THE USE OF SELF AND PEER ASSESSMENTS IN A MIDDLE SCHOOL

PHYSICAL SCIENCE LABORATORY SETTING

by

Joshua Allen Dennis

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

Masters of Science

in

Science Education

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2012

STATEMENT OF PERMISSION TO USE

In presenting this professional paper in partial fulfillment of the requirements for a master's degree at Montana State University, I agree that the MSSE program shall make it available to borrowers under rules of the program.

Joshua Allen Dennis

July 2012

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

In order to determine the effectiveness of self and peer assessments for semi-structured labs in an eighth grade physical science classroom setting, students had the option of using a universal lab checklist over the course of five treatments. Lab scores significantly improved for all students, with scores increasing most for students who had used the checklist compared to those who had not.

INTRODUCTION AND BACKGROUND

Project Background

For the previous four years I have been teaching science at Dover Intermediate School in Dover, PA. The intermediate school contains only seventh and eighth grade students with approximately 275-300 students in each grade. On an average school year there are between 130 and 150 students enrolled in my 6 sections of 8th grade physical science. Located on the fringes of suburbia and rural farmland, Dover Area School District is a public school system with a diverse student population, including learning support students, low income families, handicapped students, ESL learners, and a multicultural student body. The population of Dover Intermediate School consists of a 29.4% portion of the student body receiving free/reduced lunch as of 2010, and this portion has been increasing slightly each year over the previous 5 years (Schooldigger, 2010).

Most of my students have had little science exposure in elementary school, with student experiences varying drastically due to time constraints, teacher choices as to whether or not science is taught, a previously unstructured curriculum, and a focus on math and language arts due to state testing. All of my students were enrolled in a life science class during their seventh grade school year. During the life science course, students often participated in lab activities, but were typically making observations, constructing drawings and diagrams and writing responses to questions, with little emphasis on data collection and data analysis. All of these components are important to science, but time, material, and conceptual understanding often allow for little data collection and analysis in comparison with the eighth grade physical science curriculum.

Over the course of several years I noticed students would enjoy the hands-on portion of semi-structured labs and inquiry investigations. Students are often provided with lab activities in class, using semi-structured labs with clearly defined results and inquiry investigations that allow students to lead themselves to results. After the hands-on learning occurred, many students seemed to struggle with some aspects of the labs, including analyzing data and discussing results. Some students struggled to complete the tasks, while others were reluctant to try them. Most students who applied their efforts would show a drastic improvement throughout the year; however, this often required quite a few lab experiences.

Commonly used in language arts classes, self-and peer-assessments are often used to help develop skills needed for success, yet seem to be rarely used in the science classroom. The purpose of this action research was to expedite the process of acquiring these skills, and to improve student lab papers through the use of self- and peer-lab assessment checklists. Students completed a self- and peer-assessment survey using a checklist format set of guidelines. These guidelines allowed students to identify flaws in their lab papers and make corrections before submitting for grading. Both types of assessments were generic in an attempt to use a standardized checklist format for multiple labs. Students participated in surveys designed to gauge the effectiveness of self- and peer-lab assessments and how they influence student attitude towards the lab process.

In an attempt to begin implementing science fairs as a replacement for final exams, student confidence was analyzed at multiple intervals throughout the applied treatments. Student confidence surveys were used to determine students' level of confidence when completing various facets of the experimental process, which applies directly to the independent effort required to complete a science fair style project. The entire treatment: self- and peer- assessments, attitude surveys, and confidence surveys were used for multiple semi-structured labs and for all sections of eighth grade physical science.

Research Focus

Repeated observations occurring during multiple labs over the course of four years led me to investigate one method which may improve the quality of labs produced by students, as well as student attitude towards completing self and peer assessments. The research questions being addressed in this action research include:

Main Question: How does the use of self and peer lab assessments affect student lab achievement?

Sub Questions:

What are the affects of self and peer lab assessments on student attitude?

How do self and peer lab assessments affect student confidence towards designing their own experiments?

How will the use of self and peer lab assessments influence my teaching strategies in the future?

Support Team

Throughout this action research, in addition to MSSE classmates, I met with several colleagues who taught in the same building. Each of the members of the support team, most unknowingly, helped develop the ideas and methods used within this AR and allowed for reflection throughout this research. Prior to data collection, during data collection and at various times throughout the Action Research process, support team members were included in brief conversations, interviews and email correspondences. The results will most certainly be shared with my support team, as they were aware of the Action Research goals before the research was completed. After the research was completed, each team member was provided detailed results and was asked for input to provide more effective self and peer assessments for laboratory use. The colleagues in the support group included:

Margie Manley – A 25 year veteran English and gifted education teacher, Margie has helped provide ideas to encourage the growth of this Action Research project. Having used self and peer assessments in her classroom, she offered valuable insight and questioning skills that assisted in the creation and implementation of data collecting techniques.

Scott Oste – Being a fellow eighth grade science teacher that works in the same building as myself, Scott can be counted on to bounce ideas off of, to create new strategies that work for our students, and to try new ideas. I had provided Scott with a majority of his initial classroom resources when he joined the faculty, but he has filled in the gaps, provided innovative labs and lessons, and has helped improve my pre-existing ideas over the past three years.

Chuck Richards – Teaching social studies on the same team as myself, Chuck is also experienced in classroom data collection, using formative assessments, and thinking outside of the box. Chuck often comes up with innovative methods of improving student learning.

Julie Sterner – Being the administrator in my building, Julie had highly encouraged this project from the day I introduced it to her. She was fascinated with the project, the data collection methods and nature of assessing student learning in science.

CONCEPTUAL FRAMEWORK

Introduction and Background

The use of self- and peer-assessments seemed commonplace in writing-intensive courses such as language arts, yet seldom discussed in a science setting. This observation led to the idea of using self- and peer-assessments as a method to improve student achievement on structured labs. The existence of self- and peer-assessments can be accredited to George Jardine, who attempted to prepare his students for immersion in society, and focused on students' abilities to reason, evaluate, investigate, judge, write and speak (Gaillett, 1994). While not focused on the speaking aspect, the self- and peer-evaluations for this study were based on similar aspects as Jardine's ideas from two centuries ago.

In order to determine if self- and peer-reviews are effective, each lab assignment was self-assessed by students, followed by a peer-review. These assessments acted as a rubric for students to evaluate their own work before ever being submitted to the teacher. According to Andrade (1997):

> Rubrics can improve student performance, as well as monitor it, by making teachers' expectations clear and by showing students how to meet these expectations. The result is often marked improvement in the quality of student work and in the learning. Thus, the most common argument for using rubrics is they help define quality (p. 3).

Checklist designed rubrics were used in order to prevent the focus on the likes and dislikes of students' responses, and in preventing inaccurate and inconsistent feedback from classmates during self and peer assessments (Nilson, 2003). Surveys were used after completing several labs to determine if students felt that the self and peer assessments were effective at improving the quality of their labs papers.

Self-assessments have long been a part of teachers' repertoire, to the extent that some teachers may suggest using self-assessments for every large assignment or as Ross (2006) suggests, when:

> (i) students create something that requires higher level thinking (i.e., they interpret their performance using overt criteria); (ii) the task requires disciplined inquiry, (i.e., the criteria for appraisal are derived from a specific discipline); (iii) the assessment is transparent (i.e., procedures, criteria and standards are public); and (iv) the student has opportunities for feedback and revision during the task (e.g., by responding to discrepancies between the student's and teacher's judgment) (p. 2).

For my action research, despite using semi-structured lab activities, higher level thinking is used in order to come to conclusions and to describe the collected data. Students must also have knowledge of a specific discipline as per the topic of the class at the time of the lab. The procedures, criteria, etc., are clearly defined and shared via a teacher provided structured lab paper. Students have the opportunity for feedback, through discourse before, during, and after the lab, as well as by responding to teacher comments on their lab papers.

Direction and Focus

In some instances, teacher directed feedback is not enough feedback for students. Having students reflect on their own work, through the use of self- and peer-assessments and other type of formative assessments allows students to develop a set of skills useful for problem-solving, answering questions, and framing questions of their own (Sadler, 1998). Self-assessments have long been proven as reliable classroom evaluation tools, typically providing valid data with accuracies of 80% or more when comparing self-evaluation to teacher grading. Even without additional lessons on self-assessment, studies have shown that completing self-assessments results in increased self-efficacy and improved student achievement (Ross, 2006). Ideally, the improved student achievement and increased self-efficacy would result in improved student confidence as well.

Peer assessments are often used in conjunction with self-assessments, due to their similar nature and the opportunity to improve self-assessments by allowing the assessor to apply the same guidelines to another student's work. They have been attributed to guiding students to learn that mistakes are opportunities, teaching valuable life skills such as analyzing and evaluating. Studies have shown that despite varying responses from peers, peer-assessments are typically valid and reliable methods of evaluation (Bostock, 2000).

In an attempt to introduce a new method of formative assessment to the science curriculum at my current placement, some guidance was necessary for justification and focusing of ideas. In one case study (Black, 2001) a teacher reported on the King's Medway Oxford Formative Assessment Project (KMOFAP), a collection of twelve science and twelve mathematics teachers challenged to using new forms of formative assessment, stating that it was worthwhile to cut extension activities and reduce repetitions when teaching to students who lack confidence, who are less likely to believe they are capable. When students thought they were not capable, they usually weren't, as evidence supports. It was generally agreed upon that it was better to cover fewer concepts, but ensure student confidence increased, as their understanding of the information correlated with confidence. This justified the use of valuable class time (four classes) in order to attempt the use of the self- and peer-assessments. One such example within KMOFAP required six months of use, before formative assessments were adapted to changing classroom climates to be considered effective enough for long term use. This was a challenge for this Action Research, as the applied treatments were intended to be short term duration, and classroom climate was not monitored as a holistic approach.

Brown (2004) suggests that changes in assessment techniques have become essential due to the increasing diversity of student populations. Brown (2004) also suggests the one critical factor is not necessarily the nature of the assessment, but the timing of the application. Assessments provided early in the learning process have resulted in students having the opportunity to learn from their mistakes. Although timing is critical, the nature of assessment can vary, and current trends and data suggest the use of formative assessment is becoming more accepted due to its authentic nature. One key component to authentic assessment is to show evidence of achievement rather than regurgitation. Brown (2004) also discusses the validity of these types of assessments, asserting that the focus of the assessments need to focus on what was intended to be learned compared to what the students have learned.

In one case study, (Macdonald, 2004) a teacher observed that certain students turned in consecutive assessments with repeated, similar mistakes. In order to try to correct the situation, she had students self-assess their assignments using a very detailed self-assessment form. The teacher did face initial complaints and hesitations, but students soon learned what was expected for each assignment. Students not only used this self-assessment format to determine if they met the desired criterion, but also to assign themselves a grade on the assignment. In one example, 80% of students had assigned themselves appropriate marks. The self-evaluation format was applicable to several different types of assessments as well, making this a universal idea. I had students self-assess their lab assignments before submitting them, for the same reasons Macdonald had observed.

Framework and Methodology

Much can be learned from self- and peer-assessments, for the students and for their teachers. Students who clearly grasped the ideas of the laboratory activity should easily be able to assess other students' work to determine if they also grasped concepts such as data collection, data analysis and drawing conclusions from the data collected. In order for students to show improvements in achievement, they should at least experience the lab process, receive teacher feedback, and have an opportunity to discuss the lab with their peers and teacher, before self and peer assessments are begun (Doran, Chan, Tamir, & Lenhardt, 2002). In my AR this was accomplished by completing one structured lab as traditionally done, with student feedback being limited only to teacher comments. This lab, the Paper Airplane Lab, was used as baseline data to compare student achievement for all other labs.

Self-confidence is often correlated with student achievement, and is best analyzed when students are learning specific, unfamiliar skills or when students have repeatedly struggled with previously learned skills (Angelo & Cross, 1993). Surveys are often used to assess student confidence, which typically consist of a few questions designed to get a rough idea of the students’ level of confidence. A lack of student confidence may identify areas of weakness or indirectly students’ anxiety levels. One drawback to student confidence surveys is that some students grossly underrate or overrate their ability level and thus their confidence (Angelo & Cross, 1993). I would expect that if this was the case, over the course of multiple trials of completing the confidence survey, students would get a better grasp as to specific areas where they show strength or weakness, due to multiple lab experiences and teacher provided feedback. Student achievement, or the grades received on lab papers, along with teacher comments, should quickly help students identify these areas.

A similar study used universal lab rubrics with seventh grade life science students (N = 131) and resulted in improved student lab scores, decreased time required to score labs, and students showed improvement which extended beyond the science classroom, when examining achievement in other classes using rubrics (Mullen, 2003). Student lab scores improved from a D to a B- average in this instance, and students developed skills required to critique and analyze their own work while reaching high expectations.

In another similar study, tenth grade students showed improvement on their written lab reports (+15%) and even requested additional peer-review rubrics for labs not included in the applied treatments (Acker, 2011). Students were often very critical of each other's work, yet did not hesitate to offer suggestions for improvement. While not monitored directly, confidence in writing ability seemed to improve as well. Both of these studies support the idea that self and peer reviewed lab papers using rubrics will result in improved performance, achievement and confidence.

For my Action Research, I used a combination of ideas from several resources, each of which are designed to answer the research questions as described in this paper. The data collection methods were unique to my placement and the current practices within my department.

METHODOLOGY

With nearly 130 students enrolled in my 6 8th grade physical science classes, students are often provided with both inquiry-based labs and semi-structured labs. The labs used for this Action Research were used during the 2011-2012 school year and spanned the course of multiple units. Each of the labs provided some structure for students based on traditional lab reports, yet allowed for restricted amounts of time provided for students for teacher and peer interaction. While semi-structured, each of these labs allowed for students to come to their own conclusions about any data collected. Most semi-structured labs were designed to allow for a brief explanation of the goal(s) of the lab, provided in a question format. In an ideal setting, students collected data within one class period, which is a forty-two minute period, and then completed the lab in class the following day. Several data collection methods were employed during this action research project, including analysis of raw lab scores, two types of surveys which included both Likert scale data and qualitative responses, student interviews, and quantitative data collected in the form of students' scores.

The intent of this Action Research was to determine the effectiveness of self-and-peer-assessments when used with semi-structured labs. Each treatment and data collection method was offered to every student (max N = 124) for the course of five labs. 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.

Treatment: Self and Peer Assessments

After completing each lab, students used a standardized self-and peer-assessment, a checklist-style rubric that addressed each main component of a typical lab and other structured experiments. Students were not provided with explicit directions to evaluate scoring of the semi-structured labs due to the slight variation of style of each lab, but upon completion, scoring, and returning baseline labs to students, a description of how each portion of a lab is scored was supplied for each class. This treatment was used to address my main question: how does the use of self-and-peer lab assessments affect student lab achievement? After completing each lab activity/paper, students were given a self-assessment checklist (Appendix A). Students evaluated their own work using the self-assessment before having another randomly assigned classmate use the peer-assessment checklist. Peer-assessors were assigned randomly to ensure students were not consciously choosing friends and/or students that may provide biased or inaccurate responses.

Whether completing the self- or peer-assessments, students read the description of the item and marked accordingly, with options ranging from “Disagree” to “Agree.” For example, if Student A was assessing Student B’s lab paper, and thinks Student B did an excellent job with writing his/her hypothesis as a statement, Student A would check or mark the “Agree” box.

After a student completed the self-assessment, he or she met with his or her randomly assigned classmate in order to complete their peer-assessments. Students swapped papers, completed the peer lab assessment and returned papers. Once both the self- and peer-lab-assessments were completed, students could easily compare how their classmate reviewed their lab writings. This allowed students to seek help if there were large disagreements, and make any revisions necessary before submitting their lab paper for a grade. Students had this opportunity for each of the five applied treatments, which were spread out through nearly a semester of class.

