STANDARDS-BASED FEEDBACK IN 7TH GRADE MATH

A Dissertation

Presented to

The Faculty of the Education Department

Carson-Newman University

In Partial Fulfillment

Of the

Requirements for the Degree

Doctor of Education

Karoline Michelle Stache

May 2019
Dissertation Approval

Student Name: Karoline M. Stache / 0329101

Dissertation Title: Standards-Based Feedback in 7th Grade Math

This dissertation has been approved and accepted by the faculty of the Education Department, Carson-Newman University, in partial fulfillment of the requirements for the degree, Doctor of Education.

Dissertation Committee:

Signatures: (Print and Sign)

Dissertation Chair: Dr. Tammy Barnes

Methodologist Member: Dr. P. Mark Taylor

Content Member: Dr. Patricia Murphree

Approved by the Dissertation Committee Date: March 14, 2019
Abstract

The purpose of this experimental, quantitative study was to examine how feedback in the form of standards-based grading impacted student learning. Two aspects of grading practices were the focus of this study. First, students received standards-based scores and tracked their progress on each standard assessed. Then, the students reflected upon that progress by answering three questions regarding their feedback. The theoretical framework was based upon Hattie’s work with visible learning and feedback. The study was conducted in 6 seventh grade math classes in a rural middle school in East Tennessee during the 2018-2019 school year. There were 3 levels of classes studied: standard, advanced, and elite. Half of the classes formed the control group and the other half were in the experiment group. Each group contained one of each level of class. All students took a pre- and post-test on the MasteryConnect platform, and the post-test gains were analyzed using a variety of t-tests to determine statistical significance between the two groups studied. Overall, this study revealed post-test gains for the tested and control groups showed no statistical difference. However, differences were apparent when the different levels of math classes within the tested group were compared. The information from this study will help classroom teachers to improve pathways of feedback and diagnose areas of need in the academic classroom.
Acknowledgements

Thank you to my family for all the love and support you have provided to me throughout this process.

Thank you to Dr. Tammy Barnes for your time and encouragement you gave me this past year during this dissertation journey.

Thank you to Dr. Patrick Mark Taylor for talking through any and all issues that arose during my experimentation.

Thank you to Dr. Patricia Murphree for your guidance throughout my time at Carson Newman.

Thank you to all the Carson Newman staff for making this journey a blessing and an honor to complete.
Dedication

This dissertation is dedicated to my family. To my parents, who have always encouraged me to strive for more, and instilled in me the work ethic to see my goals to the end. To my brother and sister, who have shown me what that work ethic looks like. To my baby niece, whose constant smile and energy motivated me through to the finish.
Table of Contents

Abstract........................................................................................................................................... ii
Acknowledgements.......................................................................................................................... iii
Dedication........................................................................................................................................... iv
Table of Contents............................................................................................................................ v
List of Figures & Tables ......................................................................................................................... viii
CHAPTER 1: INTRODUCTION & BACKGROUND........................................................................ 1
  Statement of the Problem ............................................................................................................... 2
  Purpose/Significance of Study ........................................................................................................ 2
  Theoretical Foundation .................................................................................................................. 3
  Research Questions ......................................................................................................................... 3
  Limitations & Researcher Positionality .......................................................................................... 5
  Assumptions/Definitions ............................................................................................................... 6
  Organization of the Document ....................................................................................................... 7
CHAPTER 2: REVIEW OF LITERATURE..................................................................................... 8
  Introduction ...................................................................................................................................... 8
  History of Grades & Standards ...................................................................................................... 9
  History of Grades & Standards for the State of Tennessee ............................................................ 14
  How Are Students And Teachers Held Accountable? ................................................................. 16
  Visible Learning ............................................................................................................................ 22
  Feedback ......................................................................................................................................... 25
  Standards-Based Feedback ........................................................................................................... 28
  Standards-Based Feedback: Tracking Progress .......................................................................... 31
Priority Standards ........................................................................................................... 32
State Level Examples of Standards-Based Feedback ..................................................... 33
School Level Examples of Standards-Based Feedback .................................................. 40

CHAPTER 3: METHODOLOGY ......................................................................................... 47
Population and Sample ................................................................................................. 48
Description of Instruments ........................................................................................... 49
Research Procedures & Time Period of the Study ......................................................... 50
Analysis of Data ............................................................................................................ 53

CHAPTER 4: RESULTS ................................................................................................. 56
Introduction .................................................................................................................. 56
Research Question 1 ..................................................................................................... 57
  Results ....................................................................................................................... 58
Research Question 2 ..................................................................................................... 59
  Results ....................................................................................................................... 59
Research Question 3 ..................................................................................................... 60
Research Question 3a ................................................................................................... 61
  Results ....................................................................................................................... 61
Research Question 3b ................................................................................................... 62
  Results ....................................................................................................................... 62
Research Question 3c ................................................................................................... 63
  Results ....................................................................................................................... 64
Additional Results ......................................................................................................... 64

CHAPTER 5: CONCLUSIONS, IMPLICATIONS, & RECOMMENDATIONS ................. 71
Summary ......................................................................................................................... 71

Conclusions .................................................................................................................. 73

Research Question 1 ...................................................................................................... 73
Research Question 2 ...................................................................................................... 73
Research Question 3 ...................................................................................................... 73
Additional Findings ..................................................................................................... 74

Implications .................................................................................................................. 75

Recommendations ........................................................................................................ 78

References ..................................................................................................................... 79

Appendix A ..................................................................................................................... 90

Front of NS.A.3 Tracker ................................................................................................. 91
Back of NS.A.3 Tracker ................................................................................................. 92

Appendix B ..................................................................................................................... 93

Front of RP.A.2 Mastery Tracker ................................................................................ 94
Back of RP.A.2 Mastery Tracker ................................................................................ 95

Appendix C ..................................................................................................................... 96

Front of EE.A.2 Mastery Tracker ................................................................................ 97
Back of EE.A.2 Mastery Tracker ................................................................................ 98

Appendix D ..................................................................................................................... 99

Permission Form from Director of Schools ................................................................. 100

Appendix E ..................................................................................................................... 101

Permission Form from Teachers ................................................................................ 102
List of Figures & Tables

Figure 2.1: 5th Grade Math Standards-based Report Card ................................................................. 30

Figure 2.2 Front of TN Individual Student Report .................................................................................. 34

Figure 2.3: Back of TN Individual Student Report .................................................................................. 35

Figure 2.4: Standards Analysis Report .................................................................................................. 36

Figure 2.5: Class Roster Report .......................................................................................................... 37

Figure 2.6: Instructional Focus for Standard 7.EE.A.2 ........................................................................ 39

Figure 2.7: MasteryConnect Tracker .................................................................................................. 42

Figure 3.1: Front of Mastery Tracker ................................................................................................... 52

Figure 3.2 Back of Mastery Tracker .................................................................................................... 53

Table 4.1: T-Test Summary for Tested vs. Control Post-test Gains ......................................................... 58

Table 4.2: T-Test Summary for Female vs. Male Post-test Gains .......................................................... 60

Table 4.3: T-Test Summary for Standard vs. Advanced Level Classes Post-test Gains ...................... 62

Table 4.4: T-Test Summary for Standard vs. Elite Level Classes Post-test Gains ................................. 63

Table 4.5: T-Test Summary for Standard vs. Upper (Advanced & Elite) Level Classes Post-test Gains ............................................................................................................................. 64

Table 4.6: T-Test Summary for Standard 7.RP.A.2 Standard vs. Advanced Post-test Gains ............... 66

Table 4.7: T-Test Summary for Standard 7.RP.A.2 Standard vs. Upper Level Post-test Gains ............. 67

Table 4.8: T-Test Summary for Standard 7.EE.A.2 Standard vs. Elite Level Post-test Gains ............... 68

Table 4.9: T-Test Summary for Standard 7.EE.A.2 Advanced vs. Elite Level Post-test Gains .......... 69

Table 5.1: Example of Mastery Tracker Reflection .................................................................................. 77
CHAPTER 1: INTRODUCTION & BACKGROUND

As Bill Gates said in a TED talk in May 2013, “We all need people who will give us feedback. That’s how we improve.” Bill Gates was referring specifically to teachers. Teachers need to know if the practices they have employed are effective. How will teachers know if they have been effective? How will their students know?

According to Brookhart (2008), feedback is defined as just in time information provided to the student when and where it can be the most beneficial. Proper feedback is necessary for the development of both teacher and student. Most feedback for students comes in the form of grades. Students turn in work, teachers grade the work, and students see their individual scores. It is important to determine, however, if students understand what the grade means. Are they able to monitor their own progress? Do they know if they are close to mastering a topic in a class?

Hattie (2009) conducted decades of research on effective influences of student learning. Feedback was 10th on the list of over 100 influences and maintained a top 10 position on that list each year that research was conducted. The intent of this research and writing was for teachers to “know thy impact” (Hattie, 2012, pg. 23). Teachers must know what is effective. They must have a way to determine if what they are doing is helpful in helping students gain a greater understanding of a specific topic.

