The Effects of Student Reflections on Mastery of Academic Standards in a Standards Based Grading Modeling Classroom

Jesse Ruiz, Chad Jacobs, Timothy Quinn

Principal Investigator: Dr. Robert Culbertson
Arizona State University

This study was performed as an Action Research required for the Master of Natural Science degree with a concentration in physics. Further correspondence regarding this study should be directed to (Robert Culbertson), Department of Physics, Arizona State University, Tempe.

Submitted July 2018
Table of Contents

Abstract .................................................................................................................................. 2
Acknowledgments .................................................................................................................. 3
Rationale ............................................................................................................................... 4
Literature Review .................................................................................................................. 6
Method .................................................................................................................................. 14
Procedure for Treatment ..................................................................................................... 16
Data Analysis ....................................................................................................................... 20
Timeline .............................................................................................................................. 24
Results .................................................................................................................................. 25
Conclusion ............................................................................................................................ 52
Implications for Further Research ...................................................................................... 54
Works Cited .......................................................................................................................... 55
Abstract

Student learning results from a conglomeration of student and teacher efforts. However, students often view their learning in physics and mathematics as solely the teachers’ responsibility. Research has been done to show that standards based grading allows for students to continually develop their understanding of the standards taught. The investigation group theorized that having students reflect on their success while explicitly addressing the standards throughout the learning process would increase their mastery of the standards. Students were required to reflect on their learning and mastery of the standards taught throughout the semester, and wrote about their achievement and their options for improvement. The results indicate that students who reflected on the standards did not demonstrate significantly greater gains in understanding force and algebra concepts overall than those students who did not.
Acknowledgments

The writing of this research paper was an intense learning experience for the three of us. At the onset we were unsure of exactly what to expect along the way, and how many hours this research would take. We would like to thank several individuals for their assistance along the way. Thank you to Drs. Robert Culbertson and Carl Covatto, and Brenda Hacker for the work they do to improve the education physics teachers receive and thus the education received by our students. We thank Dr. Colleen Megowen for her guidance during the writing of this paper. We thank Dr. Chantelle Frazee for her help with the statistical analysis. We thank Dr. Jane Jackson for all of her support provided to the MNS program and the teachers therein. We thank Jim Archambault for helping get us started on this endeavor, and consistently giving advice on how to improve our study. Thank you to Jared Washburn, Heather Babb, Josh Cunningham, Kathy Dao, Sandra Gorton, Karen Preston, Michael Crofton, Theresa Burch, and Angela McClure for allowing us to learn from your experiences in the action research process. Finally, to our families, who have been so understanding of the late nights and endless hours of writing and rewriting, the extreme highs and disappointing lows throughout the entire action research process. We truly appreciate all of you for all the support you have provided.
Rationale

In education, student success has traditionally been measured in a style of grading that compiles a set of scores from various homework assignments, projects, quizzes, and tests, while sometimes including extra credit, and averaged to give a percentage on a scale of 0 to 100. The state and district formally list standards and objectives for students to master. The method of traditional points-based grading can sometimes be frowned upon as a grade that doesn’t accurately reflect the students’ ability to master the material. A new form of grading, referred to as Standards-Based Grading, is a type of grading that gives a score to a student based on their achievement in the standard as listed by the district. These standards can be further broken down by specific objectives and used as the “assignments” the student is graded on. Using this method implies that the student will take no penalty for when the student achieves the material. It isn’t a matter of when the student learns it but rather did the student learn it. This also improves communication to parents and students about what the “letter” or “percentage” means when receiving grades back. The students know where their understanding lacks in the content, so they can better improve on a specific goal rather than figuring out busy work to do to help improve their grade. This type of learning is referred to as “assessment for learning” and uses test results to focus on helping the students figure out what the big ideas for the learning are. Another benefit for standards based grading is allowing for a higher quality curriculum and assessment for all the learning goals. The teacher is able to align each assessment with learning goals which in turn aligns with a higher quality instruction in the classroom.

This study’s main purpose was to study the effects of student achievement while applying standards-based grading. To maximize the effects of implementing the standards-based grading, the investigators incorporated the idea of student reflections. Students using reflective practices have been proven to benefit their understanding and enhance their learning in the process. An additional contribution toward maximizing the effectiveness of the student’s success in the study includes the application of the Modeling Instruction teaching method used in science classrooms. This method allows the students to develop a scientific model using the scientific method through an inquiry based approach and data collection.
This study attempted to improve students’ academic achievement in a math and Modeling Instruction science classroom by addressing the specific standards and objectives, while applying a standards-based grading scale with the students reflecting on their personal achievement during the process. They wrote weekly journal reflections on the standards addressed during the week, as well as the strategies used to achieve this goal. They summarized their journal reflections in a reflection form at the end of each assessment, and their scores were compared to scores without the reflection and without the standards-based grading. It was theorized that having students reflect on their success while explicitly addressing the standards throughout the process of the learning would increase their mastery of the standard. This can also point toward further study into the use of standards-based grading as well as the effectiveness of students’ reflective practices.
Literature Review

Standards

Traditional points-based grading is becoming a thing of the past. Recently there has been much discussion regarding whether assigning grades according to points earned on individual assignments, quizzes, and tests fairly represents a student’s success. Problems that have been noted with traditional grading practices are that “they are imprecise and uncalibrated, include ‘non-achievement’ factors, and require the mixing of different types of grades to get an ‘average’ single score” (Ballard, 2010).

The standards-based grading method gives a way to monitor students’ progress toward mastery of the subject instead of assignments dictating a students’ achievement (Miller, 2013). Students are presented with the content and tested at the end of each unit. These tests solely assess the students on how well they have mastered the content presented to them. Should a student not achieve mastery, he or she is allowed to retest, an indefinite number of times, but in practice this number is dependent on the teacher using standards-based grading (Miller, 2013).

Standards-based grading is a powerful way of addressing teaching methods, content, assessment, and differentiation in a classroom. Students are more informed about what they are learning rather than just being given a large amount of content. They are then more able to determine what needs to be learned (McGee, 2012; Ohio Department of Education, 2012; O’Connor, 2009; Scriffiny, 2008). Teachers have information related to student performance and can identify students who are in need of assistance earlier, so they can help the students achieve mastery (Ohio Department of Education, 2012; Oliver, 2011; O’Connor & Cooper, 2008; Scriffiny, 2008).

Standards-based grading reflects students’ mastery solely on academic achievement, and not on traditional grading practices that include homework completion, effort, and participation (Marzano & Heflebower, 2011; Phillips, 2011; Guskey & Jung 2006; Walker, 2006). Standards-based grading also reduces the discrepancy between teachers who grade differently, so students who perform at the same level, in the same grade level, or subject are more likely to get the same final grade (McGee, 2012; Urich, 2012; Guskey et al., 2011; Marzano & Heflebower, 2011; O’Connor & Cooper, 2008). Teachers judge students’ performance in terms of what
they have learned and what they are able to do, regardless of how well or poorly their classmates perform. Grades are not based on overall class performance, with the best students in the class earning A’s, when in fact, those best students may have not reached mastery of the standards. (Adrian, 2012; Urich, 2012; Guskey, 2011; Phillips, 2011; O’Connor, 2007a; Guskey, 2001a; Guskey, 2001b; Krumboltz & Yeh, 1996).

Classroom competition is removed, and the focus becomes learning. In a traditional setting students may compete for grades, and thus may not cooperate or help each other because of it. (Urich, 2012; Guskey, 2011). Experts believe that the lack of a clear standards-based curriculum harms low income and minority students more than it does others. In schools with high rates of student mobility, standards-based grading helps ensure that children are not left behind academically due to a move to another school. (Ohio Department of Education, 2012; Brookhart, 2011; Craig, 2011; Paeplow, 2011; Great Schools, 2008).

Zimmerman, a physics teacher, has four main principles that encompass the ideas behind standards-based grading: the student should not be penalized for their mistakes, but rather their successes, the student should be focused on mastering the material, student grades should reflect skill level, and the student should be allowed to retest on a skill that is not mastered yet (Zimmerman, 2017). He says, “if points are used to motivate students to show up to class and turn homework in on time, grades are no longer an accurate reflection of how well students understand the material” (Zimmerman, 2017). In addition to these principles, standards-based grading also does its part in helping the teacher in their practice. “...(I)t forces teachers to think critically about what they are teaching, why they are teaching it, and whether students truly understand what is being taught” (Ballard, 2010).