Data Collection Methods

The data collected to answer the research questions involved several items, including a post lab survey, test scores, confidence survey, and student interviews (see Table 1). The labs themselves were scored in the style as traditionally used in my current placement, in order to determine the affects of self-and-peer-assessments on student achievement in the form of percent grades. Each of the surveys included a checklist design with several open ended responses, allowing for both quantitative and qualitative data.

Table 1
Data Collection Methods Used

Collection Tools	**Student Attributes**			**Description of Utilization**
	Achievement	Attitude	Confidence	
Post Lab Survey		X	X	This survey is completed after scored student labs are returned to students.
Confidence Survey			X	This is administered to students three times; pre, during, and post all treatments.
Student Interview		X	X	Students were interviewed at the end of the treatments in order to clarify data.
Test Scores	X			Student test scores were compared with associated lab scores.
Lab Scores	X			Student lab scores were analyzed for each treatment.

Data Collection Method: Lab Survey

In order to determine student attitude towards the treatment process, each lab survey was analyzed based on the corresponding semi-structured lab. This lab survey (Appendix B) was administered only after completing the self-and-peer-assessments and submitting the lab for a score. Questions on the survey represented a range of student attitudes towards using the lab assessments, and determined if they put the lab assessments to use by revising their lab paper before submission.

Questions 1 and 2 have neither a positive or negative connotation, in part to determine if students thought the lab assessments were equally valuable, or if one was more valuable than another. Numbers 3, 5, 6, 8, and 9 were all positive statements. For these items, student responses consisting of "Agree" or "Somewhat Agree" were considered reflections of a positive attitude towards self/peer-lab-assessments. Questions 4, 7, and 10 were negative statements, where responses of "Disagree" or "Somewhat Disagree" were considered reflections of positive attitude. Students were asked to include comments about the process, including their likes, dislikes, and identifying an area where the self/peer-assessments have helped them in some way. The qualitative feedback was very useful in explaining student responses, and for uncovering student attitudes.

Each of the questions was analyzed separately in order to determine change in student attitude throughout the action research treatments. For the positive statement items on the survey, the responses were assigned a value to determine a numerical score (A = 2, SA=1, SD=-1, D=-2). The negative statement items were assigned the opposite values as the positive statements, and were used to determine change in student attitude, as well as to validate the positive statements.

Data Collection Method: Confidence Survey

Confidence surveys were utilized at three different times during this Action Research: before treatments began, half way through all treatments and after all treatments were completed. Students were asked to complete confidence surveys (see Appendix C) to assess various aspects of the lab process, which were also addressed in the self-and-peer-assessments. Those students participating in the confidence survey selected a numbered box that correlates with a scale of how confident they felt about completing the lab tasks. The idea is that if students felt increasingly confident towards completing those portions of the lab, they will feel more confident in the future when designing their own experiment in a science fair designed project.

Data Collection Method: Student Interview

At the conclusion of the last treatment, students showing distinct patterns in their achievement and/or attitude were interviewed to probe for more information and personal accounts of their actions and thoughts throughout the AR process. Six students with low or decreasing lab scores and six students with high/increasing lab scores were selected to be interviewed. Also, six students showing a negative/increasingly negative attitude and six showing a positive/increasingly positive attitude were chosen for interviews. All students meeting the criteria set forth were placed in a pool of candidates to be chosen randomly, and it was ensured that the same student was not selected for more than one interview. Interviews were not specifically structured, but were conducted in a manner to probe students' true thoughts towards the self- and peer-assessment process and any implications for changes to be made to the process in the future. Interviews were transcribed, with student responses recorded.

Data Analysis Plan

Student lab scores were the main focus of this Action Research, in order to determine the effects of self and peer assessments. Along with lab scores, test scores from two associated tests were used to determine if self- and peer-assessments had impacts on student achievement extending beyond the labs. Surveys were analyzed to determine student attitude towards the assessments and confidence throughout the treatments. Individual interviews were used to provide student responses in order to support and validate any findings.

DATA AND ANALYSIS

Lab scores, survey results, and interview responses were used to answer each of the initial questions asked for this Action Research. While each was analyzed separately, they were all used to support the effectiveness of self and peer assessments. Student lab scores were used to determine if the self and peer assessments resulted in improved grades. Each lab used for the treatments, including the baseline was valued at 30 points. The first lab of the year, the Paper Plane Lab (PPL) was used for baseline data, and was completed entirely in class as students were walked through the process of completing a lab based on the scientific method, while being provided with a thorough example of a well done lab paper. For each of the treatments, the Penny Drop Lab (PDL), Float Your Boat Lab (FYB), Gummy Bear Lab (GBL), Separating a Mixture Lab (SAM) and Rate of Solution Lab (RSL) students were given less teacher influenced answers and instead provided the self and peer assessment checklists. The data collected from these treatments was analyzed and sorted into several subgroups.

Each subgroup was defined based on whether or not students completed the self and peer assessments, and whether students were in an academic or honors class. With little experience with science, honors students are recommended by seventh grade life science teachers, or requested by parents, in order to allow for additional higher level thinking and group interactions during their eighth grade science class experiences. Some students included in the academic group could and should be in an honors placement, and vice versa for honors students, but recommendations and requests trump ideal placement.

During each treatment, the *N* values differed due to multiple reasons. Other than student absences, some students chose not to complete the self and peer assessments and submitted their labs as they were, or submitted a blank or mostly incomplete assessment checklist.

Table 2
N Values for Treatments Completed

Lab Name	***N* (Did Complete SA/PA)**			***N* (Did Not Complete SA/PA)**			**Total *N***
	Honors	**Academic**	**Subtotal**	**Honors**	**Academic**	**Subtotal**	
Paper Plane Lab	X	X	**X**	24	85	**109**	**109**
Penny Drop Lab	21	64	**85**	3	35	**38**	**123**
Float Your Boat Lab	18	73	**91**	6	26	**32**	**123**
Gummy Bear Lab	18	74	**92**	3	9	**12**	**104**
Separating a Mixture Lab	17	72	**89**	4	17	**21**	**110**
Rate of Solution Lab	16	78	**94**	5	1	**6**	**100**

Each treatment was offered to every student in each class, with a possible *N* value of 127 students. Oftentimes, students fail to complete assignments when absent, and rarely complete assignments outside of class, resulting in varied *N* values (see Table 2). In order to compromise with other courses, parents, administrators, and students, any student that missed a portion of the treatment was offered the self and peer assessments, however time restrictions often prevented them from completing them. This resulted in varying *N* values throughout the treatments, along with schedule changes, and health concerns such as several students who had concussions. Those students who were absent for any treatment were exempt due to academic concerns, in order to ensure missing assignments were completed, and due to the treatments having no pre-determined, direct effect on student grades. Due to difficulties finding time for absent students to complete assignments upon their return to class, students were exempted from the additional time consuming process of completing the self and peer assessments. Even if students were able to complete the labs in a timely manner upon their return, removing other students from scheduled class time to assist in completing the peer assessments was not a logical solution. The removal of these students from other classes may have negatively impacted multiple students' achievements in those classes, activities and assessments for my class, and attitude towards the treatments as they were applied in this Action Research.

Over the course of the five treatments, more students were completing the self and peer assessments for their semi-structured labs. The specific reasons may not be known for this particular Action Research; however data collected in the form of surveys, qualitative responses, and interviews imply that more students had given up on resisting the additional time and work required, or had recognized the benefit of completing the assessments and were attracted to the idea of improving their grade on assigned labs.

Student Achievement

Data from each of the five treatments were compared to the baseline lab scores to determine if there was a difference between lab scores for students that had completed the self and peer assessments and those that had not. The baseline lab scores were used only as a comparison score for each of the five treatments. The Paper Airplane Lab (PPL) was used as a baseline score to compare with each treatment (Appendix E). The lab featured simple measurements, graphing techniques, and data analysis.

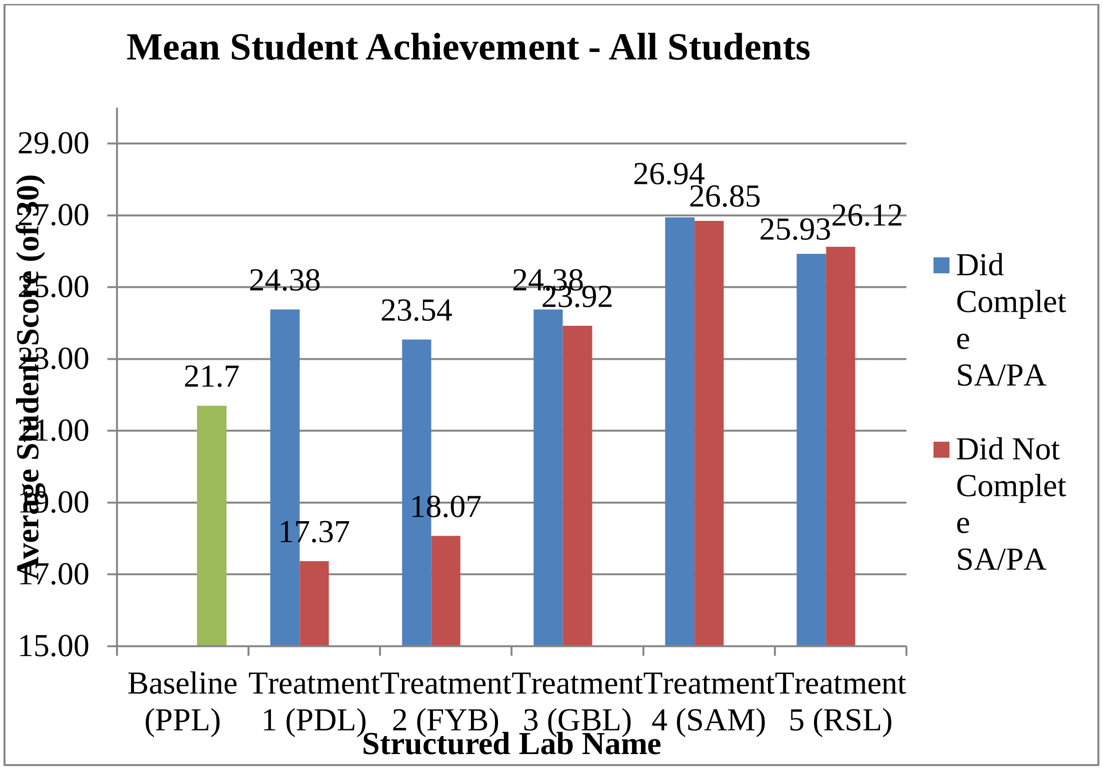

Figure 1. Student achievement showing all participants' and non-participants' lab scores for each treatment. *Note.* Paper Airplane Lab= *N*= 109, Penny Drop Lab= *N*=123, Float Your Boat= *N*=123, Gummy Bear Lab= *N*= 104, Separating a Mixture Lab= *N*= 110, Rate of Solution Lab= *N*=100.

Mean lab scores of all students clearly showed there was a difference between students that had completed the self and peer assessments and those that had not (Figure 1). After the first treatment using the Penny Drop Lab (Appendix F), the mean lab scores of students that had completed the self and peer assessments were higher (23.4%) than those students that had not completed the self and peer assessments. The Penny Drop

Lab (PDL) was similar in nature to the baseline lab in that it required simple measurements to be made using a repetitive process.

Each treatment was analyzed using a two tailed t-test, assuming equal variances. Students who had completed the self-and-peer-assessments were compared with students who had not completed them for each of the five treatments. Lab scores were also compared with the baseline lab scores, in order to determine how the mean, median and spread of the scores changed throughout the treatment applications.

Students that had completed the assessments also showed scores that averaged higher than (8.9%) the baseline lab scores, showing significant improvement $t(37)=5.72$, $p<0.001$, without significant teacher input. This same trend continued throughout the following treatments, when comparing students that had completed the self and peer assessments with those that hadn't, the gap between the two groups decreased until both groups of students had similar mean lab scores.

Lab scores for all students were shown using a box and whisker plot (see Figure 2), where the middle 50% of the students' scores are shown using a box. Whiskers were then added to the top and bottom of the box to represent one point five times the limits of the box, with any outlier lab scores represented as dots.

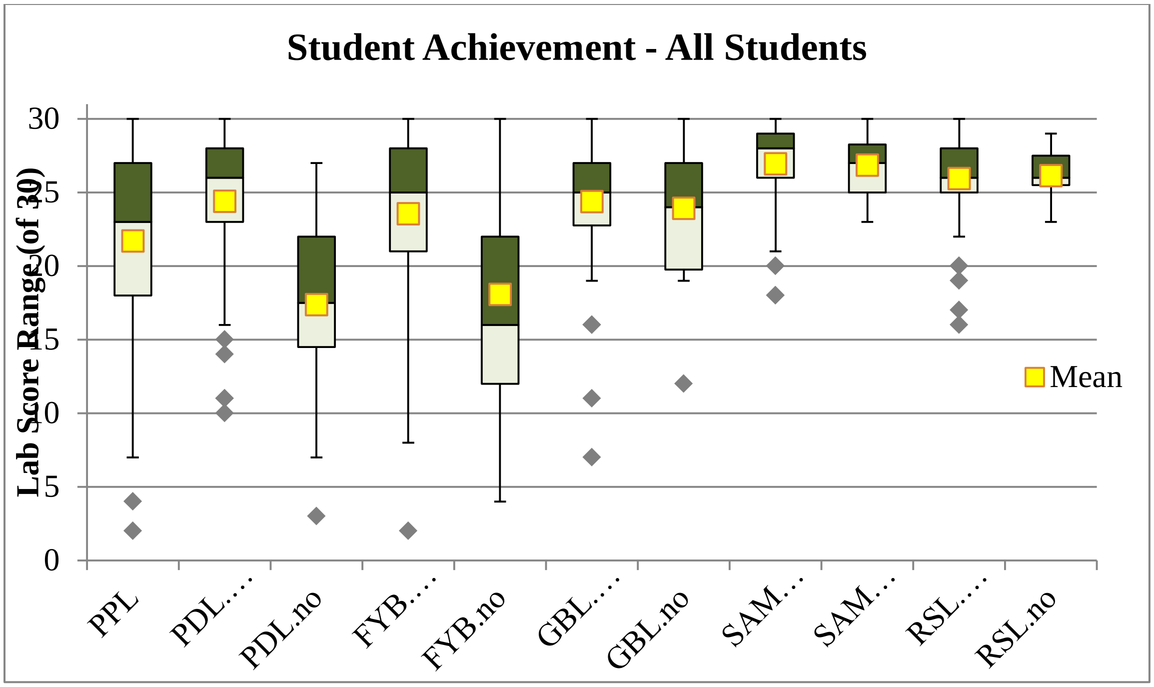

Figure 2. Student achievement showing the middle 50% of the lab scores, with outliers, for each treatment, with 'yes' indicating students had completed self and peer assessments. *Note*. PPL= *N*= 109, PDL= *N*= 123, FYB= *N*=123, GBL= *N*= 104, SAM= *N*= 110 , RSL= *N*= 100.

Not only did the mean scores rise and the gaps close over the course of five treatments, but the spread of the data decreased in range and also showed improvement in student achievement (Figure 2). The baseline lab (PPL) had the largest range of student scores (28) with a generally decreasing value during each treatment, and for each subgroup, with the smallest range occurring during the fourth (7) and fifth treatments (6). During the fourth and fifth treatments, the range of scores was smaller for students that had chosen not to complete the self- and peer-assessments than those that had completed them. This suggests that these students did not need the self and peer assessments in order to complete a quality lab paper, and have learned the expectations of a thoroughly completed lab assignment.