Marzano scrutinized different types of feedback teachers could provide their students. He worked often with standards-based feedback. Instead of teachers providing students with a single numerical score on an entire assignment, Marzano suggested breaking the assessment up into different topics of studies, or standards, and providing students with feedback on each standard (Marzano & Heflebower, 2011). Instead of a student receiving, for example, an “85” on a test, each student would have a breakdown of that score based on each standard covered. A level of
proficiency should be provided instead of the standard percent correct score. Instead a student receiving an “85” on a test that covered three standards, the student would receive a proficient score on two topics and a near proficiency score on another topic. This would enable the student to understand the specific area that needs improvement.

Statement of the Problem

Current grading practices include a single numerical score associated with a myriad of learning targets. Neither the student nor the teacher can properly diagnose an area of weakness when such a small amount of feedback has been provided.

Teachers must not only change their grading practices, they must scrutinize their entire system. Students need to know how each assessment they take will help them prove their understanding over each specific learning target established in their learning. Students need standards-based feedback on their assessments.

Purpose/Significance of Study

The U.S. has experienced copious amounts of educational reform in the past three decades. There have been numerous state and national mandates implemented throughout the years with minimal improvements. Reform must start at the classroom level, not the state or national level. Teachers and students must work in tandem to identify necessary areas of improvement to increase student understanding.

The types of feedback provided to students by their teachers should be a primary area of focus. Teachers need more than what the state provides them at the end of the school year. Both teachers and students need checks for understanding throughout the school year. When students receive standards-based feedback from teachers, both student and teacher can better identify the path that must be taken to improve upon areas of weakness.
**Theoretical Foundation**

Hattie (2009) studied 800+ meta-analyses that had been previously conducted on effective teaching strategies for student learning. The result was a list of 100+ influences on student learning. He calculated the average effect size of each influence by calculating the pre-test/post-test difference, divided by the standard deviation of the two tests. He found that the average of all the influences studied was 0.40. This was established as his “hinge point.” At 0.40 effect size, a year’s growth occurred in a year’s time. If an influence had an effect size of larger than 0.40, then the student evidenced more than a year’s growth in a year’s time. Feedback has an effect size of 0.75, meaning almost two years growth can occur in a single year when feedback is used effectively by teachers. Hattie’s work will not only be used in this study’s calculations, but also in its implementation. Hattie emphasized two essential components of feedback: when it occurs and what type of feedback the student used. The student must be able to answer these questions: “Where am I going?” “How am I going there?” and “Where to next?” Four types of feedback students can receive were identified: task, process, self-regulation, and self-feedback.

**Research Questions**

This study attempted to answer one main quantitative question by scrutinizing pre-test and post-test scores of students who received standards-based feedback compared to those students who received single grade feedback. The quantitative analysis investigated if the post-test score gains of the experimental group from the pre-test were significantly different from those in the control group.
Research Question 1: Is there a significant difference in post-test gains from students who receive individual standards-based feedback and students who receive traditional evaluation feedback on formative assessments in the 7th grade math classroom?

Hypothesis:

H₀: There is a significant difference in the post-test gains from students who receive individual standards-based feedback and students who receive traditional evaluation feedback on formative assessments.

Null Hypothesis:

H₀: Post-test gains are equal in the experimental group and the control group.

Research Question 2: Is there a significant difference in post-test gains from female students who receive individual standards-based feedback and male students who receive individual standards-based feedback on formative assessments in the 7th grade math classroom?

Hypothesis:

H₀: There is a significant difference in the post-test gains from female students who receive individual standards-based feedback and male students who receive individual standards-based feedback on formative assessments.

Null Hypothesis:

H₀: Post-test gains are equal in the female group and the male group.

Research Question 3: Is there a significant difference in post-test gains from students in standard level classes who receive individual standards-based feedback and students in honors level classes who receive individual standards-based feedback on formative assessments in the 7th grade math classroom?
Hypothesis:

$H_a$: There is a significant difference in the post-test gains from students in standard level classes who receive individual standards-based feedback and students in honors level classes who receive individual standards-based feedback on formative assessments.

Null Hypothesis:

$H_0$: Post-test gains are equal in the standard group and the honors group.

**Limitations & Researcher Positionality**

This study was implemented at one school in a semi-rural Appalachian school district. Therefore, a smaller sample size was used for this study that may not be representative of a more diverse population. There were six classes involved: 3 in the control and 3 in the sample. The method of sampling will be discussed further in Chapter Three.

There was also a limitation of the fidelity of implementation of this new grading system. The research school had no prior experience with standards-based grading. Therefore, the students and the teachers were unfamiliar with this type of feedback and the fidelity of implementation had to be identified as a limitation of the study. However, the researcher created a student mastery tracker for each of the priority standards covered during the five-week unit used for the experiment. This tracker was used to ensure both students and teachers understood the standards-based feedback provided within the class.

For the experiment group, teachers created their own assessments and provided standards-based grades for all assessments during the unit. Most of these assessments were administered on the MasteryConnect platform, which provides a breakdown by standard, as well as a level of mastery. However, there were some assessments given on paper as well as on other platforms. For these other assessments, standards-based grades were provided by the teacher.
Fidelity of implementation of using standards-based grades was ensured by use of the mastery tracker. After each assessment, students would record their level of mastery for each standard assessed. The researcher is also the instructional math coach for the school tested and visited the research school on a weekly basis.

**Assumptions/Definitions**

The researcher assumed that all teachers involved would provide standard-based feedback on all assessments implemented over a five-week period covering the topic of ratios and proportional relationships. The researcher also assumed that students would be allowed time in class after each assessment to properly plot their latest proficiency levels in their mastery tracker, as well as answer reflective questions about the tracking process.

The following definitions are noted because they are used frequently in this study:

- **Benchmark/Interim Assessments**: medium scale assessments falling between formative and summative assessment that serve to (1) evaluate students’ knowledge and skills relative to a specific set of academic goals, typically within a limited time frame, and (2) are designed to inform decisions at both the classroom and beyond the classroom level, such as the school or district level (Perie, Marion, Gong, & Wurtzel, 2007).

- **Feedback**: “information allowing a learner to reduce the gap between what is evident currently and what could or should be the case” (Hattie & Yates, 2014).

- **Priority Standards**: “a carefully selected subset of the total list of the grade-specific and course-specific standards within each content area that students must know and be able to do by the end of each school year to be prepared for the standards at the next grade level or course. Priority standards represent the assured student competencies that each teacher
needs to help every student learn, and demonstrate proficiency in, by the end of the current grade or course” (Ainsworth, 2013).

- Standards-based Grading: an assessment system in which students are graded solely upon their proficiency about well-defined standards (Marzano, 2010a).

Organization of the Document

The first chapter presents the background of the study, purpose of the study and statement of the research question. Chapter 1 also identifies the theoretical framework used in this phenomenological mixed methods study, as well as the definition of terms, assumptions, and limitations of the research. The second chapter includes a review of literature investigating the history of feedback and grades, the legislation passed in the U.S. for accountability and testing, John Hattie’s work on Visible Learning and feedback, Marzano’s contributions to standards-based feedback, specific case studies that have identified positive results, as well as areas for improvement for proper implementation of the new grading system. The third chapter details the methodology implemented for the collection of data for the quantitative question posed. The fourth chapter discusses the statistical outcomes from the data collection. Finally, the fifth chapter summarizes of those results and offers recommendations for further study.
CHAPTER 2: REVIEW OF LITERATURE

Introduction

American education has always been a work in progress. There has always been a desire to learn more and determine how to do better. Educators began developing and implementing grades based on agreed upon standards. In 1965, Lyndon B. Johnson signed the *Elementary and Secondary Education Act* (ESEA), with the intent of improving the educational experience of lower income students. It was not enough for the country to just have an education system in place, there needed to be a proper environment for all students to learn, not just the elite.

In 1983, Ronald Reagan addressed education again in *A Nation at Risk*. Providing an education for all students was not adequate. Instead, a rigorous education was needed to be provided for all students. Thus, the improvement process continued into the ‘90s and into the turn of the 21st Century. Americans always wanted to improve their education system and to ensure that there was a way to hold their teachers and students accountable.

However, America wanted to have the best education system, but as time passed, it continued to suffer in comparison to other countries.

The United States used two primary assessments to compare student achievement. They are: the National Assessment of Educational Progress (NAEP) and the Program for International Student Assessment (PISA). NAEP was given first nationally in the U.S. starting in 1969. By 1990, NAEP was used by all states on 4th, 8th and 12th graders. This common assessment allows for comparison of achievement between states within the country (National Center for Education Statistics, 2009). It is used specifically to show progress within the U.S. According to the 2017 results, students in the U.S. have made very little progress in math and reading since the last assessment in 2015 (Camera, 2018).
Many countries also use PISA as a benchmark to compare students’ academic performance. In 2015, the U.S. ranked 35th in Math, 24th in Reading, and 25th in Science out of the 72 countries who took the PISA exam (Jackson & Kiersz, 2016). NAEP showed that each state has demonstrated little to no gains in the last two testing windows and PISA indicated that the U.S. does not compare favorably with other countries in education. How can this gap be closed?