How should students who are going to take advantage of the system and try to pass courses without completing any homework be handled? Some students will have no problem achieving mastery, and showing this mastery on assessments without completing homework assignments, but others may try and fall behind in their understanding. With standards-based grading, these students technically don’t “fall behind”, because they are allowed to retest, and their grades are cumulative. The grades change throughout the semester, and as students master the standards, grades
Effects of Student Reflections

will improve; however, Miller handles her lazy students by entering a zero in the gradebook under each standard in which the students are not completing their work. Without completing assignments, the students do not show mastery, so as Miller (2013) says, “(t)he procrastinators get the message when they see scores related to process and responsibility fall to six or seven on a ten-point scale.” (Miller, 2013)

Tests and quizzes are still an important aspect of learning as it is a great method for students to keep testing their abilities, and for the teacher to assess where the students are at, as far as mastering the standards. Sometimes this gives rise to some issues where a student either doesn’t study or suffers from test anxiety. The use of a second assessment opportunity helps both of these student situations, as it gives several chances for the former student, and questions of familiarity for the latter student (Miller, 2013).

Reflection Practices

While standards-based grading is a method for measuring a student’s achievement, there are many methods in the field of education that have been studied to reach this ultimate goal of what education is all about. Educators want nothing but to help guide their students on the path that will help them grow and blossom into self-sustaining adults who can think and process information on their own. “...The goal of education should be the emancipation of students’ thinking in order to empower them to take control of their lives” (Ornstein & Hunkins, 1998). This goal of emancipation for the students is achieved by guidance from the instructor as they develop their thinking in different learning environments. Instructors rarely think of this as the goal for the students because they are too worried about getting the students to perform well on the standardized test. “...(I)n our culture we are obsessed with external evaluation and because of this teachers need to consciously encourage students to reflect upon their learning” (Bond, 2003). “...(There is) strong emphasis on the responsibility of teachers to create opportunities for reflective thinking, to monitor the appropriate development of student learning, and to intervene to avoid inappropriate constructions of knowledge” (von Glasersfeld, 2000). This idea of reflective thinking can be achieved by allowing the
students to construct their own understanding and write down their thoughts of the
learning that is going on.

The act of reflecting on one's learning requires the student to think outside of the
ideas presented and more on how the process of learning has made them feel or think.
This can be a way for students to express what they have already accomplished and
isolate what needs more attention. It is about “thinking about thinking and emphasizes
the individual’s self-awareness of his or her thinking patterns” (Bond, 2003). It was
theorized by Albert Bandura (1997), “the reflective processes mediate the information
about perceived efficacy that learners acquire from different sources of their learning
experience.” It can be a way for the student to organize their thoughts in a manner that
makes sense to them while also achieving a few other goals. “The purposes of
reflection journal writing include: to critically review the behaviors (e.g. strengths and
weaknesses; learning styles and strategies); learning of self and others; setting or
tracking learning goals (i.e. how learning took place); and exploring connections
between knowledge that was learned and students’ own ideas about them” (Lew, 2011).
The main idea here is the learning that is happening between the student and ‘oneself’.
“(J)ournal writing can improve self-regulation strategies if it is structured around self-
reflections in terms of learning goals, learning strategies, observations, understanding,
feeling, and dialogues with oneself and others” (Al-Rawahi, 2015).

The idea of Learning Journals gives students this opportunity to take notes not
on the content itself, but the way in which their thinking about the content is developed.
“Students who kept learning journals also showed more sophisticated conceptions of
learning, greater awareness of cognitive strategies, and demonstrated the construction
of more complex and related knowledge structures when learning from text...(t)hey also
performed significantly better on the final examination for the course” (Lew, 2011). The
cognitive strategy mentioned here is directly relating reflections as dealing with the
psychology concept of metacognition. Flavell (1976) defined metacognition as “one’s
knowledge concerning one’s own cognitive processes and products or anything related
to them.” Bond (2003) also mentions many other definitions of metacognition as a
branch from cognitive psychology that focuses on a student's ability to actively be
participating in their thinking and learning process, or a way to remember, think, and
Effects of Student Reflections

act. Some might be confused about the difference between what cognitive and metacognitive are; however he states, “...cognitive as being involved in doing and metacognitive as being involved in choosing and planning what to do and monitoring what is being done” (Bond, 2003). Although both are important in a science classroom, the use of metacognition is an internalized thought process for the student to work through. Bond found that “the inclusion of metacognitive strategy caused an increase in student performance.” This led Bond to conclude that the reflective assessment strategy succeeding in an increase of achievement for students who practiced it (Bond, 2003).

The students might not always do this through words, but rather in the form of an assessment. “Reflective assessment is a formative process through which students can experience assessment as part of learning, rather than as a separate evaluative process” (Bond, 2003). The reflective piece should be a way for students to ask themselves how they are doing so far and what they need to do to get better. Metacognition, or the act of reflecting on one's practice, is a psychological concept in cognition, which “focuses on the active participation of the individual in his or her thinking process” (Stewart & Landine, 1995). Simply put, it gives the students the chance to organize what they know and what they don’t know. Having the students organize their thoughts in this manner can help them focus more on what areas of the content they are struggling with, to potentially improve their overall understanding of the specified standards.

In a science classroom, a lot of different types of learning are happening from hands-on manipulations, direct instruction, visual demonstrations, and self-guided practice. “...(S)tudies have indicated that students approach the learning of science passively when they rely solely on absorbing scientific information and storing it in the memory without sufficient reflection on what has been learned” (Al-Rawahi, 2015). Science should not be about memorizing formulas and inputting numbers to mechanically crank out an answer that leaves the student with little to no meaning of the answer that was given. “Students appeared to develop better strategies in problem solving through writing as compared to mere memorizing of calculations” (Lew, 2011). Science should give the student an opportunity to engage in active learning strategies.
“Making observations, hands-on activities, dialogue with others, and self-reflection through journal writing are examples of active learning strategies” (Al-Rawahi, 2015).

The teaching strategy that will be applied in the current study is the Modeling Method as developed mainly by Arizona State University (ASU) in which the science is learned through an inquiry-based model. The students are asked an overarching question for which they use the scientific method to help analyze and collect data to develop a model to understand the situation. “When journal writing is associated with an inquiry-based approach to learning, it helps improve the quantity and quality of students’ questions, which is the first essential step in inquiry-based learning” (Al-Rawahi, 2015).

The outcome of reflective journal writing does not only benefit the student, but also the teacher as well. “...the purpose of informal reflective activity is to both encourage student reflection and to provide feedback to the teacher about their teaching... (it) can inform both students’ studying and teachers’ teaching (Bond, 2003).

Overall, reflective practices have been found to benefit the students learning process more than it has not. Bond found that, “...the mean score of students who practiced reflective assessment strategies was significantly higher (than those who did not)” (Bond, 2003). “Studies indicate that journal writing engages students’ thinking through different cognitive processes such as prediction, brainstorming, reflection, and questioning. It encourages students to express their interests, thinking and curiosity about the world around them, and discover new ideas. This in turn contributes to the enhancement of their understanding of scientific concepts” (Al-Rawahi, 2015). However, some studies have proved that there is no direct correlation between student success and reflective journal writing. Lew’s study investigated this relationship and found no significant correlation between the two. “...(T)here is this possibility that the weak relationship between self-reflection and academic performance is because students are generally poor at self-reflection” (Lew, 2011). “(Students) could be described as ‘inexperienced’...because they were first-year students in high education” (Lew, 2011). This does raise the question of how accurate students’ reflective practices need to be in order for the practice to benefit their academic achievement. Other issues that Lew found were the lack of commitment from the students on the goal of increasing academic achievement through the use of reflecting on the learning. "(S)ome students
simply do not take the task of journal writing seriously while others perhaps do” (Lew, 2011). “(Students) were skeptical about the need to reflect on how and what they have learned, citing reflection journals as ‘mechanical and meaningless’” (Lew, 2011). This could be a possible challenge dealing with the students at the high school level. Lew gives his advice that “learners may need a structure to guide the complex process of self-reflection on the content and process of their learning” (Lew, 2011). With this considered, the reflection process can be modified to ensure the students give maximum effort to evaluate the data without these possible limitations. By combining this metacognitive skill that students can develop and use with the standards-based grading, students can reflect on their success in each of the standards so they are made aware of their progress not only via a quiz or test score, but also through this self-reflective practice. This used in tandem with the modeling method for teaching science could potentially extrapolate the goals of developing their own model, analyzing it, refining it, and working toward the end goal of creating these scientific relationships that ultimately lead them to success in the objective and standard.