During a second treatment, the Float Your Boat Lab (FYB) was used and student scores' decreased in all groups, however the mean scores (18.2%) and the middle fifty percent of student scores showed a clear difference. The range of the scores may be due to the more abstract nature of the lab in comparison to other semi-structured labs typically used in a middle school setting. Students did not have a specific plan to follow (Appendix G), allowing for more abstract thought and planning to accomplish the task. The second treatment followed a similar pattern as the first, and although the spread of lab scores was more skewed to the negative, the differences between students that had completed the self and peer assessments and those that had been repeated and were significant, when analyzed using a two-tailed t-test, assuming equal variance $t(13)$, $p=0.03$. Each of the remaining treatments was more typical of the semi-structured labs used, which allowed students to return to a more familiar format.

The Gummy Bear Lab (GBL) was the semi-structured lab used for the third treatment and required students to make simple measurements in order to calculate density of a gummy bear before and after soaking in water (Appendix H). By the third treatment, most students seemed to have caught on to the expectations of the requirements for labs, as indicated by the nearly identical mean scores (<1%). The spread of lab scores decreased in both groups, with students who had chosen not to complete the self and peer assessments showing a more skewed spread (see Figure 2), although the difference was insignificant $t(13)=0.30$, $p= 0.76$. Students that had completed the assessments showed improvement (8.9%) compared to the baseline lab.

The fourth treatment (Appendix I) utilized a lab titled Separating A Mixture (SAM) and had results nearly identical to the third treatment in terms of difference between mean student scores. When examining all students who had completed the assessments with those students that had not, the distribution of scores was insignificant $t(38)=0.17$, p=0.86. The mean student scores and decreasing range of scores showed improvement compared to all previous treatments (Figure 2), with a drastic improvement in relation to the baseline lab (+17.5%).

The fifth and final treatment (see Appendix J), the Rate of Solution Lab (RSL) closely resembles the third and fourth treatments in both format and results(see Figure 2), with no noticeable difference between the two groups of students when examining mean scores (<1%) and distribution, $t(10)=-0.26$, p=0.80. This treatment was completed after an extended holiday/winter break, which may have influenced the scores of student lab papers, however the results repeated previous treatments.

Throughout each of the treatments, the gap between students who had not completed the self and peer assessments and those that had completed them got smaller in size, and increased in value. Throughout each treatment, outliers were present in each treatment, and generally increased in value (Figure 2). When examining all students, data clearly shows improvement, but does not analyze honors and academic subgroups of students. With unequal numbers of students in each group, one group could easily skew the results.

<u>Academic vs. Honors Students</u>

Of the six classes of eighth grade physical science represented when examining all students, a majority were considered academic students (*N*=101) while only one class was considered an honors level course (*N*=26*)*. Each category of student, honors and academic showed similar patterns when it came to student achievement, although the actual differences in scores between those students that completed the assessments and those that had not was quite different.

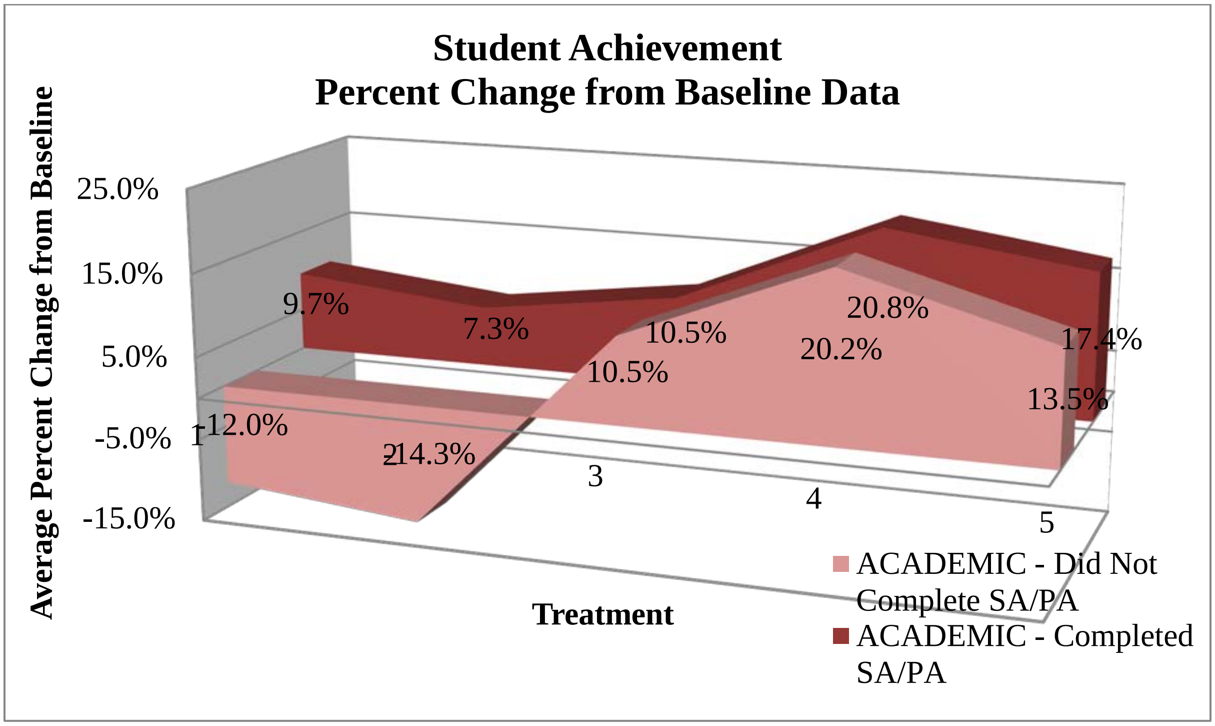

Figure 3. Student achievement for only academic students' lab scores that participated in each treatment. *Note.* Paper Airplane Lab= *N*= 85, Penny Drop Lab= *N*= 99, Float Your Boat= *N* = 99, Gummy Bear Lab = *N*= 83, Separating a Mixture Lab = *N*= 89, Rate of Solution Lab= *N*= 79.

By separating the entire population of participating students into honors and academic students, the differences in how each subgroup changed became apparent. When comparing each treatment consecutively, the mean scores of academic students show a drastic change throughout the treatments when compared to the baseline lab (see Figure 3). By the third treatment, it did not seem to matter whether or not students completed the self and peer assessments, as their mean scores were similar. This may suggest that most students learned the expectations of lab assignments and no longer required the self- and peer-assessments to double check their papers before submitting them for teacher grading. Mean scores for students who did not complete self and peer assessments did decrease during the fifth treatment, which may have been affected by the long time frame between treatments due to an extended holiday break. The decrease in scores for the fifth treatment may imply that after an extended absence from a structured routine, students may ignore or forget expectations for completing a quality semi-structured lab.

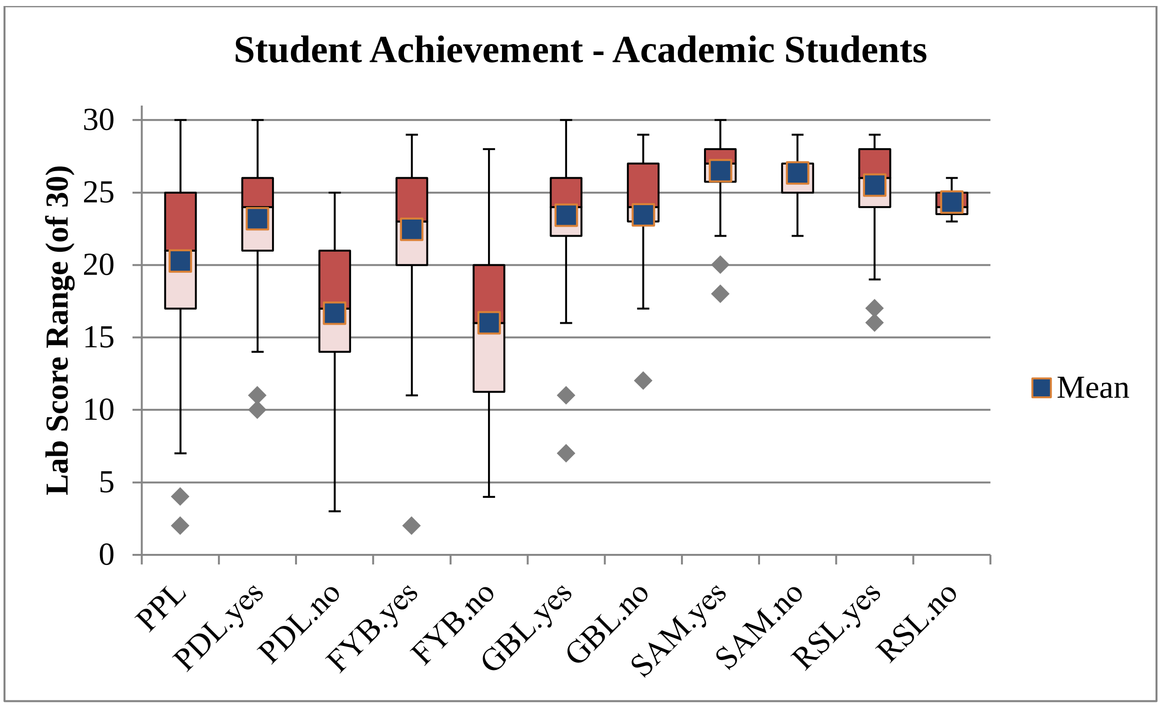

Figure 4. Student achievement showing the middle 50% of the lab scores, with outliers, for academic students only, with 'yes' indicating students had completed self and peer assessments. *Note.* Paper Airplane Lab= *N*= 85, Penny Drop Lab= *N*= 99, Float Your Boat= *N*=99, Gummy Bear Lab =*N*= 83, Separating a Mixture Lab= *N*= 89, Rate of Solution Lab= *N*= 79.

Academic students represented a majority of the students, and strongly influenced the data shown in Figure 2. The mean scores were slightly lower in comparison with the entire participating population, with the range of lab scores (see Figure 4) also lower, and mimics those of all students combined.   Outliers were found in all treatments, and also showed improvement over the course of the applied treatments. No individual student was responsible for multiple outliers, which may suggest students may have had an off day, chosen not to take the self- and peer-assessments seriously or just chose not to complete the assessments at all.

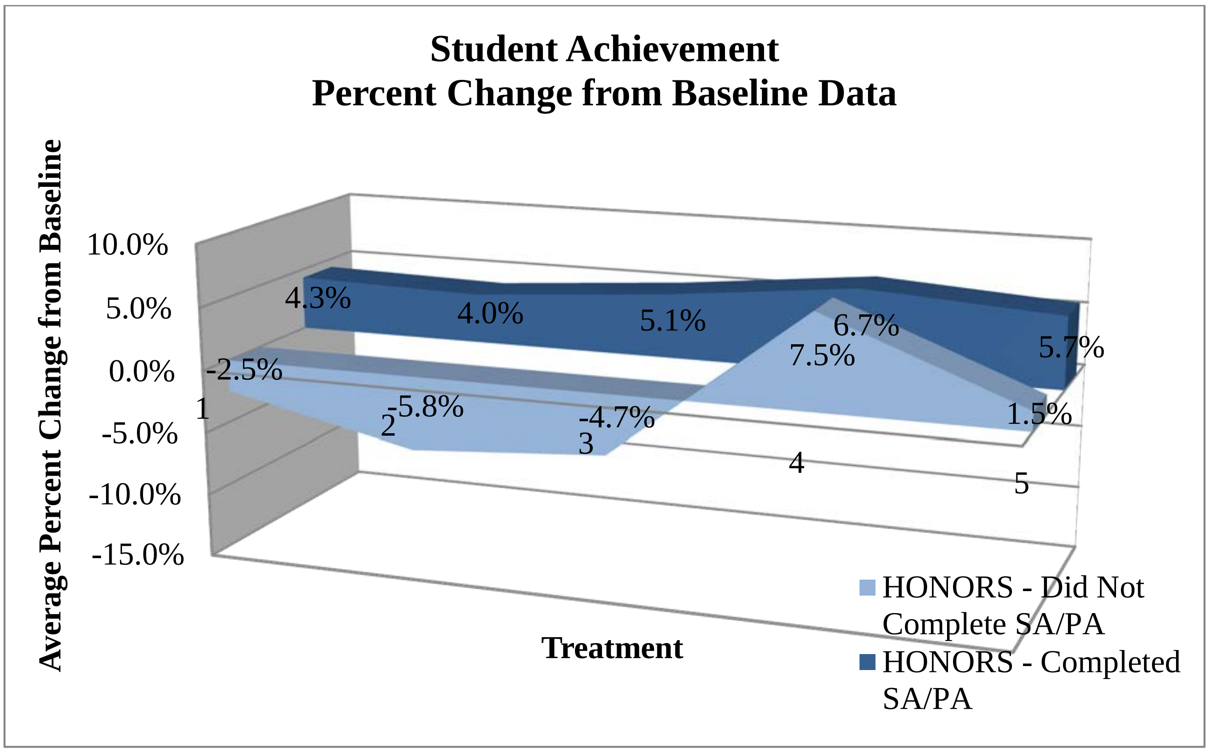

Figure 5. Student achievement for only honors students' lab scores that participated in each treatment. *Note.* Paper Airplane Lab= *N*= 24, Penny Drop Lab= *N*= 25, Float Your Boat= *N*= 25, Gummy Bear Lab= *N*= 21, Separating a Mixture Lab= *N*= 21, Rate of Solution Lab= *N*= 21.

Honors students who had completed the self- and peer-assessments showed a gradually increasing score across most treatments (Figure 5), while those students who had not completed the self and peer assessments required several labs to catch up to their peers. During the first three treatments, students who did not complete the assessments scored lower by several percent (≤5.8%), but had caught up to and surpassed their classmates' scores by the fourth treatment. As with the academic students, scores dipped on the last treatment, decreasing more for students who had not completed the self and peer assessments (4.2%). The mean scores suggest that honors students required more applications than academic students, before each group of honors students was equal. The honors students appeared to show less improvement than academic students, which is most likely due to their higher achievement to begin with, which allows for less improvement than academic students on each of the treatments.

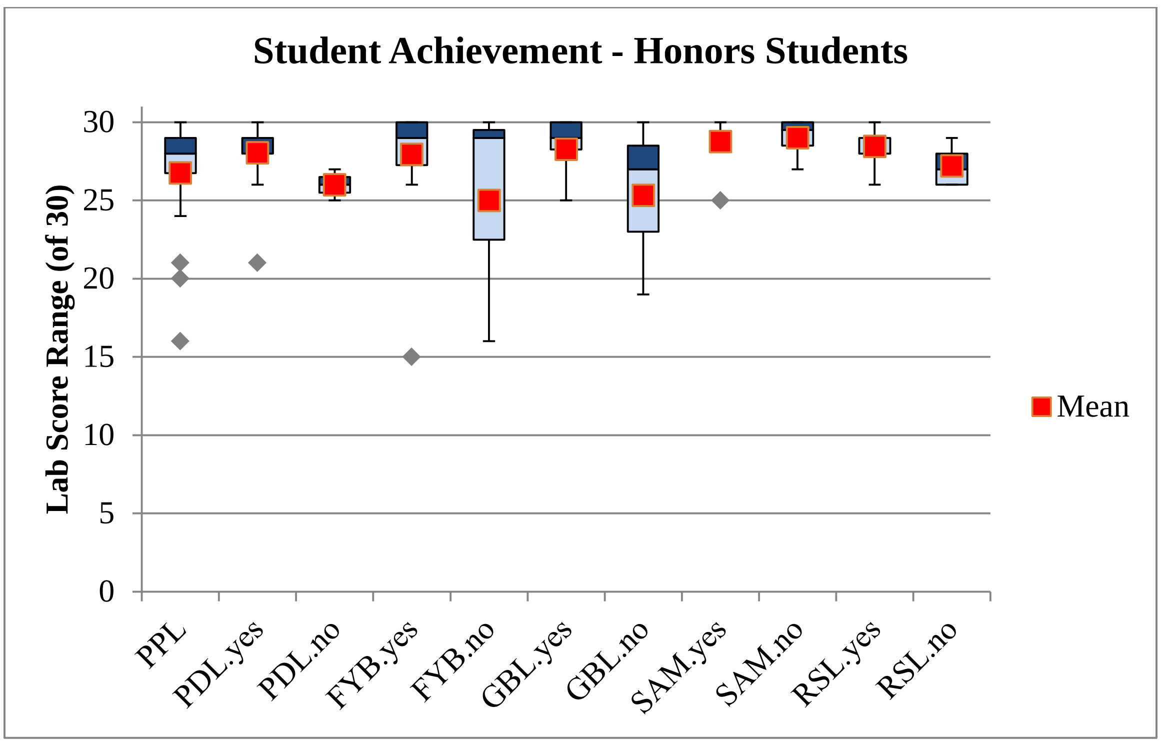

Figure 6. Student achievement showing the middle 50% of the lab scores, with outliers, for honors students only. *Note.* Paper Airplane Lab= *N*= 24, Penny Drop Lab= *N*= 25, Float Your Boat= *N*= 25, Gummy Bear Lab= *N* = 21, Separating a Mixture Lab= *N* = 21, Rate of Solution Lab= *N*= 21.