Hattie (2009) evaluated the effect size of hundreds of different influences on student achievement, through studies on over 800 meta-analyses. Once he calculated all these influences by effect size, he ordered them by highest impact. Feedback was determined to be one of the top 10 influencers on student learning and it was determined that feedback was one of the top 10 influences on student learning.

Marzano (2010a) emphasized the importance of standards-based feedback and advocated for schools to shift to more standards-based feedback instead of the traditional grading model. While there are various perspectives on standards-based feedback, all parties involved call for the absolute necessity to change how the current U.S. education system assesses student learning and understanding (Marzano, 2010a; Guskey, 2011; Hierck & Larson, 2018).

**History of Grades & Standards**

Before educational reform could effectively occur, there had to be a baseline of data. According to Stones (2012), the earliest use of educational standards can be traced back to Ralph W. Tyler in 1934. Teachers needed to know what standards they expected their students to learn by the end of the year, but how were they going to prove their students had mastered these standards? Teachers required a way to record how well their students performed, and thus letter
grades were introduced to the educational field. By 1971, letter grades were used in over 80% of schools (National Education Association, 1974).

While the use of letter grades provided a baseline for educators, there was minimal push for improvement prior to the publication of A Nation At Risk in 1983. This report had five main overarching themes of recommendation: content, standards and expectations, time, teaching, and leadership and fiscal support. Some recommendations had minimal staying power in the education field. For example, the report called for seven-hour school days and an increase in the number of school days from 180 to 200-220. It also called for teachers to be on an 11-month contract instead of the typical 10-month schedule. However, there were suggestions that would be later implemented in numerous states in the following decades.

In the first section of recommendations in A Nation At Risk, many of the points reference the number of credits high school students should take to graduate. It was suggested that necessary credits to graduate should include: four years of English, three years of math, three years of science, three years of social studies, and one-half year of computer science. These numbers correlate with graduation requirements for high school seniors in Tennessee. For the 2017-2018 school year, these students are required to have completed four credits of English, four credits of math, four credits of science, and three credits of social studies (Tennessee Department of Education, 2018a).

A Nation At Risk cited the absolute need for states to adopt more rigorous standards to make students more college and career ready. The report also called for schools to investigate how they were grading their students. The report called for grades to be a clear indication of academic achievement. It was insinuated that grades had become inflated and were not reliable.
It was also suggested that standardized tests be administered as part of a “nationwide (but not Federal) system of State and localized standardized tests” (United States, 1983).

Whitman (2015) cited Reagan’s report as the roots for Common Core. Hamilton, Stecher, Yuan, and Rand (2008) stated that *A Nation At Risk* was the cause for the standards-based grading shift.

Education was further scrutinized when President George H.W. Bush called an educational summit in 1990 to create the National Goals for Education, a concise list of six goals that the U.S. education system should reach by 2000. They were:

1. All children will start school ready to learn.
2. High school graduation rate will increase from 70% to 90%.
3. Students will leave grades 4, 8, and 12 having “demonstrated competency over challenging subject matter including English, mathematics, science, history and geography” (pg. 4).
4. U.S. students will be the first in the world in math and science achievement.
5. Every adult American will be literate.
6. Every school in America will be drug and violence free.

While many of the goals were not achieved by 2000, they still echo to this day. Most accountability models for high schools use some sort of graduation rate requirement and competency goals in a variety of content areas. Competencies were measured using NAEP scores. These scores remain in use presently. As of 2017, only 40% of all 4th graders nationally scored at or above proficient in math and 37% in reading. Only 36% of 8th graders scored at or above proficient in math and only 37% of 8th graders scored at or above proficient in reading (Miller, 2018).
President George H.W. Bush alluded to a need for some type of improvement on the current academic standards, but used specific language stating this address was not a requirement of a national curriculum. President George H.W. Bush provided the National Council of Teachers of Mathematics’ recent creation of new math standards in 1989 as an example of “defining what all students must know and be able to do in order to be mathematically competent” (Executive Office of the President, W.D., 1990). This mandate for states to clearly define what the success criteria should be for each student did not go abate as time passed. There was a common theme in later legislation for states to adopt more rigorous standards. What should these standards look like? At the federal level, the only real guidance that was provided was evidence that the current standards that were being used were not working. The first real example of these newer, more rigorous standards that addresses like *A Nation At Risk* and the *National Goals for Education* desired came in the year 2008 (Bidwell, 2014).

Addresses like *A Nation At Risk* and *National Goals for Education* were somewhat paradoxical in nature. They both called for newer, better standards, but both denied the need for a national curriculum. Many states were at a loss in the following decades as to what these new standards should look like if there was not going to be a national example. According to Bidwell (2014), one state did not want to adopt extremely rigorous standards and have its students falter on state assessments, while students in another state with less difficult standards had higher test scores from its students. However, Arizona Governor Janet Napolitano wrote the initiative to begin the Common Core State Standards task force in December 2008. Each content area had different experts that created the sets of standards, but math had a three-man team. This team consisted of Jason Zimba, William McCallum, and Phil Daro. They released their final drafts of the Mathematics Common Core State Standards to the public in June 2010 (Garland, 2014).
Then, President Obama created the Race To The Top (RTTT) initiative in November 2009 (U.S. Department of Education, 2009) calling for reform in four core areas:

1. Adopting standards and assessments to prepare students for post-secondary opportunities
2. Build a data system that measures growth and achievement
3. Finding a way to retain effective teachers and principals
4. Turn around low achieving schools

The first priority for reform seemed remarkably familiar to education reform in years past. What was the difference? This initiative involved a lot of money. Prior to RTTT, President Obama had signed the American Recovery and Reinvestment Act (ARRA) which provided $4.35 billion for the RTTT fund (U.S. Department of Education, 2009).

States had to earn a portion of the $4.35 billion. RTTT was not a law that mandated compliance. Rather, it was a competition in which the winners would receive large portions of the $4.35 billion. To be eligible to receive funds, states had to submit clearly defined plans on how they were to address the four core education reform areas listed. Three of the factors that determined the winning states included adopting new academic standards, creating a robust data system that measured growth and achievement, and an accountability model to measure school level conditions, and provide data on where improvements may need to occur (U.S. Department of Education, 2009).

Common Core State Standards were not mentioned. However, when states prepared to adopt their new standards, they had to meet a certain requirement. The standards must be adopted by a “consortium of states” where each state would be allowed to add their own unique standards, but those unique standards could not account for more than 15% percent of the total amount of standards adopted. How was a single state supposed to find their consortium? And
how were they supposed to determine a set of common standards across a myriad of states? The simple, and most used, solution was to adopt the Common Core State Standards and each state could make some slight changes (Jochim & McGuinn, 2016).

This was the case for Tennessee, and 45 other states in 2010 (Corona, 2015). So, what went wrong? All these states voluntarily implemented a new set of somewhat national standards. The backlash occurred in 2013. Test scores plummeted in New York and Indiana. Then, scathing media comments followed in 2014 (Garland, 2014). Race to the Top prompted 46 states to use Common Core Standards, with the agreement to implement them in the 2014-2015 school year. Only Alaska, Nebraska, Texas, and Virginia never adopted CCSS. There were two main testing vendors who offered assessments for those states who chose to use CCSS: the Partnership for Assessment of College Readiness (PARCC) and the Smarter Balanced Consortia. Initially, there were 46 states who committed to CCSS, but that number quickly dropped. Many states did not continue to utilize CCSS and its federal assessment companies. By June 2014, 19 states had either withdrawn from the tests or paused implementation (Corona, 2015). The cost of adopting Common Core was too great.

**History of Grades & Standards for the State of Tennessee**

In the years that followed, states like Tennessee went through a myriad of standards changes and assessment changes. Following RTTT, Tennessee was initially a part of the state consortium that used PARCC, a Common Core aligned state assessment (Kebede, 2017). However, the resistance following the 2013-2014 assessment season, forced Tennessee, along with many other states to change course.

In 2014, Tennessee chose to continue to implement the Common Core State Standards, despite the backlash from other states. However, Tennessee opted to continue using its current
assessment system, the Tennessee Comprehensive Assessment Program (TCAP). This assessment system was still aligned with Tennessee’s previous standards. Thus, the state had to find another system to assess that would be aligned with its new standards. So, it was decided in 2014 that Tennessee would hold a competitive bidding process for the vendor for its new, CCSS aligned assessment. While PARCC could have been that vendor, Tennessee would ultimately decide against using them, as did Florida, Kentucky and several other states (Zubrzyck, 2014).

As of 2017, 21 states used PARCC for their state assessments. Tennessee was not one of those states (Gewertz, 2017). Tennessee opted out of this plan in 2014, choosing instead to create its own assessment, TNReady (Kebede, 2017).