**Modeling**

While standards-based grading ensures student mastery of a standard, and reflective practices ensures effective student learning, the method of teaching contributes to a student’s achievement which can be done through the modeling method of science learning. The modeling theory of physics instruction has arguably had the greatest impact on high school physics instruction of any physics education reform (Brewe, 2008). This theory is centered on the idea that physicists develop mental models to help them process their thoughts. These mental models are constructed through the use of graphs, charts, diagram, and formulas. The modeling method differs greatly from the traditional lecture method. The lecture method is a didactical method, best suited to setting forth a particular line of thought to an audience which shares the preconceptions of the lecturer. Therefore, it is more appropriate for advanced courses in physics than for introductory courses where the majority of students are not yet accustomed to the modes of scientific thinking (Halloun & Hestenes, 1987). Traditional instruction often overlooks the crucial influence of students' personal beliefs on what
they learn (Halloun & Hestenes, 1987). Students’ comprehension of the physical world at the beginning of an introductory physics course is a fragmented collection of common sense generalizations which are primarily prescientific (Halloun & Hestenes, 1985). Modeling Instruction is designed to help students develop model-centered knowledge bases that resemble those of practicing scientists (Brewe, 2008). Modeling Instruction has been proven to produce significant learning gains (Desbien, 2002), which likely result from the fact that the models that students investigate have been carefully chosen based on the underlying structure of the material. Modeling Instruction produces students who engage intelligently in public discourse and debate about matters of scientific and technical concern (Jackson, Dukerich, & Hestenes, 2008). The National Science Education Standards state, "Student inquiries should culminate in formulating an explanation or model... In the process of answering the questions, the students should engage in discussions and arguments that result in the revision of their explanations." A key component of this approach is that it moves the teacher from the role of authority figure who provides the knowledge to that of a coach/facilitator who helps the students construct their own understanding (Jackson, Dukerich, & Hestenes, 2008).

There is strong evidence that implementing standards-based grading is beneficial for both the student and the teacher. It allows students the time to learn, and the opportunity to reach mastery after additional practice. Teachers benefit as they are able to make informal assessments that let them know how well students are doing, and then adjust their instruction to meet the students’ needs. The use of written reflections by students, such as journal writing and other metacognitive activities requiring students to assess, explain, make connections, and think, have been shown to increase students’ comprehension of the classroom subject matter. The use of the Modeling method in the physics classroom helps students learn by making mental models of behavior they see. These models are refined as they learn more, just as actual scientists do when they are studying the natural world. By linking standards-based grading with student reflections, in a modeling classroom, it is hypothesized that students will perform better on assessments in the classroom.
Method

Subjects and Investigators in Experimental Group

Investigator 1:
Investigator 1 teaches at a suburban school in the metropolitan Phoenix area with an enrollment of approximately 2500 students in grades 9-12. The population consists of 56% White, 35% Hispanic, 5% Black, 2% Asian, 2% American Indian/Alaskan Native, and 1% of two or more races with a total of 44% minority. The school is equally distributed by gender with females at approximately 49% of the population. There are 53% of the students enrolled in a free or reduced lunch program, classifying the school as a Title 1 school. Investigator 1 worked with 31 General Physics (SC81) students. General Physics is populated with approximately 38% seniors and 62% juniors. The General Physics class is offered for Dual Enrollment to receive credit for PHY101.

Investigator 2:
Investigator 2 teaches at the same school as Investigator 1. Investigator 2 worked with 20 General Physics (SC81) students. General Physics is populated with approximately 69% seniors and 31% juniors. The General Physics class is offered for Dual Enrollment to receive credit for PHY101.

Investigator 3:
Investigator 3 teaches at the same school as Investigators 1 and 2. Investigator 3 worked with approximately 50 Algebra 1 students. Algebra 1 is populated with approximately 91% freshmen and 9% sophomores.

Contrast Group 1:
Contrast Group 1 consists of 3 classes of General Physics students taught at the same school by Investigators 1 and 2 from the previous school year of 2016-17. Approximately 54% of the students at this school were enrolled in free and reduced lunch. The classes used for the contrast group consisted of 64% seniors and 36% juniors with a total of about 70 students. These classes were taught by Investigator 1.
and Investigator 2 in a General Physics class, also offered for college credit. These classes use some of the modeling curriculum designed mostly by Arizona State University. This allows the students to develop models from the laboratory setting and construct the physics afterward, guided by the instructor.

**Contrast Group 2:**
Contrast Group 2 was in a large suburban high school in a major Southwestern city. The school had an enrollment of approximately 3,720 students of whom about 63% were Hispanic, 28% Caucasian, 1.5% Asian, 4% African American, and 3.5% were multiracial. Approximately 65% of students came from low-income households and qualified for free and reduced lunch. Comparison group 2 consists of 2 classes of General Physics students taught at another school in the same district as the investigators. The classes used for the contrast group consisted of 55 students. These classes are taught similar to Investigator 1 and Investigator 2’s General Physics class offered for college credit. It also is similarly taught using the Modeling Method of Instruction. Contrast Group 2 only contributed FCI and Semester District Exam scores.

**Contrast Group 3:**
Contrast Group 3 consisted of 3 classes of Algebra 1 students taught at the same school as the investigators by another Algebra 1 teacher. The classes used for the contrast group consisted of 94% freshman and 6% sophomores with a total of 80 students. Contrast Group 3 were taught using Standards-based grading but not using any reflective practices. The data this group contributed consisted of Mathematics Concepts Inventory, Algebra Concepts Inventory, and Semester District Exam (AZMerit).

**Contrast Group 4:**
Contrast Group 4 consisted of 5 classes of Algebra 1 students taught at the same school as the investigators by Algebra 1 Teachers from the previous school year of 2016-17. The class used for the contrast group consisted of 93% freshman and 7% sophomores with a total of 107 students. This class is taught by the same teacher as
Contrast Group 3’s Algebra 1 class. The data this group contributed consisted of Semester District Exams (AZMerit).

**Procedure for Treatment**

**Permission**

All students, along with their parents or legal guardians, signed a consent form acknowledging their participation in the study. If permission to be in the study was not submitted from either the student or parent/guardian, they were not included in the study. The participants in the study will remain anonymous and confidential.

**Pre-assessment of Students Abilities**

Investigators 1 and 2 and Contrast Groups 1 and 2 instructors were given the necessary material for Standards-based grading and FCI exams at the beginning of each year. The Force Concept Inventory (FCI) was given at the beginning of the unit on forces. The FCI was developed by ASU as a pre-assessment to measure students’ baseline conceptual understanding of the physics that is taught within the units involving motion and forces. The FCI was administered at the end of the first semester following the conclusion of the force and uniform circular motion unit.

Similar pretests were used for the Algebra 1 students given by Investigator 3 and Contrast Group 3, called the Mathematics Concepts Inventory (MCI). The MCI was developed by the Physics Underpinnings Action Research team at Arizona State University (ASU) in June 2000 and revised six times in the next three years. It has 23 questions and is intended for 8th and 9th grade students. The first eight questions are paired, on scientific thinking skills (conservation of mass and volume, proportional reasoning, control of variables). They were recommended by Professor Anton "Tony" Lawson, ASU School of Life Sciences, from his Classroom Test of Scientific Reasoning, a widely-used research-informed instrument (Boyarsky, 2009). In addition to the MCI, Investigator 3 and Contrast Group 3 also administered the Algebra Concepts Inventory (ACI) developed by a graduate of the University of Utah (Underdown, 2015). That instrument is not research-validated. The ACI questions address algebraic concepts more abstractly using variables to develop mathematical deductions and properties.
Both were administered at the beginning and end of the school year. Contrast Group 4 did not have MCI or ACI data to compare, as this group is from a previous year.

Treatment

Investigators 1 and 2 conducted the treatment on the first semester of General Physics that incorporate concepts in motion and forces. The focus of standards for the physics classroom are set by Arizona Academic Content Standards and further detailed by their school district’s guidelines with further information on vocabulary, specific notes, integration practices, and resources of problems and lab situations. These standards were approved by the district Governing Board on May 10th, 2005. The modeling method of instruction for the sequenced units used in the study are seen in Table 1 below.

<table>
<thead>
<tr>
<th>Unit</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Constant Velocity</td>
</tr>
<tr>
<td>2</td>
<td>Constant Acceleration</td>
</tr>
<tr>
<td>3</td>
<td>Projectiles</td>
</tr>
<tr>
<td>4</td>
<td>Static Forces: $\Sigma F = 0$</td>
</tr>
<tr>
<td>5</td>
<td>Dynamic Forces: $\Sigma F = ma$</td>
</tr>
<tr>
<td>6</td>
<td>Momentum</td>
</tr>
<tr>
<td>7</td>
<td>Uniform Circular Motion</td>
</tr>
</tbody>
</table>

*Table 1: Outlined sequence of units used for Investigator 1 and 2*

Investigator 3 used data from Fall and Spring semesters of Algebra 1 that incorporate concepts in solving and graphing linear and quadratic equations and inequalities. The focus of standards for the Algebra 1 classroom are set by Arizona
Academic Content Standards and further detailed by Mesa Public Schools guidelines with further information on vocabulary, specific notes, integration practices, and resources of problems. These standards were approved by Mesa Public Schools Governing Board in December of 2016. The content used for instruction is sequenced in the units as seen in Table 2 below.