The spread of the lab scores of honors students suggest there was a very slight difference between students who had completed the self- and peer-assessments and those that had not. With very small *N* values for some treatments (*N*=3), further analysis was not completed. The pattern was similar to overall population, but it can be seen how the honors students skewed the overall population scores to the positive, as seen by examining Figures 1, 2, and 3. The spread did change slightly throughout the treatments, although it can be stated that honors students did not benefit from the self and peer assessments nearly as much as academic students. This may have been due to the higher achievement on labs to begin with, which would allow for less improvement than the lower scoring academic students.

Lab Scores and Test Scores

Within the grading structure of my classroom, labs and tests hold equal weighting, with labs often providing a hands-on opportunity for students to learn and observe concepts that may not be addressed within each test due to time and material constraints. To determine if the self- and peer-assessment checklists influenced other forms of classroom assessment, two tests were chosen that closely mirrored the desired outcomes of similarly matched labs. Labs and tests were chosen for comparison based on the nature of the assessment and the similar content. Only academic students' results were examined, due to greater improvement on lab scores in comparison to honors students. Because academic students benefitted more from the self and peer assessments when it came to lab scores, they may have also showed more improvement when it came to test scores.

The first treatment, the Penny Drop Lab was chosen to compare with the second test of the school year, a Measurement Practical (Figure 7). The Measurement Practical required students to simply make various measurements, including length, mass, and volume, which is similar in nature to Penny Drop Lab due to the requirement of making simple measurements.

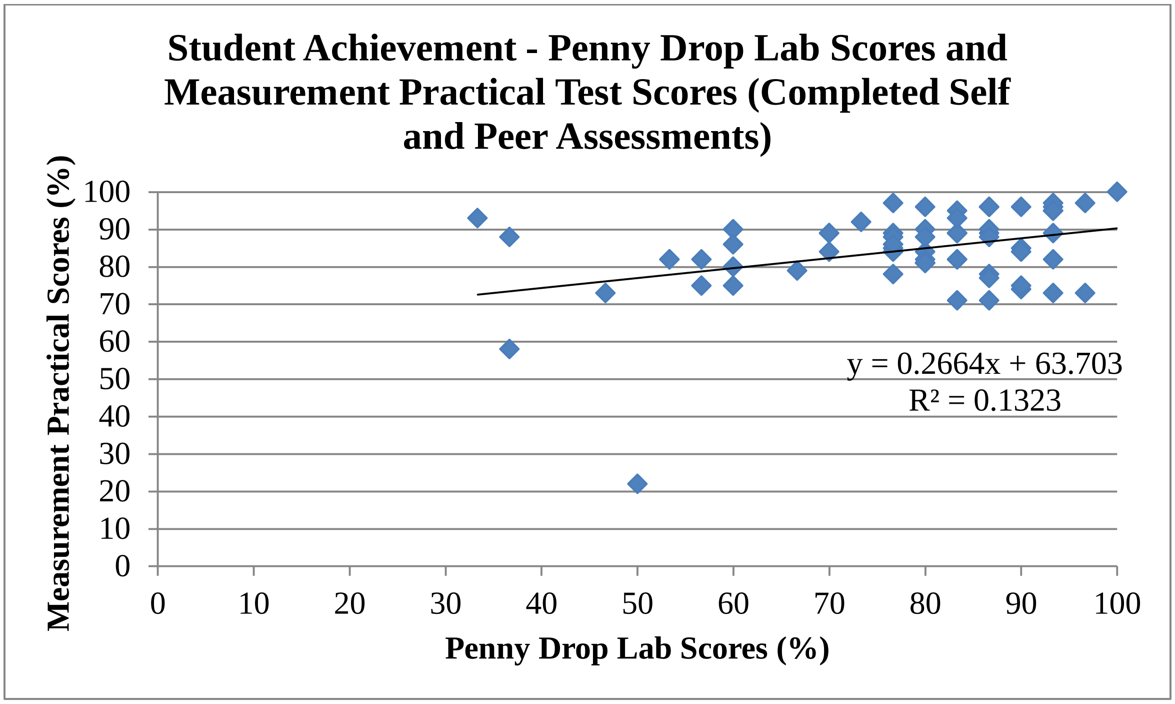

Figure 7. Student achievement showing Penny Drop Lab scores and the associated test for students that completed the self and peer assessments, (*N*= 64).

Students who scored well on the Penny Drop Lab, and completed the self and peer assessments (*N*= 64) generally scored well on the Measurement Practical, but the scores suggest only an extremely weak association, r(62)= 0.36, p=0.003, between the two different types of assessments (Figure 7). While both types of assessments required students to make and record simple measurements, the Penny Drop Lab did not require students to make volume and mass measurements, and required additional skills not found in the Measurement Practical. Those students that did not complete the self and peer assessments (*N*= 23) showed a similar pattern of achievement, but with an even weaker association between their scores, r(21)=0.23, p=0.29 (Figure 8).

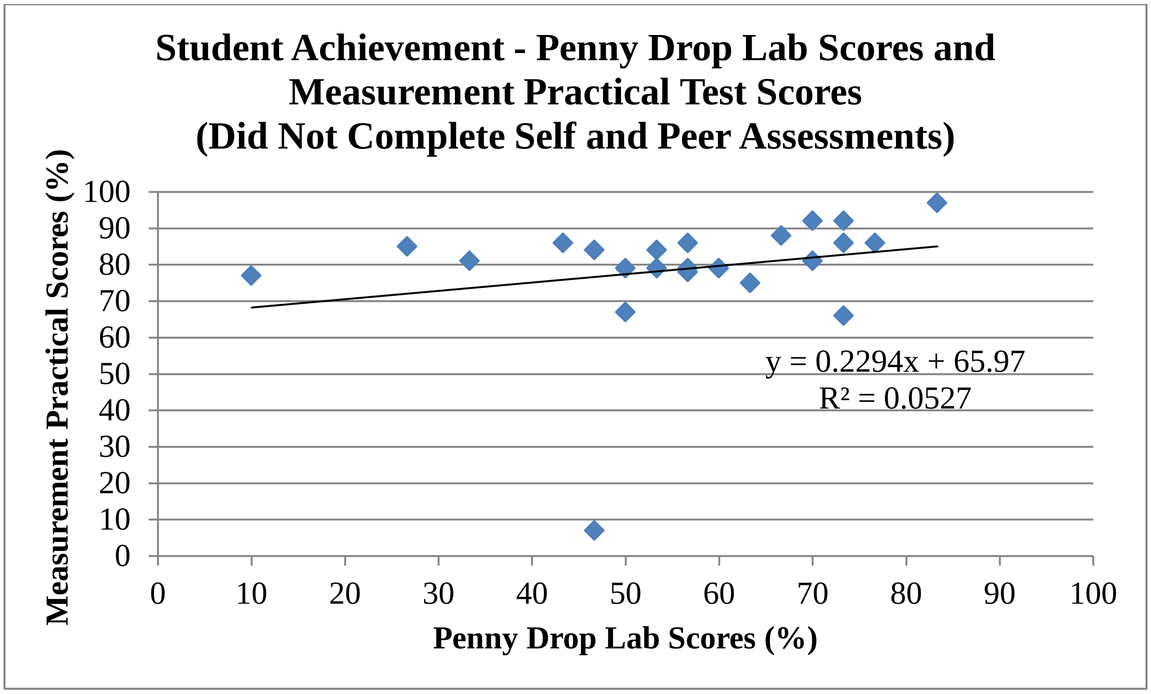

Figure 8. Student achievement showing Penny Drop Lab scores and the associated test for students that had not completed the self and peer assessments, (*N*= 23).

A second lab and test combination was examined in order to determine the impacts of self and peer assessments on tests, where the Gummy Bear Lab and Describing Matter Test were compared. The Gummy Bear Lab required students to make simple measurements and calculate the volume of a solid and density whereas the test required students to understand the concept of density and other terminology used to describe states of matter. While similar in nature, both types of assessments involved different skill sets in order to be successful, and a more intense understanding of concepts beyond simple measurement, even more so than the Penny Drop Lab and Measurement Practical comparison.

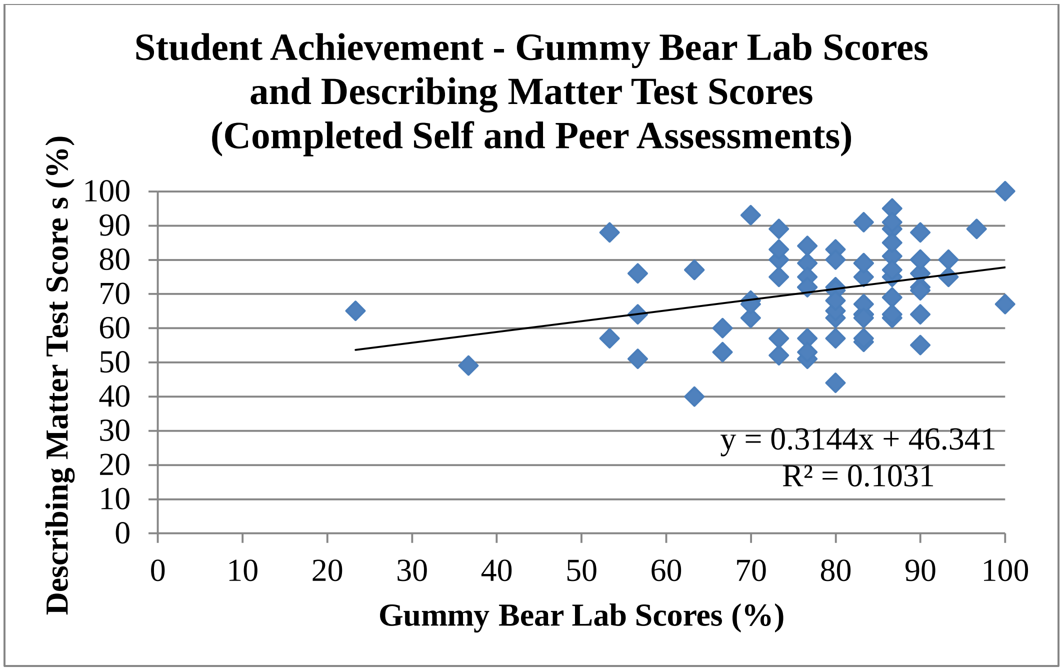

Figure 9. Student achievement showing Gummy Bear Lab scores and the associated test for students that had completed the self and peer assessments, (*N*= 73).

Those students who had completed the self- and peer-assessments for the Gummy Bear Lab (*N*= 73) showed a nearly identical correlation between lab scores and test scores as found for the Penny Drop Lab, r(71)=0.32, p=0.006 (Figure 9). While only a weak correlation, when compared with those students who had not completed the self- and peer-assessments, the self- and peer-assessments appeared to be beneficial for other forms of assessment as well. The few students who had chosen not to complete the self- and peer-assessments (*N*= 9) showed no correlation between “good” lab scores and test scores, r(7)=-0.18, p=0.65 (Figure 10).

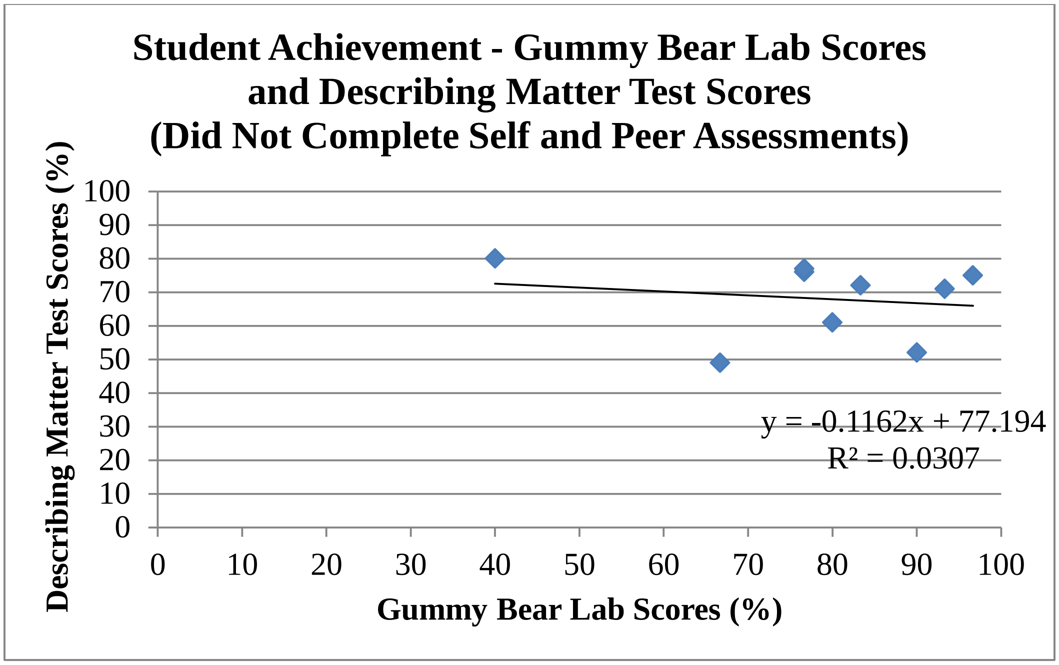

Figure 10. Student achievement showing Gummy Bear Lab scores and the associated test for students that had not completed the self and peer assessments, (*N*= 9).

The comparison of these two pairs of labs and tests indicate that there is little to no connection between the use of self- and peer-assessments for labs and student achievement for other forms of assessments. Upon examining the Penny Drop Lab and the associated test, it appeared as though there was not a substantial difference between students who had chosen to complete the self and peer assessments and those who had not. However, when examining the Gummy Bear Lab and the associated test, which required deeper understanding of concepts in order to be successful, there was a very moderate difference between the two groups of students, indicating that self- and peer-assessments may be influential in improving not only lab scores but also other forms of assessment, such as tests and quizzes.

Student Attitude

After lab papers were submitted and scored by the teacher, they were returned to the students to allow them time to review their score, teacher comments, and how their self and peer assessments matched their lab paper and the teacher's comments. After reviewing, the students were provided with a Lab Survey, which address items associated with attitude and achievement (Appendix B). For each item on the survey, students selected one of four options, Disagree (D), Somewhat Disagree (SD), Somewhat Agree (SA), and Agree (A). Based on the wording of the items in the surveys, student responses were assigned a numerical value in order to determine positive and negative responses and how the percentage of students with a given response changed over the course of the applied treatments. For some items a values were (D) = -2, (SD) = -1, (SA) = 1, and (A) = 2, in order to complete various data analyses and for others, the values were inverted, with (D) = 2.