Subsequently, Tennessee created its own assessment to be administered by another testing vendor. Tennessee wanted to avoid the issues that numerous other states had experienced with Common Core aligned assessments, such as PARCC and Smarter Balanced. For the 2015-2016 school year, Measurement, Inc., was chosen as the testing vendor. The initial plan was for students across the state to take the assessment online. Up to this point in Tennessee’s testing history, students had never taken their state assessment online (Tatter, 2016). However, the online platform, MIST, did not work. Measurement, Inc. did not have a back-up plan ready to implement. Months after it was determined that students would be testing on paper, no tests had been sent to schools, who would soon begin their testing windows. As a result, elementary and middle schools across the state did not take the new assessment (Tatter, 2017).

With the next year came a new assessment framework and a new testing vendor. The state determined students were spending too much time testing. Thus, the schedule changed from two separate testing windows (one in January/February and the next in April/May) to one testing window with shorter testing times. Subsequently, instead of Measurement, Inc., Tennessee used
the testing vendor Nextera (Boehnke, 2017). The 2016-2017 school year ran smoothly. Only the high schools in the state that volunteered would take the assessment online.

The next school year, 2017-18, however, was problematic in terms of testing. The first day of testing a fiber cable was cut by a dump truck and testing went down across the state. The fiber cable cut forced most internet activity to be redirected and thus caused connectivity issues for students in schools across the state (Aldrich, 2018). The next day, Nextera experienced what they referred to as a “hack” from an outside source. Education Commissioner Candice McQueen stated that no student data was compromised. Nonetheless, this resulted in another testing delay. The remainder of the testing window schools saw several more problems on both the online platform and on paper (Gonzales, 2018).

Teachers and students need standards and assessments. Presidents Reagan and Bush were correct in their grave warnings and lofty goals. The United States has created federal and state mandates to ensure all students are successful, not just a selected few. There has been a constant pressure in U.S. education to improve from one generation to the next. The birthplace of education reform becoming law from the federal level started in 1965 with President Lyndon B. Johnson’s *Elementary and Secondary Education Act*. This law was the first of many to ensure educators were held accountable for their individual student’s growth and achievement. Those students were then held accountable to those goals by their teachers (Klein, 2015).

**How Are Students And Teachers Held Accountable?**

There was minimal federal involvement in education prior to 1965. The *Elementary and Secondary Education Act* of 1965 was signed into law by President Lyndon B. Johnson. This was the first major reform attempt made at the federal level (Elementary and Secondary Education Act of 1965). It was certainly not the last.
ESEA was originally only 32 pages in length with six “Titles” or sections. It was expedited through Congress in 100 days and its primary focus was to provide federal funding to aid students of a lower socio-economic background (ESEA, 1965). The intention was good, but the parameters were not clear. ESEA intended to provide schools with money to close the achievement gap for those lower income students, but this was not necessarily the case. The money was not well spent because the government did not really provide any checkpoints for schools to ensure they had used the money in the proper way (Klein, 2015).

The next 50 years included over half a dozen reauthorizations to this act, and it grew from 32 pages to over 400 (ESEA, 1965). According to Wardlow (2016), there were three notable reauthorizations regarding improving academic standards: Improving America’s Schools Act of 1994, No Child Left Behind (2001), and Every Student Succeeds Act (2015).

Improving America’s Schools Act (IASA) was passed in 1994 and it called for “high standards for all students.” IASA called for all programs, not just those for disadvantaged students, to be focused “around a core of challenging state standards.” It also mandated that all states have:

- content and performance standards
- assessments aligned with those standards in one grade in the following grade spans: 3-5, 6-9, and 10-12
- An accountability system to identify when students did not meet the expectations on those assessments. (Jorgensen & Hoffman, 2003)

The next major reauthorization came seven years later with the No Child Left Behind Act (NCLB). Again, NCLB called for each state to adopt “challenging content standards” and “achievement standards” that would be applied to all children in the state (No Child Left Behind,
2001). This wording sounded very similar to IASA. The two reauthorizations diverged in their accountability models. States would have a baseline year for the 2001-2002 school year. Then every year after, schools would have to meet their “Adequate Yearly Progress,” or AYP, to reach the goal of all students meeting or exceeding the state’s proficient level of academic achievement on the state’s assessment by the year 2014. NCLB emphasized standards and assessment and had very specific and rigorous expectations for all students in the country.

Many states were fearful of the 2014 deadline. Numerous schools were failing to meet their AYP each year, and several states began requesting waivers (Wardlow, 2016). In 2015, President Obama passed ESSA. It removed AYP and afforded individual states more discretion to make their own decisions about improvement plans. It required that every state implement an accountability system that was aligned with the academic standards that included sanctions and rewards for making the state’s definition of adequate yearly progress (Every Student Succeeds Act, 2015).

ESSA was not the only initiative to occur during this time period in education reform history. Winners for Race To The Top had been announced in three phases beginning in March 2010 progressing through April 2012. By the time President Obama announced ESSA, the state of Tennessee, one of the phase one winners of RTTT, had begun the implementation of its reform plan (Finch, 2017).

Tennessee had created and adopted a new evaluation model for schools and teachers in the 2011-2012 school year called the Tennessee Educator Acceleration Model, or TEAM (Tennessee Department of Education, 2016). This model called for use of student achievement data, along with teacher evaluation data. The Tennessee Value Added System (TVAAS), based upon the William Sanders’ model created at the University of Tennessee in the 1980s, served as
a robust data system for student achievement in Tennessee since 1993. This value-added model
adopted by the state of Tennessee reported on the effectiveness of school systems. It used
Tennessee Comprehensive Assessment Program (TCAP) data for grades 3-8, End of Course
(EOC) for high school grades, ACT and annual writing assessments from grades 4, 8, and 11 to
calculate the overall effectiveness of teachers and schools (Sanders & Horn, 1998). This robust
system was not developed to only house achievement data. Sanders’ model to evaluate
effectiveness was based upon two parts: achievement and growth. The goal of TVAAS was to
determine the effectiveness of teachers and schools regarding student growth. The Sanders’
model provided schools with the unique opportunity to project student scores at the end of the
current year. This projected score was calculated based upon the statistical analysis of each
student’s cumulative testing data from previous school years. If students met their expected
projection, that meant they grew a full academic year. If students were multiple grade levels
behind their current grade, and they only grew a year, then the teacher’s achievement data would
be troublesome, but students’ growth data would indicate some achievement.

This idea of growth and achievement together was unprecedented. In the years to follow,
other states would begin using this value-added model, including Ohio, Pennsylvania, North
Carolina and at least 15 other states (Hall, 2014). Tennessee used TVAAS for over a decade
before adding TEAM. However, TEAM combined TVAAS with the new evaluation model to
provide teachers with a more comprehensive view of their overall effectiveness.

The first year of TEAM implementation demonstrated positive results. Math in grades 3-
8 saw an increase in proficiency and advanced percentage of students, growing from 34.6% in
2010 to 47.3% in 2012. This resulted in over 55,000 more students who performed at or above
grade level in 2012 compared to 2010. Science evidenced similar improvements, growing from
51.9% to 60.5% (more than 38,000 students) in the proficient or advanced categories (Tennessee Department of Education, 2012).

Teachers and administrators had to be properly trained on the new evaluation rubric. Teachers would be evaluated under three different domains: planning, environment, and instruction. These evaluation scores would be half of the teacher’s overall level of effectiveness score. The other 50% of their level of effectiveness score consisted of student achievement information: 35% from TVAAS and 15% based upon additional student achievement chosen by the individual teacher. This “additional student achievement” could have been the school’s state assessment score, school-wide TVAAS, graduation rate, ACT scores, etc. (Tennessee Department of Education, 2012). One of the initial concerns with this evaluation model was that teachers received inflated scores on their evaluation. In the first year, more than 75% of teachers received top scores of a 4 or a 5 on their evaluations and less than 2.5% received a 1 or a 2. This did not align with the TVAAS individual teacher effect breakdown, where almost a quarter of all teachers in Tennessee should have scored in the 1 or 2 range and only around 50% of teachers should have received top marks of a 4 or a 5 (Tennessee Department of Education, 2012). School administrators received training on TEAM evaluations and would continue to undergo evaluation training annually to receive their recertification (Tennessee Department of Education, 2016).

There has been a greater emphasis on higher academic standards and a greater accountability for schools and teachers in education in the United States (Wardlow, 2016). Many laws, starting with ESEA, have been passed to better inform schools on the effectiveness of their teaching the previous year. Likewise, states know more about the progress schools have made in closing the achievement gap for all students. For Tennessee, these tools include the TEAM evaluation model, as well as their growth and achievement measures through TVAAS and
TNReady, respectively (Tennessee Department of Education, 2012). The benefits for students, however, can be questioned. Much of the legislation was designed to be in the best interest of the student, and it has helped educators better discern teacher effectiveness over the course of the year. However, all this data and information is issued after the student leaves the teacher’s classroom. Also, students are infrequently involved in the process. Students receive their scores at the end of the school year, and their breakdown reports the following year (Tennessee Department of Education, 2017c). However, students are rarely asked to apply and utilize this information.