<table>
<thead>
<tr>
<th>Unit Number</th>
<th>Unit Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Solving Linear Equations</td>
</tr>
<tr>
<td>2</td>
<td>Solving Linear Inequalities</td>
</tr>
<tr>
<td>3</td>
<td>Graphing Linear Functions</td>
</tr>
<tr>
<td>4</td>
<td>Writing Linear Functions</td>
</tr>
<tr>
<td>5</td>
<td>Solving Systems of Linear Equations and Linear Inequalities</td>
</tr>
<tr>
<td>6</td>
<td>Exponential Functions and Sequences</td>
</tr>
<tr>
<td>7</td>
<td>Polynomial Equations and Factoring</td>
</tr>
<tr>
<td>8</td>
<td>Graphing Quadratic Equations</td>
</tr>
<tr>
<td>9</td>
<td>Solving Quadratic Equations</td>
</tr>
<tr>
<td>10</td>
<td>Radical Functions and Equations</td>
</tr>
</tbody>
</table>

*Table 2: Outlined sequence of units used for Investigator 3*

For all of the Investigators' classes, the students reflected on their learning goals and strategies used to reach achievement every Friday near the end of class. They hand-wrote a reflection on the learning goals and content addressed over the course of the week in a private learning journal. They were given a list of specific questions to be addressed in each reflection, which include the specific standard and objectives that
were focused on that week, the learning strategies applied throughout the week to work toward achieving those standards, any observations made while reaching their goals, how their understanding of the topic has developed over the course of the week, how the learning has made them feel, as well as any personal dialogue with oneself and others that was had regarding the learning in the class.

To ensure the students are providing meaningful feedback in their reflections and are aware of an effective method for reflecting the investigators provided the students with exemplar reflections for what they are expecting from their reflections. The first two full weeks were used as practice reflections for the more inexperienced students. For the first and second weeks the students modeled how reflecting should look, sound, and feel through whiteboarding sessions involving teacher and student input. Following the first couple weeks, the reflections were counted toward a grade in a writing standard.

At the end of each assessment, the students were asked a series of reflection questions about their achievement of each learning objective within each standard that the assessment covered. Using their weekly journal reflections, they summarized their reflections and responded according to the growth, or lack thereof, over the course of the unit. These Likert Scale questions are asked at the end of each assessment as a way to let the students reflect on how they believe they met each goal after taking the assessment. Being a one-to-one school, meaning each student has an individual laptop provided by the school, technology is a great tool to use for distributing the reflection forms as well as keeping track of the students’ work. With assignments that can be submitted online, work can easily be organized and dated to use for conferences or feedback when assessing student mastery.

Assessment

Each unit for all investigators had a summative assessment that covers the standards learned in each unit. These assessments are the same as the assessments given to the contrast groups in the previous year with a notable exception. The summative assessments given during the study have the specific standards, or performance objectives, embedded in the assessments. The assessments were followed by the students answering a series of reflection questions, which included their
perception of how well they mastered the standards. The questions are in Likert form of responses scored along a range of choices from strongly disagree to strongly agree. These questions consist of students reflecting on their performance of the assessment after completion. They also include specific Likert Scale questions involving their success on each specific objective or standard addressed in that unit. In addition, there are a couple of open ended, free response questions that ask the students to reflect on specific goals achieved in each unit in a short paragraph. For the purposes of this study, re-assessments are be limited to a specific number of times, but are restricted by the time allotted during the quarter.

At the beginning and end of the semester in the physics courses, the students took the FCI test. This was the same for both Investigator 1 and 2 as well as Contrast Group 1 and 2. For the Algebra 1 classes involved in the study for Investigator 3 and Contrast Group 3, students took the MCI and ACI at the beginning and end of the year. The unit assessments were the same for both groups as well and also include Contrast Group 4. Contrast Group 4 had similar unit assessments but no data for the MCI and ACI. Following each unit assessment, the same reflective questions were answered by the students, with the only modification being the standards addressed in each class.

**Data Analysis**

**Quantitative Data Analysis**

The FCI, ACI, and MCI pretest and posttest scores were used to calculate the normalized gain score for the subjects

\[
\text{Normalized Gain Score} = \frac{\text{Posttest score} - \text{Pretest score}}{\text{Max Possible Score} - \text{Pretest score}}.
\]

These normalized gain scores shows the students' potential growth they could have made from the pre to the posttest (Hake, 1998). The district Criterion Referenced Test (CRT) for physics at the end of the fall semester would be used for comparison between the physics treatment and contrast groups. For the district final in math, the students were tested by the AZMerit test which has their own score distribution that ranges from the minimum of 3577 to a maximum 3787. Between this range, there are
Effects of Student Reflections

four proficiency levels of Minimally Proficient (3577-3660), Partially Proficient (3661-3680), Proficient (3681-3719), and Highly Proficient (3720-3787). The scale scores were set when the test was developed and vertically scaled across all grade levels. The items were organized from low to high difficulty by content area expert teachers and then these same teachers determined where the cut off for points should be between the items for the performance levels.

Qualitative Data Analysis

The qualitative data in the study consisted of the written reflections provided by the students each week. These weekly reflections were tasked to the students but then ranked by the investigator on a scale of 0 to 4 based on how well they reflected on their progress. A breakdown of the rubric used to grade the quality of reflections can be seen in Table 3 below.

<table>
<thead>
<tr>
<th>Reflection Score</th>
<th>Qualifications</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No reflection</td>
</tr>
<tr>
<td>1</td>
<td>Student wrote something but did not reflect on his or her progress</td>
</tr>
<tr>
<td>2</td>
<td>Student reflected on progress but failed to elaborate on standards, current strategies, or future strategies to implement</td>
</tr>
<tr>
<td>3</td>
<td>Student reflected on progress toward mastering the standard and mentioned a few strategies currently being used, or that could be used for further achievement</td>
</tr>
<tr>
<td>4</td>
<td>Student reflected on progress toward mastering the standard while mentioning current strategies being implemented to study as well as discussing improvements that can be made for furthering their achievement</td>
</tr>
</tbody>
</table>

*Table 3: Scale for scoring written reflections*
The students were provided a set of sample questions to answer when reflecting to help guide them to a high quality reflection.

**Reflection Questions:**
1. How much did you know about this subject before we started?
2. Did you meet the standards or learning objectives?
3. Would you do anything different to help learn the material?
4. What grade would you give yourself?
5. What could you do to improve?

At the end of every unit or chapter, the day of the summative assessment, the students were asked to fill out a short survey when completed with the exam. The first five questions were set on a Likert Scale and the same each week, 1 representing Strongly Disagree and 5 representing Strongly Agree. These were asked to see how the students felt about their final performance with the material and how well the material aligned with the standards they have been reflecting on.

**End of Chapter Exam Questions:**
1. I felt confident before beginning the exam.
2. The study material provided before the exam was thorough enough to prepare me for the exam.
3. I believe the questions asked in the exam accurately reflected the learning objectives encompassing this material.
4. I needed more time to prepare for this exam.
5. I will need to re-study the material and retake this exam to master the learning objectives covered by this material.

In addition to these questions every week, the second set of questions covered how confident the students felt about mastering each of the specific standards in that unit. The questions were also in the Likert form with the same scale as the previous questions. For example, for the Dynamic Forces Unit, all of the Arizona State Science
Standards that applied to the topic of dynamic forces were used and the students ranked themselves.

**Standard Confidence Questions:**

1. I felt confident about the performance objective that states: Using Newton’s 2nd Law of Motion, analyze the relationships among the net force acting on a body, the mass of the body, and the resulting acceleration.
2. I felt confident about the performance objective that states: Represent the force conditions required to maintain static equilibrium.
3. I felt confident about the performance objective that states: Describe the nature and magnitude of frictional forces.

The final set of questions the students were asked were given as free response questions that the student had to give a written response to. These were used to gauge the different understanding of the standards and learning strategies each individual student used.

**Free Response Questions:**

1. How did your understanding of the standards addressed in this unit progress throughout the weeks leading toward the assessment?
2. What were some of the learning strategies you applied throughout the previous weeks to prepare you for the assessment?
3. What classroom methods helped you work toward mastery of the standards addressed?
4. What kind of dialogue (if any) did you have with yourself or others to help you reach achievement in the standard?
5. Are there any other questions or comments you have about these standards or the unit covering them that have not been addressed yet?
Timeline

1. Treatment

The treatment was performed during the first semester for units 1 through 7 for the physics classes, and for first and second semester for unit 1 through unit 9 for the Algebra 1 classes. The emphasis was on making students aware of the standards they were to master during a unit of instruction. Each week students were asked to make written reflections in a composition book. At the end of each unit, students filled out a Google Form that asked students various questions were they could rank their understanding, and perceived standards mastery using a Likert Scale. The Google Form also had free response questions that prompted students to reflect on what learning strategies, classroom methods, peer dialogues, and anything else that helped them learn.

