhood. The line is divided into segments. The middle segment is horizontal. What does the horizontal segment tell you about Mrs. Gwartney and her dog's progress between minute 6 and minute 8?



APPENDIX N  
GRAPHING SPEED PERFORMANCE TASK

Name: \_\_\_\_\_ #: \_\_\_\_\_ Homeroom Teacher: \_\_\_\_\_

**Speed Performance Task**

**Bunny Jumping Competition**

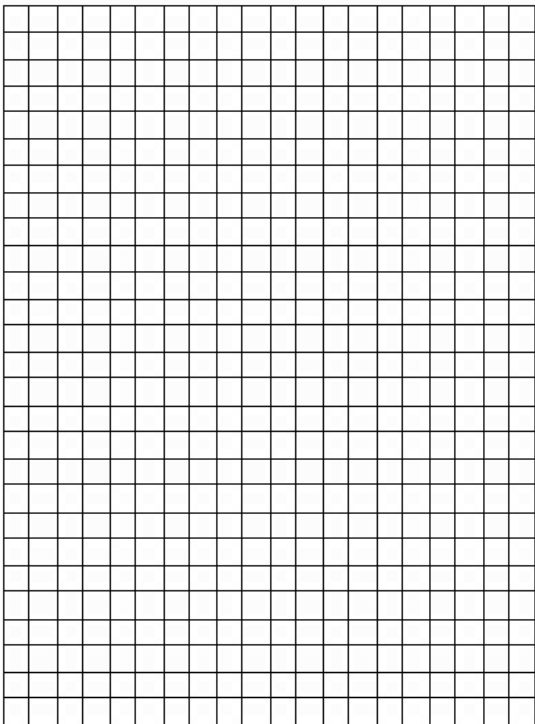
**Bunny #1: Flora (Swedish bunny owned by Tina)**

Time: \_\_\_\_\_ (round to the nearest whole)

**Bunny #2: Cherie (Swedish bunny owned by Magdalena)**

Time: \_\_\_\_\_ (round to the nearest whole)

The distance of a bunny jumping course is approximately 30 meters. Using the graph below, graph both Flora and Cherie's speeds. Use the graph grading rubric to assist you.

<p><b>Graph Grading Rubric</b></p> <table border="1" style="width: 100%; border-collapse: collapse; margin-top: 10px;"> <thead> <tr style="background-color: black; color: white;"> <th style="padding: 5px;">Item Needed</th> <th style="padding: 5px;">Score</th> </tr> </thead> <tbody> <tr> <td style="padding: 5px;">Label x, y, &amp; origin</td> <td style="padding: 5px;">• 3 pts</td> </tr> <tr> <td style="padding: 5px;">Title (who and what)</td> <td style="padding: 5px;">• 1 pt</td> </tr> <tr> <td style="padding: 5px;">Distance Units (Named &amp; Well-Spaced)</td> <td style="padding: 5px;">• 2 pts</td> </tr> <tr> <td style="padding: 5px;">Time Units (Named &amp; Well-Spaced)</td> <td style="padding: 5px;">• 2 pts</td> </tr> <tr> <td style="padding: 5px;">Flora's Speed Line (Labeled)</td> <td style="padding: 5px;">• 1 pt</td> </tr> <tr> <td style="padding: 5px;">Cherie's Speed Line (Labeled)</td> <td style="padding: 5px;">• 1 pt</td> </tr> <tr> <td style="padding: 5px; text-align: center;">Total:</td> <td style="padding: 5px;"></td> </tr> </tbody> </table> <p style="margin-top: 10px;">What was Flora's speed?</p> <p>_____</p> <p>What was Cherie's speed?</p> <p>_____</p>	Item Needed	Score	Label x, y, & origin	• 3 pts	Title (who and what)	• 1 pt	Distance Units (Named & Well-Spaced)	• 2 pts	Time Units (Named & Well-Spaced)	• 2 pts	Flora's Speed Line (Labeled)	• 1 pt	Cherie's Speed Line (Labeled)	• 1 pt	Total:		
Item Needed	Score																
Label x, y, & origin	• 3 pts																
Title (who and what)	• 1 pt																
Distance Units (Named & Well-Spaced)	• 2 pts																
Time Units (Named & Well-Spaced)	• 2 pts																
Flora's Speed Line (Labeled)	• 1 pt																
Cherie's Speed Line (Labeled)	• 1 pt																
Total:																	