Educational change must be initiated at the student level. Change cannot be a “top down” process, because, unfortunately, in “top down” change, students are the last ones to know what change occurred. Change must occur at the federal, state and district levels. Subsequently, school administrators share this change with teachers, who then share with students. Students must be partners in their own education. They need to hear and understand what it is they know or do not know. They need to internalize this information and process it throughout a school year. Students do not need to be told at the end of the year they did not earn a high enough score to be proficient for their grade, and they are now behind their peers. Students need feedback regarding academic progress from teachers and should work in tandem with the teacher to determine how to better progress.

According to Frey, Fisher, and Hattie (2018), schools must work to create “assessment-capable learners,” not students who are “adult-dependent learners” or “compliant learners.” They stipulated:

Assessment is conventionally thought of as something we do to students. We measure their progress and report it to them, their families, and the public. Enlightened teachers
take it a step further, using assessment results within a systematic formative process to make decisions about future instruction. Groups of these educators may even band together in professional learning communities to make data-driven decisions for their grade level or department. But there is one person whose input is left out of these processes—the student. The person at the center of the discussion is relegated to a passive role. (pg. 47)

**Visible Learning**

Before teachers can help students to become assessment-capable learners, they must first understand visible learning. John Hattie, Professor and Director of the Melbourne Education Research Institute at the University of Melbourne, Australia, as well as an honorary professor at the University of Auckland, New Zealand, is the author of *Visible Learning* (Hattie, 2012). While his first book on this topic was released in 2009, the research on the topic began decades prior. Hattie (2012) researched thousands of educational research papers to scrutinize various instructional strategies to better learn what impacted student learning. He calculated effect sizes for every influence and used pre- and post-test scores for every study analyzed. He took the difference between the two averages and divided it by the average standard deviation of the two groups.

A common thread was determined. The average effect size was 0.40. Using this average, Hattie made the connection that a year’s worth of learning occurred in that years' time if an influence had an effect size of 0.40. Anything above a 0.40 effect size, the student would grow more than a year. This 0.40 was referred to as the “hinge point” (Hattie, 2012, pg. 15). This effect size calculator was used to score and categorize over 100 different influences on student learning.
Many educators tout Hattie’s research as the Holy Grail of education (Evans, 2012). However, there has been a small but vocal group opposed to his vast generalizations found from his meta-analyses and calculated effect sizes. Snook, O’Neill, Clark, O’Neil and Openshaw (2009) proposed a myriad of arguments with Hattie’s research and effect size findings. They noted his bias toward published studies. Published studies tend to have more positive results, thereby inflating Hattie’s potential influence. They also argued with the vast differences in some of the studies that Hattie used for a single influence. For example, they cited Hattie’s effect size for homework was listed as 0.29, but upon further inspection. This effect size was the average for all students. Homework’s effect size for elementary was 0.15, but 0.64 for high school. They also discussed the overwhelming complexities of some of Hattie’s singular influences, such as age group and ability level. They noted that the effect size of 0.29 does not properly encompass all the different possibilities of such an ambiguous topic. Some of Hattie's definitions where also deemed as vague. For example, class size has an effect size of 0.20, but what was the definition for “small” versus “large?” What one study defined as “large” was not necessarily the same definition used in another study.

Hattie (2012) determined that approximately 10% of effect sizes were negative. The effects causing the most positive change in student learning occurred after that 0.40 hinge point. Other strategies and influences below the 0.40 score were also impactful. For example, in the 2012 release of influences, it was stated that drama and arts programs only had a 0.35 effect size for student achievement. It is essential for teachers to continually re-evaluate daily classroom practices. Teachers should be cognizant of their impact on students.

This list of influences could easily be a checklist of strategies to use (or not use) in the classroom. Shanahan (2017) recommended it to be used as a reference but encouraged teachers
to scrutinize Hattie’s original research when he compiled his 800+ meta-analyses. It was not
Hattie’s intention for educators to use this as a list. In fact, he wanted to empower both the
student and the teacher.

After he released his 2009 *Visible Learning* and his 2012 *Visible Learning for Teachers*,
Hattie collaborated with Nancy Frey and Douglas Fisher to write *Developing Assessment-
Capable Learners*. This book focused on students. According to Hattie, Fisher & Frey (2018),
five factors impact assessment-capable visible learners:

1. I know where I’m going
2. I have the tools for the journey
3. I monitor my own progress
4. I recognize when I’m ready for what’s next
5. I know what to do next

When students are asked to do these things, they struggle. Students find it difficult to verbalize
what they know and do not know. They have trouble identifying the level of difficulty of
problems. In a 2009 study conducted on Physics and Biochemistry students, at the University of
Sydney in Australia, Lingard, Minasian-Batmania, Vell, Cathers, and Gonzales (2009) asked
students to rate each of their quiz questions with a designation of difficulty. Results indicated
that students who consistently misaligned the question difficulty had a higher likelihood of
getting the question incorrect. Devoting the time each day for students to analyze their own
progress on meeting learning goals and asking students to verbalize how they are going to use
what they know to progress is vital for the assessment-capable visible learner.

According to Hattie, Fisher and Frey (2018), students need teachers who have clearly
defined the success criteria for their learning. They need teachers to provide them time in the
learning to examine their progress toward those success criteria. Teachers need to cultivate the skills of motivation, goal setting, self-regulation, and feedback. All four of these skills work together to create an assessment-capable learner.

**Feedback**

Brookhart (2008) defined feedback as just in time information provided to the student when and where it can be the most beneficial. Hattie and Yates (2014) defined feedback as the information that allows the learner to close the gap between current understanding of a topic and the goal of mastery for that topic. Hattie (2012) identified feedback as one of the top influences on student achievement; feedback had an effect size of 0.75, meaning its use could result in a student growing nearly two years’ worth in a year’s time. Marzano (2010a) identified what types of feedback were most effective. His work on different types of feedback led to Marzano’s work involving the standards-based grading reform movement.

According to Hattie (2012), students should ask themselves the following questions during a lesson: “Where am I going? How am I going? Where am I going to next?” (pg. 130). The characteristics of an assessment-capable learner and questions that visible learners should ask themselves are very similar. For a student to be assessment-capable, he/she must first be active participants in his/her own feedback. Four different types of feedback were described: task, process, self-regulation, and self.

These four types of feedback built upon one another. A student may start with task feedback. This type of feedback simply told the student if he/she got the question right or wrong and how successfully the task was completed. Process feedback asked the student what strategies he/she used to determine an answer, or what strategies will the student need to solve the next problem. A study conducted by Harks, Rakoczy, Hattie, Besser & Klieme (2014) scrutinized 9th
grade Algebra I students who received process-oriented feedback versus grade-oriented feedback. The students who received process feedback found it more useful and thus had a more positive effect on changes in student achievement.

Self-regulation allows the teacher to provide support but prompts the student to question the next step in learning. For example, a student works out a problem and checks his or her answer. The student finds the answer to be incorrect. The teacher provides the self-regulation feedback of asking the student if he or she can deduce where the error was made. This type of feedback places more emphasis on the student enhancing his/her skills of self-evaluation. Self-feedback occurs when a student personally evaluated his/her role in learning. Self-feedback asks the learner to step outside the content and picture the overall learning process. Hattie (2012) noted the importance of the involvement of the student. These latter two levels were not focused on the teacher; the student asked the questions and selected a path.

According to Hattie, Fisher, and Frey (2016), the goal of providing feedback is to ensure that students hear the teacher. To ensure maximum effect, teachers must first establish clear success criteria. Then, teachers must provide different kinds of feedback and maximize efforts to promote students’ listening skills. Clear success criteria are essential if teachers are to offer any type of effective feedback. Hattie and Marzano both agree that for feedback to occur, clear success criteria must be in place (Hattie, 2012; Marzano & Haystead, 2008).

Killian (2015) noted that both Marzano and Hattie highlighted the importance of feedback being a “two-way street.” Both the teacher and the student must be involved. According to Killian, Marzano suggested that teachers must provide feedback when there is still time for the student to learn. For example, students should not only receive feedback the
following school year regarding a state exam they took the previous year, but they should also receive effective feedback from their teachers when they can utilize the feedback for growth.

Feedback is vital, but not all forms of feedback are equally effective (Marzano & Association for Supervision and Curriculum, 2006). Four recommendations for effective feedback in the classroom were provided by Dean & Marzano (2012):

1. Identify what the student did correctly and let the student know what to do next.
2. Provide the feedback in time for the student’s needs.
3. Feedback should be criterion based.
4. Students should be a part of the feedback process (pg. 14)

Praise was not considered to be an effective form of feedback. Teachers must provide students with information on their performance on a specific task. The feedback must be “criterion based,” meaning the feedback must be linked to a specific topic or standard. Another type of based feedback can be norm-based feedback (Marzano, 2010a). Guskey (2011) argued against this type of feedback. For instance, if a student received a “C” on an assessment, he/she would interpret his results as average among his peers. If the assessment was criterion-based, it would mean he/she had reached approaching proficiency. The perception shift causes students to focus more on their individual understanding instead of their ranking among their peers. Nitko (1994) defined criterion-based scoring as using a set up performance standards to define the level of mastery, rather than norm-referenced scoring, when a student's score is based upon how he/she did in comparison with peers.