2. Assessment

The FCI was given to the students at the beginning of the Force unit as pre-assessments. The MCI and ACI were given at the beginning of the year in the Algebra 1 classes as pre-assessments. The FCI was given again at the end of the first semester, and the MCI and ACI were given again at the end of the school year. First semester district CRT scores were collected for Investigators 1 and 2 and Contrast Group 2. AZMerit Test scores were collected for Investigator 3 and the corresponding Contrast Groups 3 and 4.
Results

The normalized gains measured by each investigator show that student reflections had a positive influence on student achievement as measured using pre and post data for the FCI, MCI, and ACI tests. However, the data statistically showed the scores had no significant difference between the gains from the treatment groups to the contrast groups. First, the physics treatment was looked at as a whole, and then separated for further analysis by looking at each investigator individually.

FCI

The FCI was used to monitor student understanding of motion and forces. Based on a non-directional paired samples t-test at alpha = 0.05, the null hypothesis was rejected that the population mean of the pretest takers is equal to the population mean of the posttest takers, \( t(50) = -9.11, p<0.000 \). This concluded that there was a significant difference between students understanding of motion and forces as measured on the FCI before the treatment (Mean = 7.29, SD = 3.02) and after the treatment (Mean = 13.1, SD = 5.50). This test shows a 95% confidence that the interval -7.04 and -4.49 contains the true population mean difference. The correlation was 0.571.

<table>
<thead>
<tr>
<th>FCI Pretest Skyline this year</th>
<th>FCI Posttest Skyline this year</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Frequency</strong></td>
<td><strong>Frequency</strong></td>
</tr>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>5</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>11</td>
</tr>
<tr>
<td>13</td>
<td>15</td>
</tr>
<tr>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td><strong>Number Correct</strong></td>
<td><strong>Number Correct</strong></td>
</tr>
<tr>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>19</td>
<td>22</td>
</tr>
<tr>
<td>25</td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Skyline FCI Pretest</th>
<th>Skyline FCI Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Mean</td>
</tr>
<tr>
<td>7.29</td>
<td>13.1</td>
</tr>
<tr>
<td>Std Dev</td>
<td>Std Dev</td>
</tr>
<tr>
<td>3.02</td>
<td>5.5</td>
</tr>
<tr>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>51</td>
<td>51</td>
</tr>
</tbody>
</table>

*Figure 1: Comparison of histograms for treatment FCI pretests and posttests*
The above histograms show that the pre FCI scores are skewed right, and that the post FCI scores are centralized. The t-test confirms that these two populations are different and that students that received the treatment showed significant gains in their understanding of motion and forces.

The gains by the treatment group with that of the first contrast group were compared. This group consisted of the students taught by Investigator 1 and Investigator 2 the previous year. It was first checked to see if this contrast group was from the same population of students. Based on a non-directional independent samples t-test at alpha = 0.05, the null hypothesis is rejected that the population mean of the treatment group is equal to the population mean of Contrast Group 1, \( t(115) = 3.03, p<0.003 \). This concluded that there was a significant difference between the treatment sample (Mean = 7.29, SD = 3.02) and contrast group 1 (Mean = 9.08, SD = 3.25). This test shows a 95% confidence that the interval 0.617 and 2.95 contains the true population mean difference.

The gains by the treatment group with that of the second contrast group were compared. This group consisted of students from another school that is similar in size and demographics as the treatment group. It was first checked to see if this contrast group was from the same population of students. Based on a non-directional independent samples t-test at alpha = 0.05, the null hypothesis that the population mean of the treatment group is equal to the population mean of Contrast Group 2 failed to be rejected, \( t(108) = -1.056, p=0.293 \). This concluded that there was no significant difference between the treatment sample (Mean = 7.29, SD = 3.02) and Contrast Group 2 (Mean = 6.66, SD = 3.23). This test shows a 95% confidence that the interval -1.82 and 0.555 contains the true population mean difference.
Effects of Student Reflections

*Figure 2: Comparison of histograms for FCI pretests*

Inspection of the above histograms shows that the distribution of scores among the treatment group and Contrast Group 1 look similar, but their averages are much different. The histograms also show that while the distribution of scores among the treatment group and Contrast Group 2 look different, the averages are quite similar. This confirms what the t-tests show, that the treatment group and Contrast Group 1 are not from the same population, yet the treatment group and Contrast Group 2 are from the same population.
Effects of Student Reflections

Following the comparison of the pretest scores, the normalized gains made on the FCI by the treatment and contrast groups were then analyzed. The physics treatment group had a gain of $<g>_{51T} = 0.260 \pm 0.217$. This was larger than Contrast Group 1 $<g>_{66T} = 0.190 \pm 0.179$, but less than that of Contrast Group 2 $<g>_{53T} = 0.416 \pm 0.294$.

![FCI Gains](image)

*Figure 3: Histogram showing average FCI gains*

One method of comparing the significance of normalized gains is comparing the number of standard deviations between the gain of the treatment group and that of the contrast groups.

\[
\frac{\text{Treatment} <g> - \text{Contrast} <g>}{\text{Standard Deviation of Treatment} <g>}. \]

The average gain of the treatment group was 0.32 standard deviations above the average gain of contrast group one. This is significant as this means that the treatment group performed better than Contrast Group 1. However, the average gains of the treatment group were not nearly as significant as Contrast Group 2. The treatment group was 0.73 standard deviations below that of Contrast Group 2.
Because the gains of Investigators 1 and 2 and those of Contrast Groups 1 and 2 were being compared, a one-way ANOVA test was run. This test makes multiple comparisons between groups. There was a statistically significant difference between groups as determined by one-way ANOVA (F(2,167) = 14.3, p=0.000). A Tukey post hoc test revealed that the average gain experienced by the treatment group (0.260 ± 0.217, p=0.233) was not statistically higher than that of Contrast Group 1 (0.190 ± 0.179). Yet the treatment group was statistically lower than that of Contrast Group 2 (0.416 ± 0.294, p=0.002).

Another method of comparing how well students in the treatment group did was to look at the fall semester CRT scores. This district wide criterion referenced test was given at the end of the fall semester. In looking at the CRT scores this year, it was interesting how closely they were aligned, especially since Contrast Group 2 had higher gains on the FCI test than Investigators 1 and 2. This shows that the treatment group for Investigators 1 and 2 performed better on the CRT test than Contrast Group 2.

![Figure 4: Histogram showing average CRT Scores this year](image)
independent samples t-test at alpha = 0.05, the null hypothesis that the population mean of the Investigator 1’s treatment group is equal to the population mean of Investigator 2’s treatment group failed to be rejected, t(56) = 0.436, p=0.664. This concluded that there was no significant difference between Investigator 1’s Treatment group (Mean = 71.2, SD = 13.3) and Investigator 2’s Treatment Group (Mean = 72.7, SD = 13.0). This test shows a 95% confidence that the interval -5.53 and 8.61 contains the true population mean difference.

Since both investigators’ treatment groups were from the same population, it was worthwhile to know if their classes from last year were also from the same population. Based on a non-directional independent samples t-test at alpha = 0.05, the null hypothesis that the population mean of the Investigator 1’s Contrast Group is equal to the population mean of Investigator 2’s Contrast Group failed to be rejected, t(67) = 1.17, p=0.244. This concluded that there was no significant difference between Investigator 1’s Contrast Group (Mean = 68.9, SD = 14.6) and Investigator 2’s Contrast Group 1 (Mean = 73.2, SD = 13.5). This test shows a 95% confidence that the interval -3.03 and 11.7 contains the true population mean difference.

Since the treatment groups were from the same population they were combined. The investigator’s contrast groups were also combined because they also were from the same population. To compare these with Contrast Group 2 a one-way ANOVA test was run. There was no statistically significant difference between the groups as determined by a one-way ANOVA (F(2,169) = 2.46, p=0.088). A Tukey post hoc test revealed that the average score of the treatment group (71.8 ± 13.1, p=0.799) was not statistically higher than that of Contrast Group 1 (70.3 ± 14.3). It also was not statistically different than that of Contrast Group 2 (66.0 ± 12.7, p=0.08).

**Investigator 1 Report**

To further analyze, the two investigator’s data were separated and analyzed them against each of the contrast groups individually. For Investigator 1’s Treatment group, a comparison of pre and post test scores on the FCI was made. Based on a non-directional paired samples t-test at alpha = 0.05, the null hypothesis that the population mean of the pretest takers is equal to the population mean of the posttest takers was
rejected, $t(30) = -6.706$, $p=0.000<0.05$. This concluded that there was a significant difference between student understanding of forces as measured on the FCI before the treatment (Mean = 6.903, SD = 2.561) and after the treatment (Mean = 11.677, SD = 4.672). This test shows a 95% confidence that the interval -6.228 and -3.320 contains the true population mean difference. The correlation is 0.530.