For example, item number three from the Lab Survey was written "I think the self/peer assessments helped me improve my lab paper" where students who had responded with a "Disagree" were assigned a numerical score of negative two and students who had responded with an "Agree" were given a numerical score of positive two. Item number four was nearly identical as item three, but worded in a negative manner. Student responses of "Disagree" were then assigned a numerical score of positive two and "Agree" with a negative two. This was done in order to assist in analyzing student responses in an efficient manner and provided evidence for sub question number one and supporting evidence for the main question for this AR.

Table 3

Lab Survey Results Showing Percent Positive Responses and Percent Change from the Previous Treatment for Both Honors and Academic Students

			Lab Survey Item										
			1	2	3	4	5	6	7	8	9	10	11
ACADEMIC	1st Treatment	% Positive	60.6%	69.1%	71.3%	74.5%	35.1%	75.3%	37.6%	95.7%	79.6%	75.3%	68.8%
		% Change	-3.2%	4.3%	-3.2%	1.1%	-3.2%	-3.5%	6.8%	3.2%	-1.9%	1.3%	3.9%
	2nd Treatment	% Positive	57.4%	64.9%	74.5%	73.4%	38.3%	78.7%	30.9%	98.9%	77.7%	76.6%	64.9%
		% Change	5.8%	-1.7%	7.8%	-1.0%	10.0%	6.3%	0.2%	-0.1%	-1.8%	-5.3%	6.4%
	3rd Treatment	% Positive	63.2%	63.2%	82.3%	72.4%	48.3%	85.1%	31.0%	98.9%	75.9%	71.3%	71.3%
		% Change	10.3%	11.5%	-0.4%	7.1%	2.3%	6.5%	2.7%	-0.1%	-6.0%	5.8%	4.6%
	4th Treatment	% Positive	73.5%	74.7%	81.9%	79.5%	50.6%	91.5%	33.7%	98.8%	69.9%	77.1%	75.9%
		% Change	-2.1%	0.6%	2.5%	2.3%	-1.3%	-4.5%	6.5%	-2.7%	2.8%	-8.3%	5.9%
	5th Treatment	% Positive	71.4%	75.3%	84.4%	81.8%	49.4%	87.0%	40.3%	96.1%	72.7%	68.8%	81.8%
HONORS	1st Treatment	% Positive	56.5%	69.6%	91.3%	78.3%	69.6%	82.6%	52.2%	100.0%	78.3%	87.0%	69.6%
		% Change	6.0%	22.1%	-8.0%	5.1%	-2.9%	13.2%	-6.3%	0.0%	0.9%	4.7%	9.6%
	2nd Treatment	% Positive	62.5%	91.7%	83.3%	83.3%	66.7%	95.8%	45.8%	100.0%	79.2%	91.7%	79.2%
		% Change	-0.6%	-20.2%	-11.9%	-7.1%	-14.3%	-5.4%	-7.7%	0.0%	-3.0%	-1.2%	-3.0%
	3rd Treatment	% Positive	61.9%	71.4%	71.4%	76.2%	52.4%	90.5%	38.1%	100.0%	76.2%	90.5%	76.2%
		% Change	6.3%	14.9%	14.9%	14.7%	-2.4%	5.0%	7.4%	0.0%	5.6%	5.0%	10.2%
	4th Treatment	% Positive	68.2%	86.4%	86.4%	90.9%	50.0%	95.5%	45.5%	100.0%	81.8%	95.5%	86.4%
		% Change	20.1%	2.5%	-3.0%	-7.6%	5.6%	4.5%	4.5%	0.0%	-9.6%	-6.6%	-3.0%
	5th Treatment	% Positive	88.2%	88.9%	83.3%	83.3%	55.6%	100.0%	50.0%	100.0%	72.2%	88.9%	83.3%

positive change
negative change
no change

Note. 1st Treatment=*N*= 117, 2nd Treatment=*N*= 118, 3rd Treatment=*N*=108, 4th Treatment=*N*=105, 5th Treatment=*N*=95).

The frequency of student responses was determined for each survey item, which was then used to calculate the percentage of participating students had selected a particular response (Appendix D). Student responses were then determined to be either positive or negative based on the wording of each item on the survey. This was done for each of the five treatments, with the first treatment being considered baseline data for comparison with the following treatments.

In order to determine which types of assessment may have had the largest impact on student scores, items number one and two were examined. The design of the survey allowed for students to choose responses which would indicate that both were helpful, neither were helpful, or only one was helpful. Over the course of five treatments, student perceptions of usefulness varied, but did increase for most treatments (Table 3). For both subgroups, self-assessments showed a greater increase in usefulness among students, with honors students (+31.7%) showing a much larger increase than academic students (+10.8%). Honors students (+19.3) also showed a larger increase than academic students (+6.2%) in regards to the usefulness of peer-assessments. After all five treatments were completed; nearly 75% of academic students, with a slightly higher percentage of honors students, provided responses indicating both the self- and peer-assessments were useful.

Varying changes in student perceptions of the usefulness of the assessments may be influenced by other factors, such as the nature and difficulty of the lab, time available, and student pressure to complete overlapping tasks and assignments for other classes. Several student comments during research process may have shed some light on the reasoning for the changing responses, with multiple students indicating they did not like the extra work required to complete the self- and peer-assessments. Student responses suggest this was not the case for every treatment, and changing responses were more common with the academic subgroup. With the data collected, it is not possible to determine the specific reasons for the changes in responses, but it could suggest that some labs were more difficult than others, and students felt overwhelmed with the additional work load of completing the self- and peer-assessments.

At times students may not have appreciated the additional work of completing the self- and peer-assessments after finishing a lab, but many students did not seem to think about them before the lab. Items five and seven from the survey addressed student thoughts about the treatments before the treatments themselves are completed (see Table 3). Student responses indicate that a majority of academic students (>50%), and a large portion of honors students (55.6%) did not think about the checklist style assessments during labs. This would indicate that even though students knew they would be completing the assessments, they did not think about the lab requirements while they were working. This could suggest that the assessments were used by students as a last chance method to double check lab papers before they are submitted for grading, as opposed to a proactive guide to improve lab paper quality. Item number seven verified the results from item five, with student responses differing only slightly (<10%). Over the course of five treatments, the academic subgroups showed slight increases in positive responses (14.3%), which indicates that some students increasingly thought about the assessments while they were completing the labs as opposed to after completion, and suggests that students gradually recognized the assistance the assessments provided throughout the course of the applied treatments. However, the honors subgroup started with much higher percentages (69.6%) of positive responses, and slowly decreased throughout the treatments (-14%). This could suggest that honors students were confident they could successfully complete the assigned lab without the use of the self- and peer-assessments.

Other items from the survey, such as item number six, can be used to infer about student confidence, such as a students' comfort level with their peers reviewing their completed labs (see Table 3). Responses showed some variation over the course of all treatments, with an overall increasing trend, suggesting that students became increasingly more comfortable having their peers evaluate their lab papers. The academic subgroup increased less (11.7%) than the honors subgroup (17.4%) which may suggest that the honors students were more comfortable due to a higher confidence and ability level in comparison with their peers.

Even if students were not thinking about their self- and peer-assessments before their labs, or were not comfortable with others reviewing their work, a majority of students (70-80%) responded that they had put forth more effort into their lab paper due to being aware that other students would be reviewing their lab papers (see Table 3). These values fluctuated throughout the applied treatments, but did not stray much from the 70% range. Even if the self- and peer-assessments weren't directly responsible for improved lab scores, the additional effort alone may account for increased student achievement.

In order to show a relationship between student thoughts towards the self- and peer-assessments and lab scores, the data were plotted using lines for each treatment (see Figure 11). Using responses from the lab surveys, students could select from four options, ranging from negative to positive, with no neutral response available. Each treatment was plotted as its own line, with the first treatment being the Penny Drop Lab (PDL), the second treatment being the Float Your Boat Lab (FYB), third treatment being the Gummy Bear Lab (GBL), fourth being the Separating A Mixture Lab (SAM), and the fifth and final treatment being the Rate of Solution Lab (RSL).

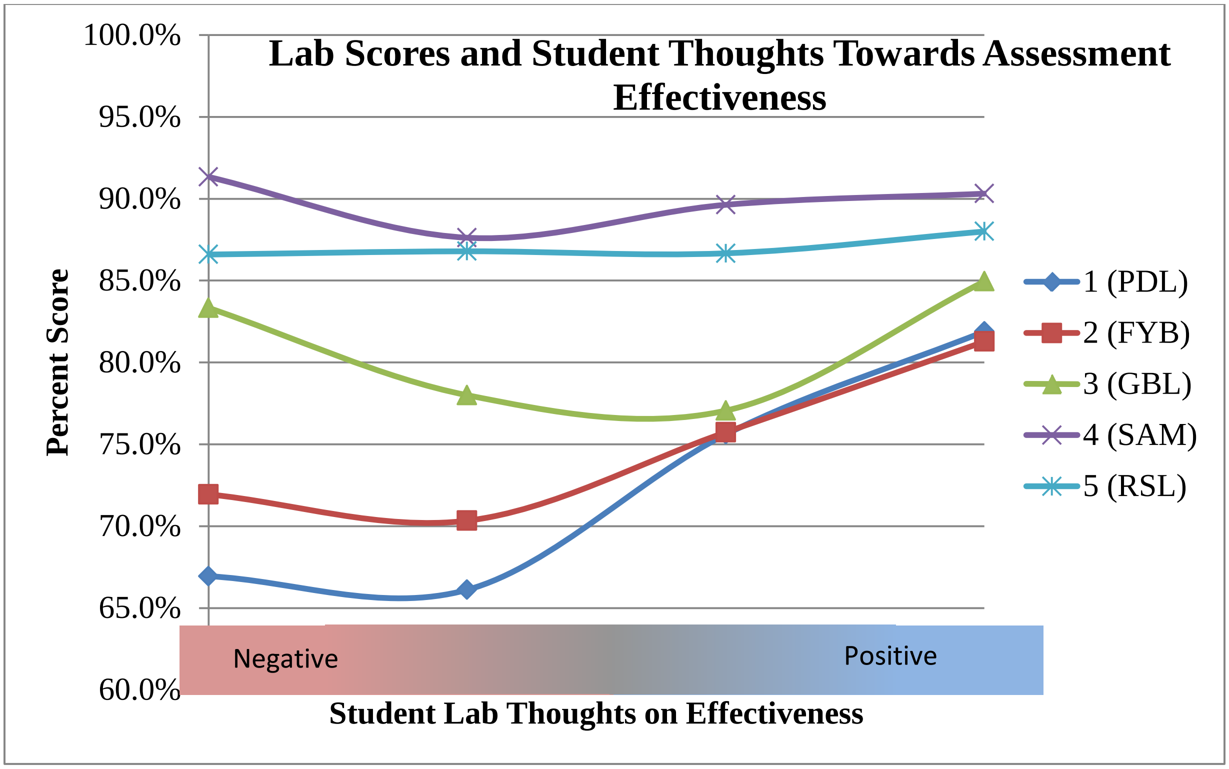

Figure 11. Mean lab scores for all students that were offered self and peer assessments, based on student responses evaluating the effectiveness of self and peer assessments.

Two items from the survey (item 3 and 4) addressed attitude by determining students' thoughts towards the effectiveness of the self and peer assessment (Table 3). Over the course of the applied treatments, students that thought the assessments were useful scored higher on labs during the first and second treatments and eventually leveled out over the course of the remaining treatments (Figure 11). When including every student that was offered self- and peer-assessments, those students that had chosen not to complete them typically had a negative outlook regarding their usefulness, and had substantially lower mean scores (58.3%) when examining the first treatment application.

Students who had a negative outlook, and completed the assessments had mean scores identical to their positive outlook peers (81.7%) and nearly all negative outlook students showed improvement (+21.3%) from the baseline scores for the first treatment. Students with a negative outlook students that had completed the self- and peer-assessment made up only a small portion of all students sampled (*N*= 11), yet showed great improvement. During the same treatment, students with a positive outlook showed less improvement (+6.8%). This could indicate that students who chose not to use the self- and peer-assessments may have done so due to a pre-existing negative connotation to completing similar tasks, which were in a few cases described as “unnecessary”, “a waste of time”, and just as “extra work”, according to responses provided by students, and suggests that a negative outlook does not necessarily result in less improvement. This same trend continued throughout the treatments, with a gradually decreasing number of negative responses from academic students. Honors students did not necessarily follow the same trend, and became increasingly negative towards the self- and peer-assessments throughout the treatments, suggesting they may not be needed or as useful as for academic students.

Some students may have provided responses suggesting they did not think the self- and peer-assessments were useful, which is supported by item number ten from the lab survey. One quarter (25%) of academic students believed the assessments were a waste of time, with even fewer honors students sharing the same belief. Across each group of students, and during each treatment, the values fluctuated but did not stray much from the baseline data. Over the course of all applied treatments the academic subgroup showed a slight increase (6.5%) in the number of students that thought the assessments were a waste of time, despite the improvement of lab scores. Until the final treatment, the percentage had remained fairly consistent, indicating that more students had given up on the assessments after the long holiday/winter break. The honors subgroup showed a similar trend, but had a smaller increase (<2%) of students that thought the assessments were a waste of time. The percentage also changed drastically during the last treatment, indicating even honors students had increasingly given up on the assessments after the long break.

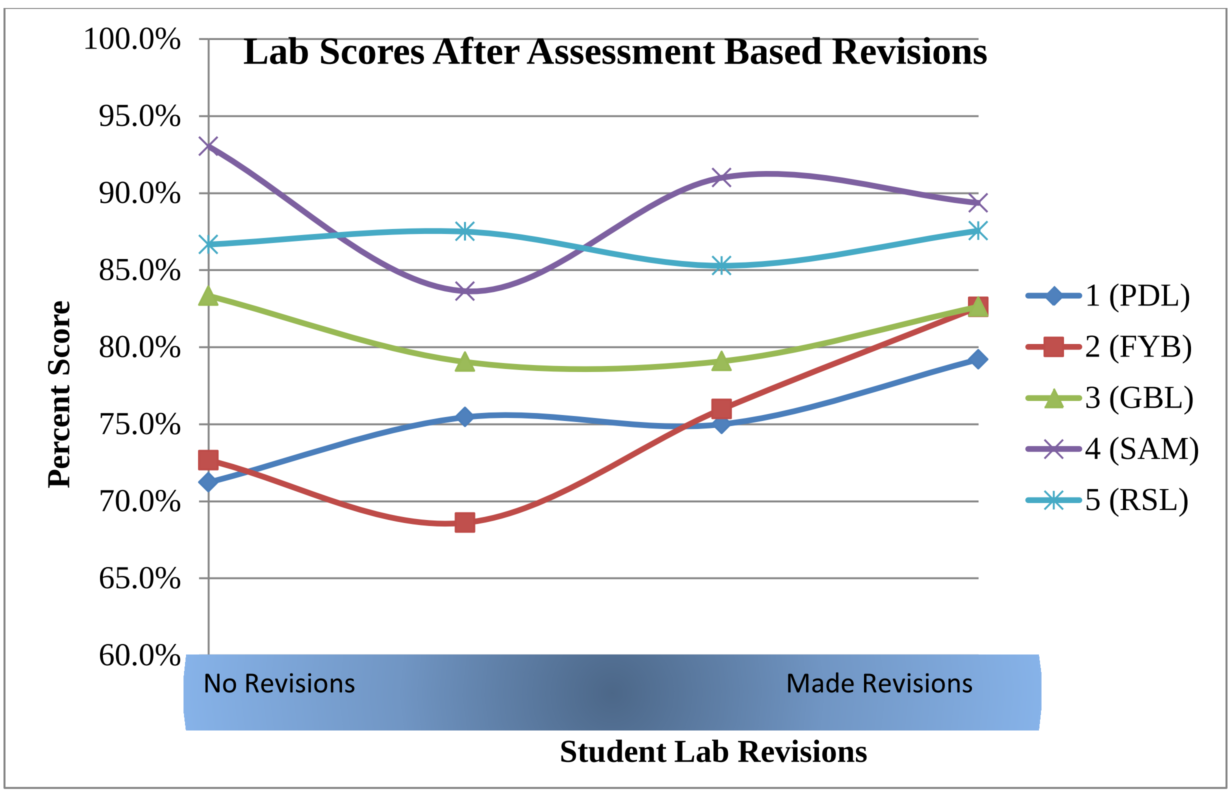

Figure 12. Average lab scores for all students based on student responses indicating revisions to lab papers had been made according to the applied assessments.