Marzano and Heflebower (2011) noted how traditional grading systems have students earn points for a variety of different activities, assignments and behaviors. Students who were graded on this system had no way of knowing what an overall percentage score of a 62.9 meant.
It could mean the student lacked fundamental understanding on a foundational topic, or it could mean the student lost points for tardiness or missing assignments. The following recommendations were detailed regarding how to provide grades that best indicate student knowledge:

1. Get rid of the single overall grade.
2. If teachers cannot get rid of the single, overall grade, provide a breakdown of scores by topic in addition to the single grade.
3. Provide other assessment options for students.
4. Provide re-testing opportunities.

Standards-Based Feedback

Marzano & the Association for Supervision and Curriculum (2006) highlighted different studies on the different type of assessment feedback. Bangert-Drowns, Kulik, Kulik & Morgan (1991) found the following effect sizes for these different types of feedback: right/wrong (-0.8), provide correct answer (0.22), explanation provided (0.53), repetition until correct (0.53). Fuchs and Fuchs (1986) demonstrated that when students utilized the assessment feedback, the effect scores increased drastically. When assessment results were displayed graphically, the effect size was 0.7.

According to Marzano & Association for Supervision and Curriculum (2006), standards-based grading is an entirely new assessment system compared to the traditional grading model. Teachers should first determine what “measurement topics” will be addressed each grading period. These “measurement topics” could be referred to as criterion or standards. Irrespective of the term, these topics the teacher chooses will be referenced whenever feedback is provided.
The teacher would design assessments to assess each of these topics. Assessments could encompass a variety of topics, but they must be designed in such a way that students know exactly how they did on each topic, not just their overall grade. Students and teachers could then use each assessment to track a student’s approaching mastery of each topic (Marzano & Association for Supervision and Curriculum, 2006, pg. 59-62). By using this new model, these types of assessments encouraged learning because they enabled students to track their own progress, encouraged self-reflection, and focused on the learning at the end of the grading period (Marzano & Association for Supervision and Curriculum, 2006, pg. 89-94).

The final component of standards-based grading is identifying the level of proficiency a student receives per topic covered. This is a critical component of Marzano’s standards-based grading approach. Students receive a numerical level of proficiency, 1-4, with 4 meaning the student has completely mastered the content. As seen in Figure 2.1, an example is noted. This report card shows the level of proficiency a student shows in 5th grade (Marzano, 2008, pg. 50-51). Similar to the traditional report card, Figure 2.1 shows a breakdown of grades. However, the report card differs in that the breakdown is not sorted by grade type; for instance, quiz, test, homework, etc., it is sorted by topic. The report card shows the student progress on the different types of topics in 5th grade math. According to Marzano, the parent and student would be able to quickly identify that “Data Organization and Interpretation” is a weak point for the student in question.
Figure 2.1 shows an example of a standards-based report card. The letter grade conversion for this card is shown:

$3.00 - 4.00 = A$; $2.50 - 2.99 = B$; $2.00 - 2.49 = C$; $1.50 - 1.99 = D$; $1.49$ and below $= F$

(Marzano & Haystead, 2008, pg. 44)

A concern with using this model has been the difficulty to convert to traditional letter grades. An example of this concern is evidenced in the study on secondary students' perception of standards-based grading conducted by Peters, Kruse, Buckmiller, and Townsley (2017). These students received standards-based grades and believed that while the grades may have better reflected what they mastered, their grades declined. The concern was that universities would scrutinize their GPA, note their decline in grades, and students would lose out on scholarship opportunities.
Standards-referenced grading has been an “attractive option” to standards-based because of issues of retention. According to Marzano & Mid-Continent Regional Educational Lab (1998), in a standards-based system, a student must demonstrate a certain level of proficiency at one level before the student can move onto the next level. Thus, standards-referenced grading is a slightly smaller shift than transitioning from a traditional grading system to standards-based. In standards-referenced system, students would still receive a performance level based upon the pre-determined rubric from the school or teacher. The main difference is that a student may progress to other standards even if he/she has not demonstrated mastery on the initial standard.

**Standards-Based Feedback: Tracking Progress**

Marzano (2010a) also investigated the effects of students tracking his/her own progress towards mastery. In the 14 studies on students tracking their own progress, the average effect size was 0.92, which translated into a 32-percentile point gain. According to Marzano (2010b), this 32-percentile point gain is not necessarily a given. Three main requirements for success on students tracking their own progress toward mastery were noted.

1. Students tracked mastery on a single goal for all assessments. A single assessment may address multiple goals, but a student would receive a score for every goal assessed.
2. Teachers used rubrics instead of points. In the studies where students just tracked their scores based upon percent correct, there was not as high of effect size. Students improved more when they were scored on a rubric, meaning they received a level of proficiency on each assessment, instead of a raw percentage score.
3. Students received different types of assessments. They did not just receive paper and pencil tests.
According to Marzano (2010b), having students track their own progress using rubrics is beneficial. Standards based grading and students tracking their own progress work together. Teacher use of a consistent standards-based grading platform makes it easy for a student to track his/her progress toward mastery. Using this type of tracking, students have clear guidance on how to improve their own learning.

Before a school or district can implement an entirely new grading system, it must first work to determine what the topics each content area will cover. Numerous American school teachers have found it problematic to cover the wide variety of required standards each year. What do these educators do? They have two options, either cover everything quickly, or cover few things deeper (Ainsworth, 2013).

**Priority Standards**

Ainsworth (2013) defined priority standards as a carefully selected subset of the total list of the grade-specific and course-specific standards within each content area that students must know and be able to do by the end of each school year in order to be prepared for the standards at the next grade level or course. Priority standards represent the assured student competencies that each teacher needs to help every student learn, and demonstrate proficiency in, by the end of the current grade or course. (pg. xv)

According to Ainsworth (2013), these priority standards should be promises teachers make to their students, that no matter what happens during the year, every student will know how to do these standards. Ainsworth defined how schools should go about defining their priority standards using what he calls the “R.E.A.L.” criteria. The R.E.A.L. criteria is an objective selection criteria. The “R” stands for readiness: is this standard need for the next grade? The “E”
stands for endurance: will this standard last beyond one grade and could be used later in life?
The “A” stands for assessment: will this standard be a focus on the end of the year assessments?
Finally, the “L” stands for leverage: does this standard have any crossover with other content areas within that grade level?

Ainsworth & Viegut (2006) suggested exactly how these priority standards should be assessed in a standards-based assessment system. Schools should utilize more common formative assessments (CFAs). These CFAs should be a blend of priority standard items that are analyzed within the confines of a PLC prior to the administration to all students in the content area. When teachers use CFAs, they are providing evidence of their effect, as Hattie (2012) recommended in his Visible Learning framework. CFAs also provide regular, timely feedback to students and have a predictive value of how students are progressing with mastery of priority standards.

State Level Examples of Standards-Based Feedback

Numerous reports that Tennessee provides for its students and teachers are full of examples of standards-based feedback. At the end of each testing window, the state works to provide schools and districts with a variety of reports. These reports, as a part of TNReady, have been released near the start of the following school year. While there are a variety of reports provided, three are relevant to the topic of standards-based feedback: the individual student report, the standards analysis report, and the class roster report (Tennessee Department of Education, 2017a).

The individual student report (see figures 2.2 and 2.3) provides the students with a multitude of information. Every student receives one of these reports for every subject in which he/she was tested the previous year. On this report, instead of a percentile rank, which would be a type of norm-based feedback, the report shows a scale score and a performance level.
There are four performance levels the student could earn: below, approaching, on track, and mastered. These proficiency levels provide the student with a quick determination if he/she demonstrated his/her total understanding of the grade’s subject area. The report also provides the student with a sub score for each category tested. This provides a more in depth analysis at either what the student mastered or what needs remediation in a specific category.

**Figure 2.2 Front of TN Individual Student Report**

![Figure 2.2](image)

Figure 2.2 shows the scale score and performance level the student received, as well as an explanation of each level.
Figure 2.3 Back of TN Individual Student Report

Figure 2.3 provides the student with sub-scores on different categories in each subject tested. The student can see points earned as well as how the student did in comparison with all other students who took the test.

The standards analysis report is provided to the teacher of record for every class tested the previous year (see Figure 2.4). This standards-based report provides the teacher with class level information on how students performed by standard. While Tennessee does not release all of the actual items used in the assessment, teachers can discern how many items were assessed per standard. They can also examine how their class scored when compared with their school, district, and state.
Figure 2.4: Standards Analysis Report

Figure 2.4 shows how each teacher can see how many items were assessed per standard, as well as how the class, school, district and state performed on that standard.

Tennessee also provides a class roster (see Figure 2.5) to teachers. The class roster provides information for the entire class regarding performance level, scale score, and sub-score
category scores. Teachers can quickly ascertain how their entire class performed on the previous year’s assessment.