![Investigator 1 FCI Pretest](image1)

### Investigator 1 FCI Pretest
- **Mean**: 6.9
- **Std Dev**: 2.56
- **N**: 31

![Investigator 1 FCI Posttest](image2)

### Investigator 1 FCI Posttest
- **Mean**: 11.7
- **Std Dev**: 4.67
- **N**: 31

*Figure 5: Comparison of histograms for Investigator 1’s treatment FCI pretests and posttests*

The above histograms show that the pre FCI scores are skewed to the right, and that the post FCI scores are still skewed to the right but more centralized. The t-test confirms that these two populations are different and that students that received the treatment showed gains in their understanding of motion and forces.

Next, a further analysis was conducted to compare Investigator 1’s treatment group with his own students in Contrast Group 1. This creates a direct comparison of the instructor from one year to the next. First, it was checked to see if this part of the contrast group was from the same population of students as Investigator 1. Based on a non-directional independent samples t-test at alpha = 0.05, the null hypothesis that the
population mean of the treatment group is equal to the population mean of the Contrast Group 3 was rejected, $t(74) = 2.266, p = 0.026$. This concluded that there was a significant difference between the treatment sample (Mean = 6.90, SD = 2.56) and Contrast Group 3 (Mean = 8.38, SD = 2.93) pretest populations. This test shows a 95% confidence that the interval 0.178 and 2.771 contains the true population mean difference.

Although the pretest populations for the treatment group and contrast group are not comparable, there was still a much larger normalized gain between the two groups as seen in Figure 6.

![FCI Gains](image)

*Figure 6: Average FCI percentage of normalized gains for Investigator 1 and 2 compared to Contrast Group 1 split by each Investigator*

Based on a non-directional independent samples t-test at alpha = 0.05, the null hypothesis that the population mean of the treatment group is equal to the population mean of the Contrast Group 1 failed to be rejected, $t(74) = -1.083, p = 0.282$. This concluded that there was no significant difference between the normalized gains for treatment sample (Mean = 0.166, SD = 0.024) and Contrast Group 1 (Mean = 0.209, SD = 0.182). This test shows a 95% confidence that the interval -0.122 and 0.036 contains the true population mean difference.
A final comparison could be made between Investigator 1 and Contrast Group 2; however, due to the large differences between the two instructors, it was found to be an unfair comparison. The instructor for Contrast Group 2 has several more years of teaching and modeling experience than that of Investigator 1.

**Investigator 2 Report**

For Investigator 2’s treatment group, a comparison of pre and post test scores was made. Based on a non-directional paired samples t-test at alpha = 0.05, the null hypothesis that the population mean of the pretest takers is equal to the population mean of the posttest takers was rejected, \( t(19) = -6.549, p<0.000 \). This concluded that there was a significant difference between student understanding of forces as measured on the FCI before the treatment (Mean = 7.9, SD = 3.61) and after the treatment (Mean = 15.2, SD = 6.11). This test shows a 95% confidence that the interval -9.63 and -4.97 contains the true population mean difference. The correlation is 0.578.

![Figure 7: Comparison of histograms for Investigator 2 treatment FCI pretests and posttests](image-url)
For Investigator 2, a comparison of gains by the treatment group with that of Contrast Group 1 from last year was made to see if the students come from the same population. Based on a non-directional independent samples t-test at alpha = 0.05, the null hypothesis that the population mean of the treatment group is equal to the population mean of the contrast group one for the pretest failed to be rejected, t(39) = 2.42, p=0.020. This concluded that there was a significant difference between the treatment sample (Mean = 7.9, SD = 3.61) and Contrast Group 1 (Mean = 10.6, SD = 3.46). This test shows a 95% confidence that the interval .438 and 4.90 contains the true population mean difference. A comparison of posttest scores was also made. Based on a non-directional independent samples t-test at alpha = 0.05, the null hypothesis that the population mean of the treatment group is equal to the population mean of the contrast group one for the posttest failed to be rejected, t(39) = -0.005, p=0.996. This concluded that there was not a significant difference between the treatment sample (Mean = 15.2, SD = 6.11) and Contrast Group 1 (Mean = 15.2, SD = 4.97). This test shows a 95% confidence that the interval -3.52 and 3.50 contains the true population mean difference. Even though the populations were different for the pretests, they ended up being the same for the posttests.

Investigator 2’s treatment group had a gain of \( <g>_{20T} = 0.34 \pm 0.25 \). This was larger than Investigator 2's Contrast Group 1 \( <g>_{21T} = 0.24 \pm 0.21 \), but less than that of Contrast Group 2 \( <g>_{53T} = 0.42 \pm 0.29 \). Investigator 2’s treatment group contrast group one gains are also shown in Figure 6. Clearly the gains made this year by Investigator 2’s treatment group were higher than the gains their contrast group made last year.

In the comparison of Investigator 2’s treatment group with that of his Contrast Group 1 from last year, there is a noticeable increase in FCI gains with Investigator 2’s treatment group. Even though Contrast Group 2 has higher gains than Investigator 2’s treatment group, that comparison is not as important because there are so many external factors that make Contrast Group 2 different. It is taught by a different teacher at a different school who may very well be a better modeling physics instructor. The best comparison to be made here is that Investigator 2’s treatment group this year did in fact have higher gains than his contrast group from last year. This increase in gains can
be attributed to the success of the action research study. By implementing student reflections in the classroom, Investigator 2’s students had higher gains on the FCI.

Investigator 3 Report

MCI

The MCI was the first measurement of students’ understanding of mathematics. Based on a non-directional paired samples t-test at alpha = 0.05, the null hypothesis that the population mean of the pretest takers is equal to the population mean of the posttest takers was rejected, t(49) = -3.255, p=0.002<0.05. This concluded that there was a significant difference between student understanding of mathematics as measured on the MCI before the treatment (Mean = 9.36, SD = 3.99) and after the treatment (Mean = 11.52, SD = 3.99). This test shows a 95% confidence that the interval -3.493 and -0.827 contains the true population mean difference. The correlation was 0.308.

![Figure 8: Comparison of histograms for treatment MCI pretests and posttests](image)

<table>
<thead>
<tr>
<th>MCI Pretest Treatment</th>
<th>MCI Posttest Treatment</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td><strong>Mean</strong></td>
</tr>
<tr>
<td>9.36</td>
<td>11.5</td>
</tr>
<tr>
<td><strong>Std Dev</strong></td>
<td><strong>Std Dev</strong></td>
</tr>
<tr>
<td>3.99</td>
<td>3.99</td>
</tr>
<tr>
<td><strong>N</strong></td>
<td><strong>N</strong></td>
</tr>
<tr>
<td>50</td>
<td>50</td>
</tr>
</tbody>
</table>

The above histograms show that the pre MCI scores are skewed right, and that the post MCI scores are skewed left. The t-test confirms that these two populations are
different and that students that received the treatment showed gains in their understanding of mathematics.

Gains by the treatment group with that of Contrast Group 3 were compared. It was first checked to see if this contrast group was from the same population of students. Based on a non-directional independent samples t-test at alpha = 0.05, the null hypothesis that the population mean of the treatment group is equal to the population mean of Contrast Group 3 was rejected, \( t(82) = 2.310, p = 0.023 \). This concluded that there was a significant difference between the treatment sample (Mean = 9.36, SD = 3.99) and Contrast Group 3 (Mean = 11.38, SD = 3.86). This test shows a 95% confidence that the interval 0.281 and 3.76 contains the true population mean difference.

![Histograms for MCI Pretests](image)

**Figure 9: Comparison of histograms for MCI pretests**

Inspection of the above histograms shows that the distribution of scores among the treatment group and Contrast Group 3 look similar, but the average scores of the treatment group and contrast group is less similar. This is contrary to the t-tests which showed the treatment group and Contrast Group 3 are from different populations.
Following the treatment, Investigator 3 analyzed the normalized gains on the MCI for his students and Contrast Group 3. Based on a non-directional independent samples t-test at alpha = 0.05, the null hypothesis that the population mean of normalized gains of the treatment group is equal to the population mean of the contrast group failed to be rejected, t(82) = -0.807, p = 0.422. This concluded that there was no significant difference between the normalized gains of the treatment group (Mean = 0.0611, SD = 0.5603) and Contrast Group 3 (Mean = -0.0236, SD = 0.29562). This test shows a 95% confidence that the interval -0.293 and 0.124 contains the true population mean difference.