Despite some students having a negative outlook, others choosing not to complete the assessment, and varying numbers of students completing each treatment, lab scores improved, especially when viewing the first two treatments. One factor analyzed using surveys was determining if students used the self- and peer-assessments to revise their lab paper before submitting for teacher review. The results from item number eleven suggest that students' lab scores were influenced by making revisions to their lab papers during the first two treatments (Figure 12). For each of the remaining treatments, student lab scores became more similar, with those students that had not made revisions receiving similar scores as those that did make revisions.

Over the course of all treatments, both academic and honors students responded that a majority of students used the assessments to make revisions for the first treatment (69%), and both groups increased in percentage of students responding favorably. Academic students showed a gradual increase during every treatment (+12.8%), showing that more students had utilized the assessments as a way to identify areas that should be completed or revised. Honors students showed slightly greater increases in the percentage of students using the self and peer assessments as a method to identify areas for revision, yet the percentage fluctuated throughout the treatments, suggesting they not have needed the self and peer assessments for some applications. Overall, students making revisions to their lab papers showed increased lab scores for the beginning treatments, when the format and requirements of a completed lab were very unfamiliar to students. This may suggest that the self- and peer-assessments provided students the opportunity to learn the expectations of a completed lab paper without the use of teacher provided input.

The data suggests that many students used the assessments to their advantage, but viewed them as just another part of the lab that had to be completed, and often doubted the potential benefit. Even students that did not make revisions on the lab paper based on the results of the assessments, showed improvements in their lab scores eventually, which may lead to implications about the timing of the applications, and the need for self- and peer-assessments for all labs. Despite the variation in the data, the first two treatments clearly show the intended and expected results, to expedite student achievement and to provide a set of guidelines for completing labs. Many students used the assessments in such a way, and for those that had attempted the assessments and made revisions, their scores were noticeably higher than their classmates that hadn't. After the second treatment, students seemed to have had a separation from the assessments, with some still relying on the checklist to improve their lab scores, while others appeared to have felt more comfortable with the lab process and more confident with their work, which also resulted in improved lab scores.

Student Confidence

One purpose of this Action Research was to determine if the use of self- and peer-assessments influenced student confidence. This was evaluated in order to determine if the use of the assessments would be useful for future plans to initiate a science fair based performance assessment to replace the existing final exam for the course. Student confidence surveys (Appendix C) were given before all treatments, immediately after the third treatment, and after all treatments were applied. Due to varying *N* values, mean student confidence was determined in order to gain perspective on the confidence of large groups of students.

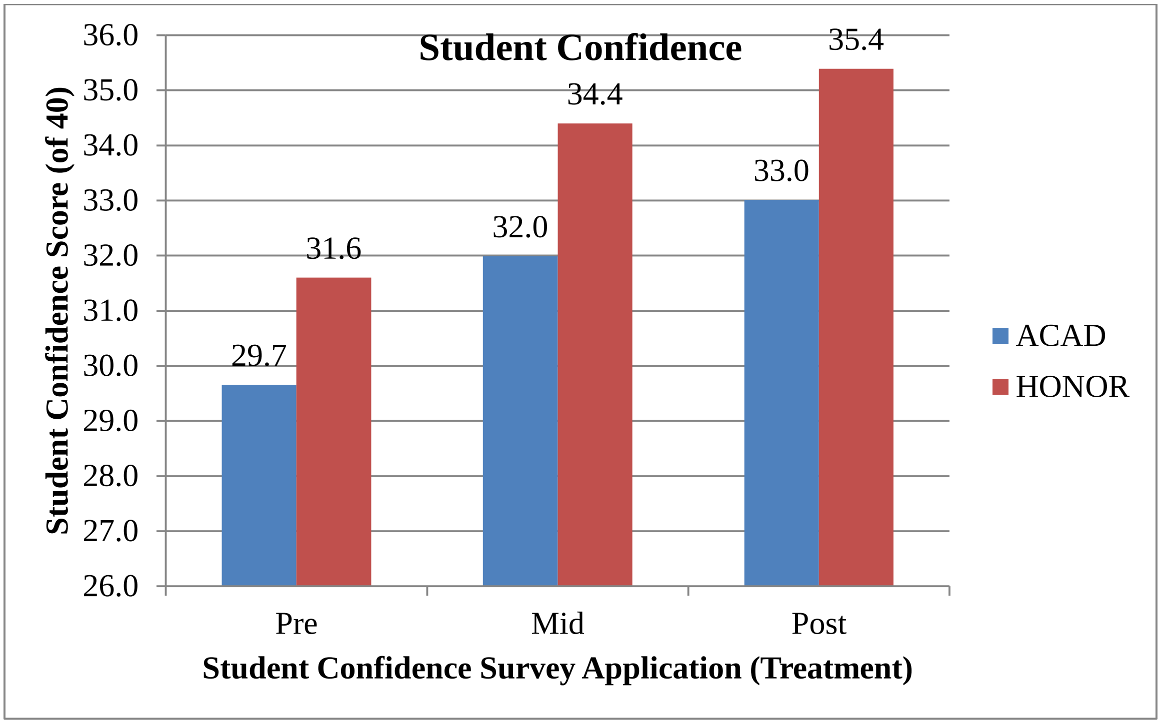

Figure 13. Student confidence throughout the applied treatments (scored out of 40).

After completing the baseline lab (Paper Plane Lab) students completed the first lab confidence survey (Figure 13), which resulted in honors students having a mean confidence score higher than (31.6) academic students (29.7). This may in part be due to higher achievement in previous activities/classes and increased ability in comparison with their academic counterparts. Throughout each application of the survey, both subgroups showed increases in their confidence level, with honors students always slightly more confident than their peers. Overall, lab scores increased during each treatment, the spread of student scores decreased for nearly every treatment, and a majority of students were using the self- and peer-assessments to their advantage in some manner. Students that typically scored well, continued to score well, and students that had a rough start and scored poorly, managed to improve their scores.

It is not entirely clear how or if the self- and peer-assessments directly influenced student confidence, as all students were included in each treatment, however, it can be assumed that the assessments did not have a negative influence on confidence. While not used in direct correlation with a science fair project, this action research addressed qualities of a well done lab paper and may transfer to a similarly designed project, where self- and/or peer-assessments could be utilized, and could feature a checklist design where students could identify areas needing revision.

INTERPRETATION AND CONCLUSION

As can be seen in the figures and tables above, student achievement, attitude, and confidence changed over the course of the applied treatments. Some results were unexpected, but ultimately, this Action Research was a success. In order to determine the true impacts of the use of self and peer assessments, each Action Research question was addressed separately.

Student Achievement

In response to my main Action Research question, "How does the use of self- and peer-lab-assessments affect student lab achievement?" lab scores were analyzed using multiple statistical methods. Scores clearly improved and the difference between those students using and completing the self- and peer-assessments was significant for the first few treatments. It would seem as though the assessments were successful in expediting the process of improving student lab scores, as those students who did not complete them required several labs to achieve scores comparable to their classmates.

Self- and peer-assessments assisted in expediting improvement in student lab scores, and may have also slightly improved student test scores when tests required students to have a deeper understanding of classroom concepts (R^2 = 0.10 for students who did completes assessments, R^2 = 0.03 for students who did not complete assessments). Academic students' scores improved more than twice as much as honors students' scores throughout the applied treatments; however, there were some notable discrepancies between the lab scores and student responses. Not only did students' lab scores improve, but the spread of scores decreased when comparing the baseline and second treatment (93%) with the fourth (40%) and fifth (46%), indicating that the self- and peer-assessments helped increase the lower scoring students by allowing them a last chance method to review their work before submitting it for grading. Scores on lab papers for honors students did not show a drastic change, but did show a slight improvement (15%).

By comparing student lab scores with closely associated test scores, there is a very weak to no correlation between the use of self- and peer-assessments on labs and increasing test scores. While this would need additional sampling and analysis to fully address, this may indicate that students are using higher level thinking skills to analyze their own work and are may be applying these skills to other forms of assessment. This appeared to be truer when more difficult concepts and connections between material and skills are required, such as those needed for the Gummy Bear Lab and the Describing Matter test.

Student Attitude

In response to my Action Research question, “What are the affects of self- and peer-lab-assessments on student attitude?” it appeared as though students had a love-hate relationship with the self- and peer-assessments. Multiple honors students provided comments about their dislike of the self- and peer-assessments throughout the process, such as, “I think these self/peer-assessments need to have different questions designed for each specific lab. It gets boring reading the same questions over and over again. It has helped me in labs though to get a better grade” (Student 77). Honors students may have been bored by the redundancy of the self- and peer-assessments, which may have some affect on the lab scores and student responses about the dislike of self- and peer-assessments. This would be difficult to determine without completely restructuring the entire process used for this Action Research.

Academic students also provided comments throughout the Action Research process, although they didn't represent the same concerns of boredom as several honors students had indicated. Instead, when academic students provided negative comments during surveys and/or interviews, negative comments were often describing the amount of work involved with the self- and peer-assessments. On multiple occasions, students acknowledged the benefits of the self- and peer-assessments, yet still disliked using them because they required an additional twenty minutes of effort to review their lab papers. One interviewee commented, "At first I thought they were not needed, useless, a waste of time, but as labs got harder the self- and peer-assessments helped me" (Student 85). Other students may not have provided feedback specifically stating they did not like the extra effort required, but did provide comments such as, "They are frustrating at times, but help you to increase your lab grade" (Student 26), indicating they were frustrated by the extra effort, or possibly the boredom of having a universal checklist style rubric.

Overall, most students agreed that the self- and peer-assessments were beneficial (67%), according to the final survey, with the remaining students nearly equally divided among being unsure if they were helpful, or sure that they were unhelpful. Throughout the process, students supported this notion by providing comments on surveys and interviews, such as, "the self/peer-assessments helped me understand what I needed to improve in my labs" (Student 66). Another student commented, "I like when my peers look over my work because they give me an honest opinion to what needs fixed and what doesn't need fixed. Most of the time when my peers look it over and when I make corrections I get a better grade" (Student 44).

Student #44's feedback demonstrated a key point in this AR when she used the terms "most of the time." Even with structured surveys, and labs that have proven to be successes in the past, student scores can fluctuate, and attitude towards any one survey or individual survey item could vary immensely. Other statistical analyses could certainly have been completed, but the focus for this AR was improving the lab process for the use with large quantities of students.

A few students did not believe self- and peer-assessments were effective in improving their labs, and it may be unclear to the specific reasons why, although one student (#13) commented, "I don't think they help whatsoever. Most people just check random boxes and don't do it truthfully therefore you get no help at all." This could very well be true to some extent, although without a thorough investigation, it may be impossible to estimate the number of occurrences in this Action Research. Despite student #13's comments and concerns about the process, her lab scores rose from a 50% during the baseline lab to a 93% on both the fourth and fifth treatments. This may suggest that even if a student does not truthfully complete the self- and peer-assessments, the process may still result in improved student achievement when used as a "last chance" to make revisions to their lab papers before submitting for teacher grading.

One student (#25) observed similar trends, and provided feedback that supported this idea; "They are a good 'second chance' tool to use as a last-minute way to correct mistakes." Other students also commented on the honesty of their peers, as if they had observed their classmates intentionally checking the same box without actually reviewing their lab papers, as mentioned by two frustrated students. Student (#42) commented, “The self- and peer-assessments help me, but the peers just need to do it accurately and take their time,” while a classmate showed similar frustration, trailing off on her response during an interview; “They’re a great idea for the people who actually use them correctly…” (Student 15).

Nearly every student (>95%) put forth more effort to some extent, when aware that others would be reviewing their work. This alone could account for score increases, which may show that self- and peer-assessments were still effective at improving student achievement. Most students were not making corrections to their lab papers ($\leq$25%), yet scores improved during each treatment. With most students not making corrections on their lab paper, student scores very well may be due to students applying additional effort. When students did make revisions, a majority of students (75%) recognized the improvement they were making to their lab paper. This could assist in redirecting focus for students to learn how and when to make corrections and revisions, which may in turn influence student attitude and achievement.

During the early stages of the applied treatments, few students expected the self and peer assessments to be useful (8.3%), but after five applied treatments, most students agreed that they were useful (67.6%) with about half of the remaining students still unsure about how the assessments helped them improve their lab papers. A small percentage of all students believed they were not helpful at improving their lab scores (16.7%). Nearly one third of all students provided evidence that they did not notice an improvement in their lab scores, yet student achievement data suggests most students showed an improvement most of the time.

According to the data collected, the increase in lab scores is not a result of students thinking about their self- and peer-assessment checklist before or during a lab. The students were aware that the self- and peer-assessment would occur after each lab, but didn't think about the requirements before or during a lab.

Student Confidence

In response to my Action Research question, "How do self and peer lab assessments affect student confidence towards designing their own experiments?", student confidence gradually increased among all students, with honors and academic students showing nearly identical amounts of increase. While a science fair style project where students design and complete their own experiments was not used with this group of students, it seems very likely that the use of self- and peer-assessments would be very beneficial for students to recognize the qualities of a thorough experiment. Semi-structured labs would not be as detailed as an independent project; the same components would be included, suggesting that students can gain valuable knowledge of how experiments can be done in a physical science classroom setting.

Future Implications

In response to my Action Research question, "How will the use of self- and peer-lab-assessments influence my teaching strategies in the future?" the data collected to describe the influences on students' scores, attitude and confidence was considered.

Based on the lab scores, survey responses, and implications of the data collected, the use of self- and peer-assessments may become the standard in my classroom. By gaining insight on students' outlook and performance, self- and peer-assessments were proven to be useful in a middle school science laboratory setting, but raised an additional question, "What could be done to improve the use of self- and peer-assessments in a laboratory setting?" Some changes may be needed to improve the effectiveness of the assessments, but the intent will remain the same.

According to student responses, the ideal self- and peer-assessment would be optional and more specific to the lab's requirements. Students seemed to recognize the universal nature and some soon became bored with the process. It may be more useful to slightly adapt the universal process to each lab, without losing the problem solving integrity of the lab itself. Some students severely disliked and resisted the additional workload of the self- and peer-assessments, which could imply the making them optional may be more productive. Many students may choose not to use them if optional, but if I were to provide evidence to the students that they are useful, most students would be interested in doing what they can to improve their grade. This would also promote the use of higher level thinking skills for a large number of students at the same time, and hold students more accountable for their assignments, as they will have opportunities to improve before submitting their labs.

Helping students find their own mistakes is one strength of the self-assessments, but it doesn't necessarily teach the students how to make the necessary corrections. A small percentage of students (8%) requested more information about how to make corrections after identifying areas for revision, and only 15% of students provided responses that they made revisions to a lab paper, yet weren't sure why. Even without being extensively trained to make corrections, self-assessments have still shown to improve student achievement throughout several studies (Ross, 2006). In the future, it is likely that a more detailed self-assessment checklist will be available for each structured lab, as well as completed examples of a lab paper from formerly used lab activities in my classroom. This should not only provide an example of a well completed lab paper, but provide students with more information about the expectations and requirements.

VALUE

When providing students with a traditional rubric in the past, my eighth grade students rarely used them to evaluate their own work. The use of a checklist style rubric provided a less intimidating, easy to use format, to which students appeared to be attracted to throughout the course of the applied treatments. Even if students were not interested in improving the quality of their work, many were interested in improving their grades, with more quality work following hand-in-hand. Using the data from this Action Research, I plan on continuing to offer checklist style rubrics for labs, and will restructure preexisting rubrics for more student friendly interaction with their own work.