APPENDIX O

GRAPHING SPEED WRITTEN SUMMATIVE ASSESSMENT

Name: \_\_\_\_\_ #: \_\_\_\_\_ Homeroom Teacher: \_\_\_\_\_

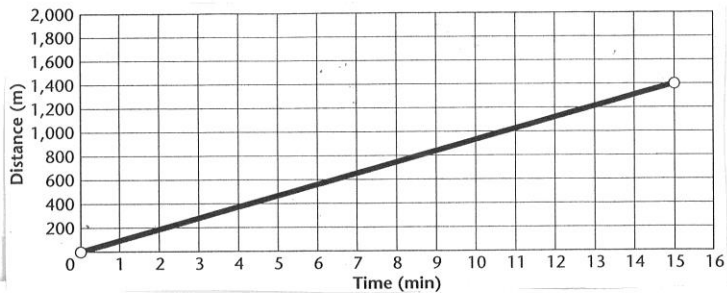
**Graphing Speed: Final Test**  
**Total points:**

1.) Write the math equation one would use to solve the speed of an object:  
 \_\_\_\_\_

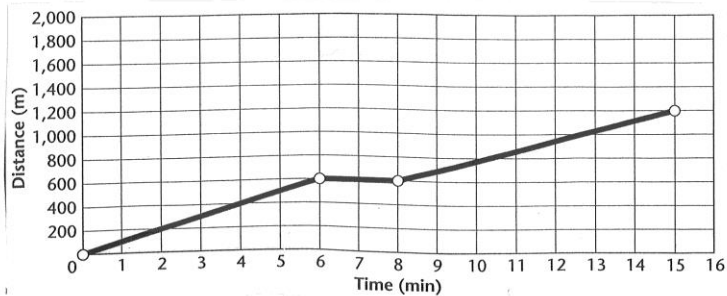
2.) "The steepness of a distance-versus-time graph depends on how quickly or slowly an object is moving."  
 Please circle one:                      TRUE                      FALSE

2.) Mr. Barnes just bought an even faster bike than his last one. He rides at a constant speed, so it is easy for him to ask Mrs. Barnes to collect data on how fast of a rider he is. What data does she collect? (Hint: think about your walking lab!)  
 \_\_\_\_\_  
 \_\_\_\_\_

3.) The distance-versus-time graph below shows the motion of Mr. Cowl on his most recent run. How far had he run at 8 minutes into his run?



4.) The distance-versus-time graph below shows the motion of Mrs. Barnes on while walking her cat in her neighborhood. The line is divided into segments. The middle segment is horizontal. What does the horizontal segment tell you about Mrs. Barnes' cat's progress between minutes 6 and minute 8?



APPENDIX P  
AFTER TEST STUDENT QUESTIONNAIRE



APPENDIX Q

SIXTH GRADE STUDENT POST TREATMENT INTERVIEW

1. What makes a performance task a performance task?
2. What makes a traditional test a traditional test?
3. What science assessments did you most enjoy this year? Why?
4. What science assessments did you not enjoy this year? Why?
5. How do you think 6th grade teachers should assess next years' students in science? Why?
6. Is there anything else you'd like to add?

APPENDIX R

SIXTH GRADE TEACHER POST TREATMENT INTERVIEW

**How did you as the instructor feel about giving the performance tasks?**

**How did you as the instructor feel about giving the written summative assessments?**

**What improvements would you make in planning for or administering PTs in the future?**

**Any other thoughts?**