Figure 2.5: Class Roster Report

The report in Figure 2.5 provides the teacher with the performance level, scale score, and sub-score category scores for all the individual students in the previous year’s class.

These reports are provided by the state at the end of each school year. According to the state’s parent’s guide for 2017-2018, these scores are not projected to be sent home until the summer following the completion of the school year (Tennessee Department of Education, 2017c). The next component of standards-based evidence the state provides is in the form of an instructional focus document. While not a report on students, it provides clarity for the standards taught during the school year.
In 2017, Tennessee also released instructional focus documents for all tested math content areas, spanning from 3rd grade to Geometry. These instructional focus documents were created to provide teachers a resource that contained evidence of learning statements for each standard and instructional focus statements for each standard (Tennessee Department of Education, 2017d).

The learning statements for each standard were divided into four levels. Level 1 statements were written to show what a student who shows “minimal understanding and nominal ability” for the standard. For each increase in level, the statements would include a higher level of understanding the student had to demonstrate to fulfill that level. Levels three and four were designated as “at” or “above” the grade level expectation. Therefore, these two levels had instructional focus statements in addition to the evidence of learning statements (Tennessee Department of Education, 2017d).

This instructional focus document, provided by the state of Tennessee, was created in the fall of 2017 as part of the Mathematics Item Writer Workshops. These documents were created to provide additional support on standards where statewide data indicated students struggled. Teachers can use this document to properly assess where their students fall in their understanding of specific standards (Tennessee Department of Education, 2017b). It also provided clarity between the state and the teachers regarding the level of state expectations for end-of-year assessments. This clarity would subsequently be used between the teacher and the student to provide a stronger understanding on the student’s level of proficiency. The student and teacher could identify the student’s current state, but also be able to plan strategies to achieve the next level of proficiency (Tennessee Department of Education, 2017d).
An example of this instructional focus document is provided in Figure 2.6. The standard depicted is 7.EE.A.2, a 7th grade standard under the Expressions and Equations domain.

**Figure 2.6: Instructional Focus for Standard 7.EE.A.2**

In Figure 2.6, the standard is listed at the top and then “Evidence of Student Learning Statements” are provided for each of the four levels. Then, additional “Instructional Focus Statements” are added for Levels 3 and 4. (Tennessee Department of Education, 2017d)
School Level Examples of Standards-Based Feedback

Many schools claim to be “data-driven,” but where are they getting their data? If schools use assessment systems that assess students traditionally, it is difficult to gather data to determine student misconceptions (Marzano, 2003). According to Wainer (2011), the focus of how to use assessment information should not be data-driven but should be on decision-driven data collection. Teachers need to think prior to their data collection about how they will use these data collected as evidence of student learning. Instead, many teachers derive data from a myriad of locations and attempt to make decisions based upon these data. Marzano (2003) indicated that using data from multiple locations to make decisions was a mistake commonly made by schools attempting to improve student achievement. Using indirect data as a predictor of student achievement was compared to attempting to determine a person’s overall health by examining only his/her weight. Although a relationship could exist between a person’s weight and overall health, there are many other factors involved that must be investigated properly.

Teachers need to grade for impact. They must plan their assessments in such a way that when students complete these assessments, both the teacher and the student learn something from the results. Hierck and Lawson (2018) worked with a school district in Wisconsin to create a target-based assessment and learning system that would better articulate for both students and teachers where learning gaps occurred and what action steps could be implemented to close these gaps.

The basis of Hierck’s & Lawson’s book came from the case study of Winneconne Community School District located in northeast Wisconsin (Heirck & Lawson, 2018). The district was semi-rural, with approximately 1,600 total students in one elementary, one middle, and one high school. The school district began using target-based assessments in the 2013-14
school year. Their overall accountability score in that initial year was 72.9, but that number grew to 83.3 in 2014-15 and to 89.3 in 2016-17.

This school district endured a long process of shifting its assessment system. Before the district could develop targets for their assessments, collaborative teams had to first identify their priority standards. Then, content level groups went through the process of taking each priority standard and identified the learning targets. For example, the Wildlife, Fish, and Forestry class had seven priority standards, but they broke those seven standards out into 29 learning targets assessed throughout their class (pg. 27).

Once these schools created learning targets, then they defined proficiency. Proficiency scales were used to assess students based upon learning targets. Instead of a student receiving an 85 on a test, he/she would receive several scores based on each learning target that was assessed on the 3-point scale (proficient, approaching, needs support). Winneconne Community School District used the MasteryConnect platform to house all these data. MasteryConnect allowed teachers to grade based upon the learning targets their collaborative teams had selected and then input scores based upon those proficiency levels. Teachers had their own gradebook within MasteryConnect, so they could see all their students at a glance. Students and parents had their own portal where they could view the individual student's progress throughout the school year (Heirck & Lawson, 2018). Figure 2.7 shows a sample MasteryConnect Tracker using current and former Green Bay Packers football players. Using this tracker view, a teacher could view an entire class and view student’s progress on each standard in view. Then, upon clicking a specific standard, the teacher can view each assessment used that aligned with the standard. In the MasteryConnect platform, the teacher could also click on a student name and view his/her overall progress on all the standards in the grade.
In Figure 2.7, a class roster is shown. For every student, a teacher has a quick view of how students have performed on each standard with a “proficient,” “approaching,” and “support” level of mastery.

MasteryConnect is an online platform used to administer formative assessments and track mastery by standard. Seaman, in Fueling Student Growth (2015) stated that "MasteryConnect provides an easy way for teachers to see and plan out the standards in terms of a curriculum map, to share and collaborate as a team on that curriculum map, to quickly identify standards, and to use that plan as they teach." MasteryConnect allows teachers to establish proficiency levels for every standard in their curriculum. Teachers can use their state standards, or they can create their own standards. This was the case with the Winneconne Community School District. This district uploaded its learning targets and created assessments aligned with those targets. With each assessment, teachers could set the “proficient,” “approaching,” and “needs support” levels.

In 2010, Kentucky also developed a standards-based report card through an initiative that involved three diverse school districts. Each of these school districts had groups of teachers who
had all been working toward a standards-based report card model separately, prior to the Kentucky initiative, but they collaborated during a three-day summer workshop on standards-based report cards that was led by researchers with expertise about standards-based grading. The workshop was divided into two parts: the first focused on the challenges of standards-based grading, recommended practices, and methods of applying the established report card to students with disabilities and English learners. The second part of the workshop emphasized creation. The teachers and district leaders worked together to created two report cards: one for grades K-5 and another for grades 6-12. These pilot groups in the three Kentucky school districts would implement these report cards in the 2010-2011 school year. The report cards listed six reporting standards for each content area. For example, K-5 Math included: Operations and Algebraic Thinking, Numbers and Operations – Base 10, Numbers and Operations – Fractions, Measurement and Data, Geometry, and Mathematical Practice (Guskey, Jung, & Swan, 2011). This report card looked similar to Marzano’s standards-based report card shown previously in Figure 2.1 (Marzano, 2008, pg. 50-51).

These reporting standards were developed through the domains of Kentucky’s state standards. Through these reporting standards, students were given a level of proficiency, rated 1-4. Students, parents, and teachers had a positive reaction to these new report cards. Teachers believed the new standards-based grading was a more accurate measure of the student understanding, and the families found them easy to understand. Due to these positive findings, Kentucky decided to accelerate its implementation for the 2011-2012 school year. Several schools in the three pilot school districts decided to implement the standards-based report cards school-wide, as opposed to their original pilot groups. The revised report card forms were
presented to leadership teams from as many as 20 other Kentucky school districts in hopes of increasing participation in the initiative (Guskey, Jung, & Swan, 2011).

Another example of an individual teacher moving to a standards-based grading system and finding positive results was Myron Dueck. Dueck (2011) changed the way he graded his history classes. He structured his new framework around three questions: Where am I going? Where am I? How do I close the gap? At the beginning of each unit, he asked his students, “where am I going?” To answer this question, he would create learning targets under four categories: knowledge, reasoning, skill and product. These questions mirrored Hattie’s (2012) 3 questions on feedback: Where am I going? How am I going there? Where to next?

Following an assessment, he would ask students, “where am I?” This prompted him to change his process of assessment. Determining student deficiencies was problematic. He scrutinized the arrangement of his tests. Previously, he would group questions by type, but he subsequently grouped questions by topic. This allowed him to create an amended version of the test for students who would take a re-test to improve the overall level of mastery. When he passed back the tests, students would record their own scores within each category to determine their level of mastery on different learning targets. Students could then identify areas of weakness within the content. By this point, students were ready to answer his third question: How do I close the gap? On his tracking sheet, students would record their scores and identify if they wanted to take a retest on certain portions. He would also ask students what they did to prepare for the initial assessment and what they planned to do for the retest. Dueck noted that his lower achieving students appreciated being able to discern small areas where they did excel, and his higher achieving students noted they felt less tempted to cheat since they knew they would have a retest. He also indicated it was easier to pinpoint areas that were difficult for students.
Battistone (2018) conducted a phenomenological study on first- and second-year teachers entering a standards-based grading district. These teachers were new to their content, and recently graduated from college. They struggled with the new system. The importance of early and frequent professional development for all teachers to ensure consistency in grading and alignment with the standards was noted.