In order to improve the checklists used, I will be adapting them to be more specific to each lab, as not every lab requires exactly the same skills and concepts to be applied. A universal checklist may be too good to be true, but the modification of the checklist would not be too time consuming from my end, and would provide students more details about the requirements, without removing the idea of having students evaluate and revise their work before submitting it for teacher review.

Self- and peer-assessments received mixed reviews throughout the process, with most students eventually recognizing that they can lead to improved quality by providing a last chance effort to catch mistakes and missing portions/items. These will be made available for all labs in the future, but will not be required, as there were definitely students that were able to carefully read and follow instructions, complete the tasks at hand, and submit quality lab papers without their use. This will place accountability on the students, depending on their choices and participation in lab activities.

Many students were able to improve their lab scores without the use of self- and peer-assessments, but did require a few additional labs to catch up to their peers which had completed the assessments. Students also recognized that the self and peer assessment checklists were more beneficial for some labs than others, which reinforces my ideas of making them available, but optional for each lab.

By using the data collected I was able to learn quite a bit about the mindset of my students when completing labs. Most students want to complete their tasks as quickly as possible, as though the assignments were a race. By using the checklists, students had the opportunity to change gears, slow down, and review their work. While some students took shortcuts, and will likely continue to take shortcuts to try to cheat the system, once students recognized the benefit, most were on board to at least attempt a task they believed would lead to improving their grades.

Because of the additional review of their own work, and the work of their peers, students quickly recognized the qualities of a well done assignment, improved their lab scores, and may have had effects that spilled over onto other forms of assessment. The checklist style rubrics and self- and peer-assessments have found a permanent home in my eighth grade classroom, and hopefully my colleagues' as well.

REFERENCES CITED

Acker, M. (2003). *Does Peer Review Improve Lab Report Quality in High School Science Students?* Unpublished Manuscript. Retrieved May 20, 2012from http://etd.lib. montana.edu/etd/ 2011/acker/AckerM0811.pdf.

Andrade, H. (1997). Understanding Rubrics. *Educational Leadership*, *54*(4).

Angelo, T. A. Cross, P. K. (1993). *Classroom Assessment Techniques: A Handbook for College Teachers*. San Francisco, CA: Jossey-Bass.

Black, P.. Harrison, C. (2001) Self- and Peer-Assesment and Taking Responsibility: the Science Student's Role in Formative Assessment, *School ScienceReview*, *83*(302).

Bostock, S. (2000). Student Peer Assessment, *The Higher Education Academy*. Retrieved October 17, 2011 from http://www.heacademy.ac.uk/assets/documents/ resources/resourcedatabase/id422_student_peer_assessment.pdf.

Brown, S. (2004).Assessment for Learning. Learning and Teaching in Higher Education. Issue 1, pp. 81-89.

Clarware LLC. (2012). Dover Area Intermediate School. Retrieved May 25, 2012 from http://www.schooldigger.com/go/PA/schools/0768006283/school.aspx

Doran, R. Chan, F. Tamir, P. & Lenhardt, C. (2002). *Science Educator's Guide to Laboratory Assessment.*Arlington, VA: NSTA Press.

Gaillet, L.L. (1994). A historical perspective on collaborative learning. *Journal of Advanced Composition*, *14*(1), 93-110.

Macdonald, A. (2004).Student Self-evaluation of coursework assignments: a route to better conception of quality. *Learning and Teaching in Higher Education*. *2004-05*(1), pp. 102-107.

Mullen, Y. (2003). *Student Improvement in Middle School Science*. Unpublished Manuscript. Retrieved May 20, 2012 from http://www.eric.ed.gov/PDFS/ ED477845.pdf

Nilson, L.B. (2003). Improving student peer feedback. *College Teaching, 51*(1), 34-38.

Ross, J. A.(2006). The Reliability, Validity, and Utility of Self Assessment. *Practical Assessment, Research and Evaluation,* volume 11 (10). Retrieved October 24, 2011 from http://www.springerlink.com/content/h743345073237g38/.

Sadler, R. A. (1998). Formative Assessment: Revisiting the Territory. *Assessment in Education*, 5(1).

APPENDICES

APPENDIX A

Lab Self and Peer Assessments

Appendix A: Self and Peer Lab Assessments.

LAB SELF ASSESSMENT

Name: ____________________________

Lab Title: __________________________

D = Disagree
SD = Somewhat Disagree
SA = Somewhat Agree
A = Agree

	D	SD	SA	A
The hypothesis is clearly written	☐	☐	☐	☐
Hypothesis is written as a statement	☐	☐	☐	☐
Hypothesis is testable	☐	☐	☐	☐
Hypothesis answers the question(s) asked	☐	☐	☐	☐

Data is recorded in a table format	☐	☐	☐	☐
All numerical data is recorded in the same manner	☐	☐	☐	☐
Labels are included with all data recorded	☐	☐	☐	☐

Calculations are shown (show work)	☐	☐	☐	☐
Graph is neat (ruler was used)	☐	☐	☐	☐
Data is accurately plotted	☐	☐	☐	☐
X axis is labeled (including units)	☐	☐	☐	☐
Y axis is labeled (including units)	☐	☐	☐	☐
Graph has a descriptive title	☐	☐	☐	☐

Hypothesis is accepted/rejected	☐	☐	☐	☐
Results explained using data	☐	☐	☐	☐
Other discussion questions answered appropriately	☐	☐	☐	☐

D = Disagree
SD = Somewhat Disagree
SA = Somewhat Agree
A = Agree

LAB PEER ASSESSMENT

Assessment Partner Name: ______________________________
Lab Title: ___________________________

A

	D	SD	SA	
The hypothesis is clearly written	☐	☐	☐	☐
Hypothesis is written as a statement	☐	☐	☐	☐
Hypothesis is testable	☐	☐	☐	☐
Hypothesis answers the question(s) asked	☐	☐	☐	☐

Data is recorded in a table format	☐	☐	☐	☐
All numerical data is recorded in the same manner	☐	☐	☐	☐
Labels are included with all data recorded	☐	☐	☐	☐

Calculations are shown (show work)	☐	☐	☐	☐
Graph is neat (ruler was used)	☐	☐	☐	☐
Data is accurately plotted	☐	☐	☐	☐
X axis is labeled (including units)	☐	☐	☐	☐
Y axis is labeled (including units)	☐	☐	☐	☐
Graph has a descriptive title	☐	☐	☐	☐

Hypothesis is accepted/rejected	☐	☐	☐	☐
Results explained using data	☐	☐	☐	☐
Other discussion questions answered appropriately	☐	☐	☐	☐

APPENDIX B

LAB SURVEY

Appendix B: Lab Survey.

Participation is voluntary, and you can choose to not answer any question that you do not want to answer, and you may stop at anytime.

Your participation or non-participation will not affect your grade or class standing.

Name: ______________________________

LAB SURVEY – Self and Peer Assessments

D = Disagree
SD = Somewhat Disagree
SA = Somewhat Agree
A = Agree

Complete after submitting your lab for grading. Please be honest!

	D	SD	SA	A
1. The self assessment was the most useful for this lab.	○	○	○	○
2. Peer assessment was the most useful for this lab.	○	○	○	○
3. I think self/peer assessments helped me improve my lab papers.	○	○	○	○
4. Self/Peer assessments DID NOT help me improve my lab papers.	○	○	○	○
5. I think about the self/peer assessments before I complete my lab.	○	○	○	○
6. I feel comfortable when my lab papers are reviewed by someone else.	○	○	○	○
7. I don't think about the assessments until I have to complete one.	○	○	○	○
8. I am honest when I assess others' work.	○	○	○	○
9. I put forth more effort knowing that others will review my lab work.	○	○	○	○
10. Self and Peer assessments are a waste of time.	○	○	○	○
11. I made corrections to my lab paper after the Self and Peer assessment	○	○	○	○

APPENDIX C

CONFIDENCE SURVEY

Appendix C: Confidence Survey.

Participation is voluntary, and you can choose to not answer any question that you do not want to answer, and you can stop at anytime.

Your participation or non-participation will not [illegible] g.

Name: ___________________________

1 = Not Confident At All
2 = Somewhat Confident
3 = Confident
4 = Very Confident

LAB CONFIDENCE SURVEY

How confident do you feel about your ability to…?	1	2	3	4
Ask questions	○	○	○	○
Do research	○	○	○	○
Write a good hypothesis	○	○	○	○
Make and record measurements	○	○	○	○
Make and record observations	○	○	○	○
Organize data in a table	○	○	○	○
Make accurate, correct, and detailed graphs	○	○	○	○
Explain results by using my data collected	○	○	○	○
Share results with others	○	○	○	○
Design and complete an experiment on my own	○	○	○	○

Would you be worried about designing your own experiment? Explain what you would be most concerned about.

Of the items listed above, which are you most comfortable doing?

APPENDIX D

LAB SURVEY ITEM DESCRIPTION

Appendix D: Description of each of the lab survey items.

Item Number	Item Description
1	The self assessment was useful.
2	The peer assessment was useful.
3	I think self/peer assessments helped me improve my lab paper.
4	Self/Peer assessments did not help me improve my lab paper.
5	I think about self/peer assessments before I complete my lab.
6	I feel comfortable when my lab papers are reviewed by someone else.
7	I don’t think about the assessments until I have to complete one.
8	I am honest when I assess others’ work.
9	I put forth more effort knowing that others will review my lab work.
10	Self and peer assessments are a waste of time.
11	I made corrections to my lab paper after completing the self and peer assessments.

APPENDIX E

PAPER AIRPLANE LAB

Appendix E: The Paper Airplane Lab, used as a baseline for treatment comparison.
Name: ________________________________ Date: _________ Pd: ___________

PAPER AIRPLANE LAB

PRE-LAB NOTES:

1.____

____

2. ____

____

3. ____

____

PROBLEM: In this lab we will be practicing scientific lab skills, such as hypothesizing, following directions, recording data, measuring, creating graphs and tables, and analyzing your data to come to a conclusion. We will use these skills to try to answer two questions.

Q1: How far will a paper airplane fly when made out of different materials?
Q2: How will increased mass change the distance the plane flies?

RESEARCH: The history of paper airplanes probably goes back to the first pages of papyrus in Ancient Egypt which were thrown at the trash can. The first flying devices to use paper were kites in China about 2000 years ago. Early hot air balloons, such as the first balloons made by the Montgofier Brothers in France in the late 1700's, also used paper in their construction. Their early model balloons were all paper, and their first human-carrying balloons of 1783 were cloth lined with paper.

The origin of paper airplanes is somewhat a mystery. Some references dated the first paper airplanes to the 1930's. Early paper airplanes were often made to study aerodynamics to help develop early airplane designs. During World War II, materials to make toys were in short supply and paper filled the need. Most toys were models that were assembled and then played with. (http://www.paperplane.org/History/history.html)

What materials are needed to answer our questions?

- Printer Paper	- Construction Paper
- Aluminum Foil	- Paper Towel

- Meter stick - Large Paper Clips

HYPOTHESIS: (Hint: Make an educated guess that answers the question(s) asked in the "PROBLEM" and is testable)

My hypothesis for the distance paper airplanes fly when made of different materials:

____

____

My hypothesis for the distance paper airplanes fly with increased mass:

____

____

EXPERIMENT: In this lab we will be making paper airplanes from different materials and throwing them. We will then measure the distance the plane flew and compare weighted planes with non-weighted. Data tables will be provided for you to fill in. Simply follow the instructions step by step.

Instructions:

(1) Find a partner to complete this lab.

(2) Each pair will be responsible for one and only one paper plane made of the materials they are assigned.

(3) Every group will make the same type (style) of plane. The directions and illustrations for this plane will be displayed to the class for example.

(4) When the plane is finished you are to fly the plane in the hallway or outside. You will be responsible for measuring and recording the flight of your plane 5 times. Measure to the nearest tenth of a meter (ex. 6.7 m, which is equal to 670 cm).

(5) The average distance flown will be calculated for YOUR type of plane and filled in on the data table. To find your average, find the total distance your plane flew (add the distances measured for all trials) and divide by the number of trials.

(6) We will collect all of the data as a class to determine the averages and to create our graph when finished with the data table.

<u>Collecting Data</u>: Paper Airplane Lab Results - the distances the different types of airplanes flew in meters, measured to the nearest tenth of a meter. (ex. 6.7 m, which is equal to 670 cm).   .

Table 1: The distances each type of paper airplane flew.

Paper Airplane Types	Distance Flown (m)					
	1st Trial	2nd Trial	3rd Trial	4th Trial	5th Trial	AVERAGE
Paper Towel						
Paper Towel with Paperclip						
Aluminum Foil						
Aluminum Foil with Paperclip						
Construction Paper						
Construction Paper with Paperclip						
Printer Paper						
Printer Paper with Paperclip						

<u>ANALYZE DATA:</u>

1. Using a sheet of graph paper, create a bar graph showing the results (average) of the distance the planes flew.

No written answer required. Be sure to staple graph to the back of your lab paper!

DISCUSS RESULTS:

1. What is the independent variable in this lab experiment?

2. According to the class averages, which type of plane flew the farthest distance?

3. List TWO variables (factors) NOT recorded or measured in this lab that could affect the flight of the paper planes.

4. Looking at your graph, what would you expect to happen to the distance the planes flew if we added more paperclips?

5. How could we get better results from this lab? List and explain two ways we could reduce our errors.

APPENDIX F

PENNY DROP LAB

Appendix F: The Penny Drop Lab, used for the first treatment.

Name: ________________________________ Date: _________ Pd: ___________

:

PENNY DROP LAB

PRE-LAB NOTES:

1.____

____

2. ____

____

3. ____

____

PROBLEM: In this lab we will be practicing scientific lab skills, such as hypothesizing, following directions, recording data, measuring, creating graphs and tables, and analyzing your data to come to a conclusion. We will use these skills to try to answer our question.

Q: How far will a penny bounce when it is dropped from different heights?

RESEARCH: The penny was the first type of currency of any kind that was authorized by the United States. Originally minted in 1787, the penny has had 11 design changes and has been made from several metals including copper, zinc, steel, and nickel. Nearly 300 billion pennies have been produced since the original minting of the coin. Today the penny is made of 97.5% zinc and only 2.5% copper. The current design makes each penny have a diameter of 19 mm and a mass of 2.5 grams.

What materials are needed to answer our questions?

- 3 pennies
- Piece of masking tape
- Meter stick

HYPOTHESIS: (Hint: Make an educated guess that answers the question asked in the "PROBLEM" and is testable)

My hypothesis for the distance from the target a penny will bounce as the height is increased: ____

EXPERIMENT: In this lab we will be dropping pennies from specific heights and measuring how far they bounce/roll after striking the ground. Data tables will be provided for you to fill in. Simply follow the instructions step by step.

Instructions:

(1) Place a piece of masking tape on the ground near the edge of your lab table. Put an “X” on the piece of tape using a pen or pencil. This will serve as your target.

(2) Drop the first penny from a height of 25 cm. Be sure the penny lands directly on its edge every drop. Be sure to try to get the penny to strike the same spot (“X”) on every drop.

(3) Measure the distance the penny bounces/rolls for each drop and record data in the table provided.

(4) Repeat for a total of five (5) times at a height of 25 cm.

(5) Drop penny from heights of 50, 75, and 100 cm.

(6) Create a graph showing the results of your data. Be sure to use the “rules” to creating a good graph as discussed in class.

Collecting Data: Record data for each drop in the table below.

Table 1: Distance penny rolls after dropping from various heights.