The T-Test indicates there is not a significant different between the gains on the MCI for Investigator 3 and Contrast Group 3; however, the average normalized gains appear to be drastically different. This can be attributed to a high standard deviation of MCI gains. There was also one student who scored 21 out of 23 points on the pretest, and 15 out of 23 on the posttest, giving a normalized gain of -300%. While this data cannot be omitted, this outlier dramatically increased the standard deviation of the average gains.
The ACI was also used to measure students’ understanding of algebra as measured by the ACI. Based on a non-directional paired samples t-test at alpha = 0.05, the null hypothesis that investigator’s population mean of the pretest takers is equal to the population mean of the posttest takers was rejected, \( t(49) = -4.953, p=0.000 < 0.05 \). This concluded that there was a significant difference between student understanding of algebra as measured on the ACI before the treatment (Mean = 6.86, SD = 2.416) and after the treatment (Mean = 9.10, SD = 2.757). This test shows a 95% confidence that the interval 3.15 and 1.33 contains the true population mean difference. The correlation was 0.241.

The above histograms show that the pre ACI scores are skewed right, and that the post ACI scores are generally more central. The previous t-test confirms that these two populations are different and that students that received the treatment showed gains in their understanding of algebra.

The gains made by the treatment group and Contrast Group 3 were then compared. The contrast group was then checked to see if the students were from the
same population as the treatment group. Based on a non-directional independent samples t-test at alpha = 0.05, the null hypothesis that the population mean of the treatment group is equal to the population mean of the Contrast Group 3 failed to be rejected, t(83) = 1.061, p = 0.306. This concluded that there was no significant difference between the treatment sample (Mean = 6.34, SD = 2.06) and Contrast Group 3 (Mean = 6.86, SD = 2.42). This test shows a 95% confidence that the interval -1.51 and 0.480 contains the true population mean difference. Because there is no significant difference found in the pretest scores on the ACI, it can be concluded that Contrast Group 3 can be used as a fair comparison to the students taught by Investigator 3.

![Histogram Comparison](image)

**Figure 12: Comparison of histograms for ACI pretests**

Inspection of the above histograms shows that the distribution of scores among the treatment group and Contrast Group 3 look similar, but the average scores of the treatment group and contrast group is similar. This confirms what the t-tests show, that the treatment group and Contrast Group 3 are from the same population.

Following the treatment, Investigator 3 analyzed the gains on the ACI for his students and Contrast Group 3’s students. Based on a non-directional independent
samples t-test at alpha = 0.05, the null hypothesis that the population mean of the treatment group is equal to the population mean of the Contrast Group 3 failed to be rejected, $t(83) = -1.28, p = 0.204$. This concluded that there was no significant difference between the treatment sample (Mean = .1254, SD = 0.20349) and Contrast Group 3(Mean = 0.0699, SD = 0.18642). This test shows a 95% confidence that the interval -0.142 and 0.0307 contains the true population mean difference.

**Figure 13: Comparison of average ACI gains**

Similarly to the MCI gains, the ACI average normalized gains seem drastically different even though the T-Test in Appendix 27 shows they are not significantly different.

**AZMerit**

The final method for comparing students’ understanding of mathematics was using the treatment group’s AZMerit scores. This state wide test was given at the end of the spring semester. In looking at the AZMerit scores this year, it was interesting how
closely they were aligned, especially because the treatment group had higher gains on the MCI and ACI. This shows that Contrast Groups 3 and 4 performed better on the AZMerit than the treatment group.

![Figure 14: Comparison of average AZMerit Scores for the treatment group and Contrast Groups 3 and 4](image)

To compare AZMerit scores a one-way ANOVA test was run. There was no statistically significant difference between the groups (F(2,217) = 0.527, p=0.591). A Tukey post hoc test revealed that the average score of the treatment group (3663.3 ± 26.2, p=0.925) was not significantly lower than that of Contrast Group 3 (3665.1 ± 26.4). It also was not statistically different than that of Contrast Group 4 (3667.9 ± 27.9, p=0.586).

**Qualitative Data**

The qualitative data that was collected came from two sources. First, a score was assigned from the rubric for each student’s weekly reflections. All the weekly reflection scores were averaged and the student was assigned a rank from 0 to 4 based on the quality of their reflections. The other source of qualitative data came from the students’
reflections at the end of every test about their confidence levels for each standard and overall performance on the assessment.

The weekly reflections were used to have the students take time each week to reflect on the progress they have been making as the material was covered in each unit leading up to the summative assessment. They were asked to reflect on the same list of questions each week, as stated in the qualitative analysis procedure. The following shows the distribution of scores for the number of students at each of the reflection levels.

![Reflection Score Distribution for Physics](image)

*Figure 15: The distribution of weekly reflection scores amongst the physics students*

Most students seemed to have minimal reflections as a large majority of the students received a 2 or 3. The students were able to provide a reflection of some sort at the minimum, commenting on their understanding, or lack thereof, of the material in the unit. This is what gave most students the score of a 2. Many of the students also took it a step further and were able to comment on the standards and how they were specifically performing on each one. Very few students were able to get a 4, as most of them failed to comment on how they could change to improve their self-assessment.

Some example comments that students were making to receive a level one reflection rating sounded like, “I can only think of (music). But wowee. I know this stuff pretty well.” The student dismisses the reflection and does not make any effort toward mentioning their progress. Another student who received a one on a reflection
commented, “The rate of change in quality is a little understandable.” This comment attempts to mention the standard, however they fail to elaborate on what they mean by ‘a little understandable’. There were very few students who received a one for their reflection, and only one student received an overall grade of a one.

Moving up the scale, the students who received a two for a grade made comments like, “I learned every part of the objective. I’m not too good with acceleration yet, but I learned a lot more about it. The three are each analyzed fairly similarly. These journal entries are starting to get a little tedious.” Notice that the student was able to comment on their progress upon mastering the topic of acceleration, stating they have ‘learned a lot more’ which shows they have recognized progress within their achievement. Another interesting comment was made by the student, stating the journal entries were already getting ‘tedious’. This comment was made in the third weekly reflection of the entire year. Many students’ scores also showed a significant decline as the weeks went on in the quality of their reflections. Many of the students’ first and second reflections were the best, earning them a 3 or 4. By the end, many of them dropped to range between 2 and 3.

Some students could stay consistent were their score and continually make deep and meaningful connections about their academic achievement toward mastering the standards. A handful of students were able to do this and receive an overall score of 4 for their weekly reflections. The following is an example of a student reflection that received a score of 4:

“I felt very confident going into the test, and reviewing. Despite this though, i did not get the grade I was hoping for, and actually did much worse. This shows that I have much to improve on, and that I must do a much better job on studying and reviewing. Hopefully I will be able to fully understand what I did wrong, so that I can correct and change things accordingly. The only part of the test that confused me a big was indeed the motion maps of velocity and acceleration, meaning I should seek help on them.”

The student here is able to reflect on their test grade and how they felt they did compared to how they actually performed. They then go on to mention what they
believe to be an area of struggle and ends with stating they ‘should seek help on them’. This student shows a thorough understanding of what it means to make a meaningful reflection on improving their achievement. Another example from a different student made sure to hit all the necessary points and received a 4 while keeping it more concise:

“Balanced Forces. After reading the standards I have met at least two of them and have been taking notes to understand everything and remember it. So far I think I am doing pretty well with this section and plan to do better with more practice.”

Although the unit seems to have just started, the student is able to interpret the standards and make assumptions to assess their performance thus far. They are able to comment on how they are currently mastering the standards, as well as what they need to do to continue to do so.

Further analyzing the reflection data, it can be seen how the reflections affected the normalized gains of the students.

![Average Normalized FCI Gains v. Reflection Score](chart.png)

*Figure 16: Average normalized gains of students at each reflection level*

Figure 16 shows that no matter how well a student reflected, their FCI gains did not correlate. There were students who received a 4 on their weekly reflections but received a 0% gain from pre to post tests. And the one student who received a 1 for a
reflection score had a percent gain of 25%. These represent the averages of all the students who received each of the reflection scores so the students within each ranking had a large distribution of FCI normalized percentage gains.

In addition to the weekly reflections, students also reflected after the summative assessment at the end of each unit over several Likert Scale questions, as well as some free response questions they had to respond to. The first set of questions asked more general questions to the students about their confidence level and the overall consensus of the exam itself. Below is a table of the averages of the students’ responses for each questions during each unit.