Height	**Distance (1)**	**Distance (2)**	**Distance (3)**	**Distance (4)**	**Distance (5)**	**Average Distance**
(cm)	**(cm)**	**(cm)**	**(cm)**	**(cm)**	**(cm)**	**(cm)**
25						
50						
75						
100						

ANALYZE DATA:

1. Using a sheet of graph paper, create a graph showing the results (average) of the distance the penny rolled compared to how high it was dropped from.

No written answer required. Be sure to staple graph to the back of your lab paper!

DISCUSS RESULTS:

1. According to your data, what happened to the distance the penny bounced as the penny was dropped from greater heights?

2. Did the data collected support your hypothesis? (Was your hypothesis correct?) Explain why.

3. What was the DEPENDENT VARIABLE in this experiment? Explain your answer.

APPENDIX G

FLOAT YOUR BOAT LAB

Appendix G: The Float Your Boat Lab, used for the second treatment.

Name: ________________________________ Date: _________ Pd: ___________

FLOAT YOUR BOAT

PRE-LAB NOTES:

1.____

____

2. ____

____

3. ____

____

PROBLEM: In this lab we will be practicing scientific lab skills, such as hypothesizing, following directions, recording data, measuring, creating graphs and tables, and analyzing your data to come to a conclusion. We will use these skills to try to answer our question.

Q: How much mass will your boat keep afloat?

RESEARCH: Being less dense than water is what keeps objects afloat. From cruise ships to canoes, the only way boats float is if they are less dense than the surrounding water. The world's largest ship is currently under construction and is expected to be over a mile long. The ship, named "Freedom" will be a floating city that will be home for 50,000 people, capable of having 20,000 guests and requiring 15,000 people to keep the ship in running order. At nearly 4,500 feet (1,370 m) in length, this ship will dwarf all other vessels on the sea. If you could stand the ship on end it would be as tall as nearly 4 Empire State Buildings stacked on top of one another. Weighing in at 2,700,000 tons, this ship will dwarf all others.

What materials are needed?

- 1 Sheet of Construction Paper
- Scotch Tape
- Masking Tape
- Triple Beam Balance
- Metric Ruler
- Calculator

Design your boat.

1. In the space below, draw a simple picture of what you want your boat to look like and the approximate dimensions (in centimeters). You will then calculate how much mass your boat should be able to keep afloat until it sinks (Its density becomes greater than that of water).

HYPOTHESIS: (Hint: Make an educated guess that answers the question asked in the "PROBLEM" and is testable)

We will calculate the mass your boat should hold. This will be your hypothesis.

MASS OF YOUR BOAT: ___________________________

VOLUME OF YOUR BOAT: ___________________________

DENSITY OF YOUR BOAT: ___________________________
(Density = Mass/Volume)

PREDICTED AMOUNT OF MASS: ___________________________
(Volume of Boat – Mass of Boat = Predicted Mass Boat Will Hold)

My hypothesis for the amount of mass my boat will hold (use your PREDICTED AMOUNT OF MASS from above): ____

____

EXPERIMENT: In this lab we will be measuring the mass our boats can keep afloat.

Instructions:
(1) Once you've come up with your hypothesis you may now float your boat. Place your boat in the water and slowly add mass until the boat begins to take on water and sinks.

(2) Start with larger mass objects first, then as your boat starts to "ride" lower in the water, gently add smaller masses.

(3)The maximum amount of mass your boat floated BEFORE it sank is your actual mass. When your boat sinks, pull all of the masses from the bottom of the float tank and add them together.

Collecting Data: Record data for your boat in the spaces provided below.

PREDICTED AMOUNT OF MASS (from above): ________________________

ACTUAL AMOUNT OF MASS (from experiment): _______________________

ANALYZE DATA:

1. How much mass did your boat keep afloat before sinking?

2. How far off from your "Predicted Amount of Mass" was the "Actual Amount of Mass" that your boat floated? Was it more than, less than, etc? (ex. My boat held 45 more grams than my prediction)

DISCUSS RESULTS:

1. What could you have done differently to get your boat to float more mass?

2. How would increasing the volume of your boat affect the amount of mass it could float? (Hint: you may want to use easy numbers to try sample calculations where you change the volume)

3. How would increasing the mass of your boat affect the amount of mass it could float? (Hint: you may want to use easy numbers to try sample calculations where you change the mass)

APPENDIX H

GUMMY BEAR LAB

Appendix H: The Gummy Bear Lab, used for the third treatment (adapted from www.sciencespot.net).

Name: ________________________________ Date: _________ Pd: ___________

Gummy Bear Lab

Pre-Lab Notes:

1. ____

____

2. ____

____

3. ____

____

PROBLEM: In this lab we will be practicing scientific lab skills, such as hypothesizing, following directions, recording data, measuring, creating graphs and tables, and analyzing your data to come to a conclusion. We will use these skills to answer our question:

Q1: What will happen to a gummy bear's mass when soaked in water?
Q2: What will happen to a gummy bear's volume when soaked in water?

RESEARCH: Hans Reigel invented gummy bears and other gummy candy during the 1920's. Reigel was the owner of a German candy company. Edible gelatin is the basic ingredient in gummy candy because it gives candy elasticity, the desired chewy consistency, and a longer shelf life. Gelatin is also found in soft caramels, marshmallows, foam-filled wafers, licorice, wine gums, pastilles, chocolate coated mallows and a host of other sweets,. Gummy candies are a blend of corn starch, corn syrup, sugar, gelatin, color, and flavor.

What materials are needed?

- 1 Gummy Bear
- Small Cups
- Water
- Triple Beam Balance
- Metric Ruler
- Calculator

HYPOTHESIS: (Hint: Statement, Testable, Answers Question)

My hypothesis for Q1:

____

____

My hypothesis for Q2:

EXPERIMENT: In this lab we will be measuring the mass of a gummy bear, calculating the volume, and then using those two values to find its density. We will then soak the gummy bear in water overnight and re-measure the bear to see what changed.

Instructions:

Day 1:

1. Zero the triple beam balance. Place the gummy bear on the tray and measure the mass of the bear.
2. Using a ruler, measure the length, width, and height of the bear.
3. Calculate the volume and record your measurements. Be sure to follow the teacher's directions for calculating volume.
4. Calculate the bear's density by using the density formula (Mass ÷ Volume).
5. Write your initials and period number on the bottom of a small cup. Pour just enough water in the cup until it becomes approximately half full.

Day 2:

1. CAREFULLY RETREIVE THE BEAR FROM ITS CUP, SLOWLY POURING OUT THE WATER BY PLACING YOUR FINGER OVER THE OPENING UP THE CUP AND POURING OUT THE WATER.
2. Zero the triple beam balance. CAREFULLY place the gummy bear on the tray and measure the mass of the bear.
3. Using a ruler, measure the length, width, and height of the bear.
4. Calculate the volume and record your measurements. Be sure to follow the teacher's directions for calculating volume.
5. Calculate the bear's density by using the density formula (Mass ÷ Volume).

Collecting Data: Show your work in the "Calculations" and record the data for your gummy bear in the "Data Table" spaces provided below:

Calculations:

Day 1:

*Use the **Volume = Length x Width x Height** equation to figure out your bear's volume. Remember to include units! Show your work below:*

a. Write out the equation needed to solve the problem:

b. Plug in your measured values in the equation:

c. Write your answer with appropriate units:

*Use the **Density = Mass / Volume** equation to figure out your bear's density. Remember to include units! Show your work below:*

a. Write out the equation needed to solve the problem:

b. Plug in your measured values in the equation:

c. Write your answer with appropriate units:

Day 2:

*Use the **Volume = Length x Width x Height** equation to figure out your bear's volume. Remember to include units! Show your work below:*

a. Write out the equation needed to solve the problem:

b. Plug in your measured values in the equation:

c. Write your answer with appropriate units:

*Use the **Density = Mass / Volume** equation to figure out your bear's density. Remember to include units! Show your work below:*

a. Write out the equation needed to solve the problem:

b. Plug in your measured values in the equation:

c. Write your answer with appropriate units:

Table 1: Gummy Bear Lab Data

Day	**Bear Color**	**Mass (g)**	**Length (cm)**	**Width (cm)**	**Height or Thickness (cm)**	**Volume (cm^3)**	**Density (g/cm^3)**
1							
2							
Difference *(Day 2 – Day 1)*							

ANALYZE DATA:

1. Using a sheet of graph paper, create a double bar graph showing the results of Day 1 Mass and Volume vs. Day 2 Mass and Volume.

No written answer required. Be sure to staple the graph to the back of your lab paper!

DISCUSS RESULTS:

1. Do you Accept (agree) or Reject (disagree) your hypothesis H1? Circle your choice and explain your decision using the data.

ACCEPT REJECT

Why? Explain using your data:

2. Do you Accept (agree) or Reject (disagree) your hypothesis H2? Circle your choice and explain your decision using the data.

ACCEPT REJECT

Why? Explain using your data:

3. Which piece of information recorded in the data table do you think is least important? Explain why you think this is and why it had no effect on the results of the experiment.

4. Which change was greater, volume or mass, from Day 1 to Day 2? How did this affect the density of your bear? Explain how mass and volume affects the density of an object using data from your data table.

5. How could we get more accurate results from this lab? List and explain one way we could reduce our errors.

6. What other experiments could be done using gummy bears? Describe at least one new experiment you could do.

APPENDIX I

SEPARATING A MIXTURE LAB

Appendix I: The Separating a Mixture Lab, used for the fourth treatment.
Name: ________________________________ Date: _________ Pd: ___________

SEPARATING A MIXTURE

PRE-LAB NOTES:

1.____

____

2. ____

____

3. ____

____

PROBLEM: How can you separate a mixture of wood, gravel, salt, sand, and water? In this lab we will be practicing scientific lab skills, such as hypothesizing, following directions, recording data, and analyzing your data to come to a conclusion.

RESEARCH: Mixtures include several substances in the same container or space, but are NOT chemically combined. We can use different methods for separating them into their individual components. Mixtures can be classified as either homogeneous or heterogeneous. Heterogeneous mixtures do not have a uniform composition and the different substances can easily be seen. Some examples of a heterogeneous mixture include concrete, soil, and a salad. Homogeneous mixtures have a uniform composition and are difficult to see the different substances in the mixture. Some examples of homogeneous mixtures include air, sugar water, and other solutions.

Solutions have two parts to them, solute and solvents. The solute is the substance that is dissolved, such as salt, and the solvent is the substance that does the dissolving (usually a liquid), such as water. The solute and the solvent together make solutions, such as salt water, or Kool-Aid.

What materials are needed?

- 1 slotted spoon
- 1 plastic colander
- 1 coffee filter
- 1 mixture container
- Water
- 1 hot plate
- 5 empty beakers

HYPOTHESIS: (Hint: Make an educated guess that answers the question(s) asked in the "PROBLEM" and is testable)

My hypothesis for separating wood: ____
____

My hypothesis for separating gravel: ____
____

My hypothesis for separating sand: ____
____

My hypothesis for separating salt: ____
____

My hypothesis for separating water: ____
____

EXPERIMENT: Instructions: Your lab group will be given a mixture of the following: wood chips, aquarium gravel, sand, salt, and water. Your job (you may decide as a group but should record your answers ON YOUR OWN PAPER) is to determine how to separate all of the components of the mixture into its separate parts.

(1) Decide how you would like to attempt to separate the mixture, using only the materials listed above.

(2) Attempt to separate your mixture by removing ONE substance at a time by using the tools provided. Place the separated substance into one of the empty beakers.

(3) Record what you used to separate each substance.

Collecting Data: WHAT DID YOU USE TO SEPARATE?
GIVE A BRIEF EXPLANATION!! (What tools did you use/What did you do?)

Wood:

Gravel:

Sand:

Salt:

Water:

ANALYZE DATA/DISCUSS RESULTS:

1. Which of your hypotheses (plural) worked the BEST (had the best results of separating your mixture)? EXPLAIN WHY:

2. Which of your hypotheses (plural) worked the WORST (had the worst results of separating your mixture)? EXPLAIN WHY:

3. If metal shavings (iron) were added to the mixture, how might you separate them from the other substances?

4. How does soil compare to our mixture? Briefly explain TWO ways.

APPENDIX J

RATE OF SOLUTION LAB

Appendix J: The Rate of Solution Lab, used for the fifth treatment.
Name: ________________________________ Date: _________ Pd: ___________

RATE OF SOLUTION

PRE-LAB NOTES:

1.____

____

2. ____

____

3. ____

____

PROBLEM: How does temperature affect the amount of sugar that can be dissolved in water? In this lab we will be practicing scientific lab skills, such as hypothesizing, following directions, recording data, and analyzing your data to come to a conclusion.

RESEARCH:
WHAT IS A SOLUTION? A solution is composed of two portions, a solute (the substance dissolved) and the solute (the liquid that does the dissolving). In a solution one substance dissolves within another to become a heterogeneous mixture. In chemistry, a heterogeneous mixture is when substances are not similar, but evenly distributed. A solution can be made using two liquids or a liquid and a solid or gas. When a solution becomes supersaturated, it contains more solute than it should be able to, and will fall out of solution (form a solid) due to temperature or pressure changes.

What materials are needed?

- 5 Empty Beakers	- Water
- Sugar Packets	- Ice Bath
- 3 Hot Plates	- Stirring Rods
- Thermometer	- Large Glass Dish
- Graduated Cylinder	

HYPOTHESIS: (Hint: Make an educated guess that answers the question(s) asked in the "PROBLEM" and is testable)

How will the temperature of water affect the amount of sugar that can be dissolved in water?

____

____

EXPERIMENT: In this lab, we will be determining how water with different temperatures affects the amount and rate of sugar that can be dissolved.

Instructions:

(1) Turn on the three hot plates, one set to (LOW), one set to (3), and one set to (HIGH). A beaker containing 100 mL of water should be placed on each hot plate.

(2) While the hot plates are warming the water, create an ice bath by filling the provided glass dish with ice and water. Place a beaker containing 100 mL of water in the center of the ice bath.

(3) Begin with the cold, ice water bath. Measure and record the temperature. Then open one packet of sugar and pour into the beaker. Stir until dissolved. Continuing adding sugar until it does not dissolve in the beaker of cold water (when you see solid sugar particles in the bottom of the beaker that will not dissolve). Record data.

(4) When finished with the cold water, repeat using the warm, medium, and hot water beakers. (You may want to split this task among lab group members). Record data.

(5) When all data is collected, pour out the sugar solution, and refill the large beaker.

Collecting Data: Complete the table below.

Table 1: Rates of Solution data collection.

per 100 mL of water	Temperature			
	Hot (High)	Medium (3)	Warm (Low)	Cold (Ice)
Temperature (o C)				
# of sugar packets dissolved				
Mass of sugar dissolved (g)				

ANALYZE DATA:

1. Using a sheet of graph paper, create a graph showing how temperature affects the amount of sugar dissolved.

No written answer required. Be sure to staple the graph to the back of your lab paper.

DISCUSS RESULTS:

1. Do you Accept (agree) or Reject (disagree) your hypothesis? Circle your choice and explain your decision using the data.

ACCEPT REJECT

Why? Explain using your data (include the measurements/observations collected during the lab to explain your decision):

2. If you were making a batch of Kool-Aid, describe how you could make sure there are no solids (sugar and flavoring) left in the bottom of the container after stirring.

3. Describe/Compare the amount of sugar dissolved in the hottest and coldest water used in the lab.

4. By examining your graph, what do you think would happen to the sugar dissolved in the hot water if cooled greatly using ice, or if only water was removed?

5. Describe and explain one way in which we could improve the results/accuracy from this lab.

6. Calcium carbonate ($CaCO_3$) is produced by dissolved coral remains in the ocean, and is the exact opposite of sugar. It dissolves best in cold ocean water and becomes **supersaturated** in warm ocean water. What will happen to calcium carbonate when cold ocean currents collide with warm ocean currents (see research section)? *Hint* The Bahamas are a great example of this type of activity.

7. What other experiments could be done with solutions and dissolving substances? Describe at least one new experiment you could do.