<table>
<thead>
<tr>
<th>Average Likert Scale Response</th>
<th>Constant Velocity</th>
<th>Constant Acceleration</th>
<th>Balanced Forces</th>
<th>Dynamic Forces</th>
<th>Momentum</th>
<th>Uniform Circular Motion</th>
</tr>
</thead>
<tbody>
<tr>
<td>I felt confident before beginning the exam</td>
<td>3.8</td>
<td>3.8</td>
<td>3.7</td>
<td>3.3</td>
<td>3.6</td>
<td>3.2</td>
</tr>
<tr>
<td>The study material provided before the exam was thorough enough to prepare me for the exam.</td>
<td>4.3</td>
<td>4.4</td>
<td>4.0</td>
<td>4.0</td>
<td>4.0</td>
<td>3.3</td>
</tr>
<tr>
<td>I believe the questions asked in the exam accurately reflected the learning objectives encompassing this material.</td>
<td>4.4</td>
<td>4.6</td>
<td>4.1</td>
<td>4.0</td>
<td>4.2</td>
<td>3.8</td>
</tr>
<tr>
<td>I needed more time to prepare for this exam.</td>
<td>2.4</td>
<td>2.5</td>
<td>2.9</td>
<td>3.0</td>
<td>2.9</td>
<td>3.3</td>
</tr>
<tr>
<td>I will need to re-study the material and retake this exam to master the learning objectives covered by this material.</td>
<td>2.3</td>
<td>1.9</td>
<td>2.6</td>
<td>3.0</td>
<td>2.6</td>
<td>2.9</td>
</tr>
</tbody>
</table>

Table 3: Likert Scale questions students answers post unit exam
Nothing unusual pops out when initially looking at the data other than a slight decline in confidence level as the class approached the Dynamic Force unit. Most students agreed the study material was a thorough enough and a fair representation for the questions that were asked on the exam. More students leaned toward disagreeing when asked if they needed more time prepare for the exam or needing to retake it.

The second set of questions were specific to each standard within each unit. Below are a set of example graphs that were taken from the Investigator 1 and 2 after the Dynamic Forces unit covering the three standards that apply.

![Number of Students at Each Confidence Level for Standard 1 in Dynamic Forces](image)

*Figure 17: Confidence level of students mastering Standard 1 of the Dynamic Forces unit*
Figure 18: Confidence level of students mastering Standard 2 of the Dynamic Forces unit

Figure 19: Confidence level of students mastering Standard 3 of the Dynamic Forces unit
It was a common trend that most students felt neutral to strongly agree for feeling confident about each of the performance objectives.

The last set of questions asked the students to respond to a set of questions that had them give written responses. For the first question regarding the student’s confidence level as the unit progressed, almost all students responded that their knowledge and ability continued to increase toward the assessment. Some student responses included, “My understanding of the standards increased as we progressed toward the test,” or, “We started with a lab that helped cement the ideas into the mind and then moved to practice sheets.” The second and third question were both in reference to different learning strategies and methods that were applied which had a wide variety of responses including working with peers, worksheets, labs, homework, studying notes, paying attention, and practice. Finally, the students were asked about the type of dialogue they engaged in to help work toward the mastery of the standards and most responded with asking questions in class or talking to their peers about how to solve something. These were all examples coming from the Dynamic Forces unit and there was no reference to the act of reflecting helping the students. However, the main goal was to have students thinking about what they do and how they work toward reaching that achievement, which was the main purpose of these questions. Having the students think about what they do and what they can do to get to their goal is what was hoped would help them reach it.

**Investigator 3**

The following shows the distribution of scores given for weekly reflections for the number of students at each of the reflection levels.
Several students showed mid-level reflections, as the vast majority received 2’s and 3’s. The students were able to provide a reflection of some sort commenting on their understanding, or lack thereof, of the material in the unit. This is what gave many students the score of a 2. Many students took it a step further and commented on the standards and how they were specifically performing on each one. Very few students were able to get a 4, as most of them failed to comment on how they could change to improve their self-assessment.

Histograms of the scores students gave themselves using the Likert Scale developed for the standards covered by the Chapter 2 test in Algebra 1, which covered solving linear inequalities are below. Standard One stated, “Create inequalities in one variable, and use them to solve problems”, and Standard Two stated, “Solve linear inequalities in one variable.”
Effects of Student Reflections

Figure 21: Confidence level of students' mastering Standard 1 of the Solving Linear Inequalities Chapter

Figure 22: Confidence level of students' mastering Standard 2 of the Solving Linear Inequalities Chapter
By further analyzing the data, it can be determined how reflections affected gains on the MCI.

![Average Normalized MCI Gains v. Reflection Score](image)

*Figure 23: Average normalized gains of students at each reflection level*

Figure 23 displays the weekly reflection scores assigned to students compared to the average gains on the MCI for each score. It can be shown that students who reflected better had higher average gains on the MCI. The students who received low reflection scores, 1’s and 2’s achieved negative gains on the MCI, whereas the students who received 4’s achieved almost a twenty percent average gain.
Conclusion

In the education field of research, there have been many studies that show that students who reflect on their academic progress show signs of increased academic achievement. There is also the argument that traditional based grading does not reflect a student’s ability to progress and learn the material, so many schools are adopting the concept of standards-based grading -- with the school making this change and adopting this new idea for students to have the opportunity to master the material at any time, allowing an extension on homework and practice and the ability to retake any assessment later for full credit. The research was an attempt to combine these two ideas, along with Modeling Instruction, to have students reflect on the standards in an effort for them to see the goal ahead and make them think about what they can do to reach that goal.

The investigators used several forms of results taken from different pre- and post-tests such as the FCI, MCI, and ACI along with the final district exams. It was found the data to show no statistically significant improvement from the investigator groups against the contrast groups for any of the results. However, many of the average normalized gains and performance averages showed a higher achievement than the contrast groups. When looking at the treatment group against the two contrast groups in the physics classroom, the pre-test scores were very different with the students in the treatment sample scoring much poorer than the year before. It was shown the Investigators 1 and 2 to have a similar population with the second contrast group; however, they showed a much larger gain overall. When comparing the normalized gains, it showed the contrast group had a statistically higher growth than the treatment group. And even though the treatment group showed higher gains than the first contrast group, the statistics showed no significant difference. The last result was analyzing the final district-administered exam given at the end of each semester. Both investigators scored higher than the second contrast group; however, yet again, it was not a statistically significant difference.

Since the treatment group for physics encompassed two separate investigators, the data was split and analyzed individually with the contrast group. For all the groups, it was found there to be a significant difference when comparing pre- to post-test scores
showing growth for all students. The gains were what was wanting to be compared with the contrast groups, alas each showed no significant difference for any of the groups, even though when looking at the raw data, both investigators showed a higher percentage gains.

For math, the third investigator yielded similar results as the physics instructors. The pre- and post-test exam gains were positive for all groups; however when comparing the gains with the contrast groups, none showed a statistically significant difference. Even though there was no statistical difference, the raw percentage score showed the ACI had a positive 6% gain for the investigator and a negative 2% loss for the contrast group. Finally, with the final district exam for math, the average scores were too close to show any difference between the investigator and contrast groups.

There are many aspects about the research that could be changed if the project were to be done again. First, the sample size for the investigator and contrast groups would need to be larger. The study was limited to only three sections of general physics at the school to compare. Another issue was the population of the pretest takers being similar. Several of the groups had significantly different pretest scores, placing the students in different sets of populations, making it harder to compare the normalized gains. Lastly, the reflections seemed to become an overbearing assignment that many students began to lose focus on. It might be reconsidered having weekly reflections and have small checkpoints randomly throughout each unit. Also, many students’ free response reflections didn’t have enough comments to create accurate and efficient reflecting practices.

Although the data did not give statistically significant gains for student achievement, reflections are still considered to be a beneficial part about a student’s education and learning and more research should be done to address this.
Implications for Further Research

This research was used to investigate whether student reflections helped improve their understanding of physics and mathematics. It did however open several more questions. Something the group had partially considered at the start of the research was how well the students know how to reflect. The investigators showed exemplar reflections at the beginning of the research, but if the students’ reflections did not always meet the quality of the reflections shown, the treatment may not have been as effective as it could be.

Something else to be investigated would be if the spacing between reflections improved the resulting gains. Students verbally displayed their distaste for reflecting after several weeks of reflections. Maybe if the students reflected less frequently, and with more time in between, they would put more thought into the reflections.

It was also noted after the treatment that the teacher who taught Contrast Groups 1 and 2 had been teaching for several more years and had more modeling experience than Investigators 1 and 2, who were compared to the contrast group teacher. Maybe if the investigators had more similar years of experience to the contrast groups' teacher, the results would have been different.

The treatment and contrast groups also had far fewer students to compare. It would have been preferred to complete this treatment with more students, or a larger sample size.

Further research could also be done comparing the gains of the algebra 1 students and physics students. An investigation could be carried out to test if reflections had more of an effect with physics students than with math students. The age and grade level could also be checked. Do younger students reflect better or worse than older students, and does that affect the gains achieved by the students?

Works Cited

Effects of Student Reflections


Ballard, A. C. (2010). *Student achievement, personal achievement goal orientations, and perceptions of classroom goal structures in a standards-based reporting system.* University of South Carolina.


a relationship?. Advances in Health Sciences Education, 16(4), 529.