According to Peters, Kruse, Buckmiller, and Townsley (2017), many students have shown animosity towards this system. Because standards-based grading examines proficiency instead of overall percentages, students were less likely to receive an A. They performed a study on a 9-12 high school, with approximately 500 students. In this school, students were graded on a 1-4 proficiency level, where a 4 meant the student had completely mastered the topic. This 4-level proficiency translated into an A on the normal academic scale. Students were concerned with their new grades because they did not believe it was a fair comparison to other schools that were not using this method of assessment. Other evidence of grade change occurred in a large urban school district in Omaha. This district transitioned to standards-based grading in the 2009-2010 school year and determined that over a three-year period, there was an increase in percentage of B’s and C’s, but a decrease in A’s, D’s, and F’s across the district (Proulx, Spencer-May, & Westenberg, 2012).

School districts that plan to transition to this new grading system must incorporate time for training and questions. According to Pritzl (2016), district leaders must have a clear plan in place for their schools to increase likelihood of success. School leaders must take their time in the planning stages; once the plan is in place, the implementation portion of the new grading system should be the primary focus of any professional development sessions.

Summary
The history of grades and standards played a vital role in the molding of what is considered modern-day education. Educators have noted a multitude of changes to these standards. They have also seen more accountability measures instated to ensure students are learning the content expected of them. These standards and accountability measures leave teachers and students feeling overwhelmed. It is difficult to ascertain how content and material is going to be taught and retained.

This lack of clarity prompted the necessity of visible learning and its studies on the value of effective feedback. Through proper feedback, teachers and students can begin to determine where the gaps and most desperate needs are located. Standards-based feedback is a primary example of this. Students and teachers can both view the progression of learning and understand what needs to occur next. This type of feedback seeks to answer the following questions from Hattie (2012). Where am I going? How am I going to get there? Where to next?
CHAPTER 3: METHODOLOGY

This quantitative study compared the differences between the post-test gains of 7th grade math students who received standards-based feedback and those who received traditional grade feedback. There were two groups of students: those who received the treatment of standards-based feedback and the control group. The control group received the traditional feedback per the norm.

All students took pre- and post-tests on the MasteryConnect platform with Certica item bank questions. As part of the treatment group, during the experimentation phase, students also received standards-based feedback from assessments on the MasteryConnect platform, as well as teacher created assessments. All assessments used during the experimentation window were standards-aligned and students received standards-based feedback for each.

There were three research questions posed:

1. Is there a statistically significant difference in the post-test gains from students who receive individual standards-based feedback and students who receive traditional evaluation feedback on formative assessments in the 7th grade math classroom? The null hypothesis stipulated that the post-test gains for the control group and the experimental group would not be significantly different.

2. Is there a significant difference in post-test gains from female students who receive individual standards-based feedback and male students who receive individual standards-based feedback on formative assessments in the 7th grade math classroom? The null hypothesis stated that the post-test gains in the female group and the male group would not be significantly different.
3. Is there a significant difference in post-test gains from students in standard level classes who receive individual standards-based feedback and students in honors level classes who receive individual standards-based feedback on formative assessments in the 7th grade math classroom? The null hypothesis specified that the post-test gains in the standard and honors group would not be significantly different.

Population and Sample

One middle school used for this study, referred to as Middle School A. It was chosen from the four middle schools in its district because it was the largest in size. Middle School A was a rural middle school in East Tennessee. Middle School A had a population of approximately 825 students, comprised of 91.5% Caucasian students, with 4.6% Hispanic or Latino, 2.1% African American, and 0.6% Native American/Alaskan, as determined from the district’s student information system. The 7th grade class consisted of approximately 250 students. The 7th grade demographics in Middle School A included: 93.1% Caucasian, 3.9% Hispanic or Latino, 2.1% African American, and 0.8% Native American/Alaskan. The school also included 25.9% students identified as “economically disadvantaged.”

A purposeful sample of 7th grade students was utilized. At Middle School A, all three math teachers participated in the experimentation. Two teachers taught 4 classes of varying levels: inclusion, standard, advanced, and elite. Teacher A taught two elite-level classes, one standard-level class, and one inclusion-level class. Teacher B taught two advanced level classes, one standard and one inclusion. Teacher C only taught two 7th grade classes; both were standard level. To ensure a comparable control group to the experiment group, inclusion classes were not included because Teacher C did not teach any of this level. Teacher A had one elite level class in the control group and one elite level class in the sample group. In the same manner, Teacher B
had a similar schedule, but her advanced classes were utilized. Teacher C had one standard level class in the control group and one in the experiment group. Each of these classes had approximately 25 students. Therefore, there were approximately 75 students in the experiment group and 75 in the control group.

Three teachers were involved in this study. These teachers are referred to as Teacher A, Teacher B, and Teacher C. Teacher A taught at Middle School A for seven years and had a total of 16 years of experience as a middle school math teacher. She holds an administrator license and an Ed.S in Instructional Leadership. Teacher B had been at Middle School A for eight years with a total of 18 years of experience. She holds a bachelor’s degree. Teacher C taught at Middle School A for five years and had a total of nine years of experience as a middle school math teacher. She holds a bachelor's degree.

**Description of Instruments**

Teachers used the MasteryConnect platform at Middle School A to collect all data. The Certica item bank was utilized for all pre- and post-test items. Teachers were also able to use these for their own formative assessments during the experimentation window.

As displayed in Figures 3.1 and 3.2, the researcher also created a mastery tracker for students to complete as they progressed through the priority standards covered during the unit of the experiment. These priority standards, along with all others for the year, were selected by all the 7th grade math teachers within the district prior to the 2018-2019 school year, utilizing the R.E.A.L. criteria: readiness, endurance, assessment, and leverage (Ainsworth, 2013).
Research Procedures & Time Period of the Study

All students took a pre-test at the beginning of the second quarter in the fall of 2018. During the five weeks of experimentation, all classes covered the ratios and proportional relationship unit. Three priority standards were covered during this time, and are listed below:

- **7.NS.A.3**: Solve real-world and mathematical problems involving the four operations with rational numbers. (Computations with rational numbers extend the rules for manipulating fractions to complex fractions).

- **7.RP.A.2** Recognize and represent proportional relationships between quantities.
  a. Decide whether two quantities are in a proportional relationship (e.g., by testing for equivalent ratios in a table or graphing on a coordinate plane and observing whether the graph is a straight line through the origin).
  b. Identify the constant of proportionality (unit rate) in tables, graphs, equations, diagrams, and verbal descriptions of proportional relationships.
  c. Represent proportional relationships by equations. For example, if total cost \( t \) is proportional to the number \( n \) of items purchased at a constant price \( p \), the relationship between the total cost and the number of items can be expressed as \( t = pn \).
  d. Explain what a point \((x, y)\) on the graph of a proportional relationship means in terms of the situation, with special attention to the points \((0, 0)\) and \((1, r)\) where \( r \) is the unit rate.

- **7.EE.A.2** Understand that rewriting an expression in different forms in a contextual problem can provide multiple ways of interpreting the problem and how the quantities in it are related. (Tennessee Department of Education, 2018b)
Next, the students who were in the experimentation group received standards-based feedback on all assessments during this quarter. To ensure fidelity of implementation, the teachers were asked to have students complete the researcher-created mastery trackers. A sample of a mastery tracker is provided below. Three mastery trackers were used by each student for each of the priority standards covered.

Students charted their progress on each mastery tracker after they took each assessment. The pre-test administered at the start of each unit was the first assessment charted. The students received their mastery level for all three priority standards. For the pre-test, students marked their initial mastery level for all three standards on each mastery tracker.

Each assessment thereafter would be plotted as well. An assessment could cover a single priority standard or cover multiple standards. Regardless of how many standards were assessed, students would plot their individual mastery for each standard on its mastery tracker. They would make the mark on the graph for their level of mastery, and they would record the name of the assessment, mastery level, and percentage score in the table to the right of the graph.

After the students plotted each score on the graph, they flipped the paper over and determined as a class the success criteria for the standard. For each subsequent assessment, the students plotted their scores and recorded their reflections. Students proceeded through Hattie’s line of questioning: Where am I going? How am I going there? Where to next? Using this reflection piece, the researcher sought to ensure fidelity of implementation. This allowed for the focus to be on the student reflection instead of the fidelity of teacher implementation of a specific teaching strategy with fidelity.

A sample of this mastery tracker is displayed in Figures 3.1 & 3.2.
Figure 3.1 shows the mastery tracker. Students filled in the table each time they received standards-based feedback and plotted their progression on the graph.
Figure 3.2 Back of Mastery Tracker

Figure 3.2 shows the back of the mastery tracker. Teachers and students created their success criteria for each standard, following the pre-test but prior to learning the actual content. Subsequently, each time they received their standards-based feedback, they answered Hattie’s three questions about feedback: Where am I going? How am I going there? Where to next?

While the experiment group received this treatment, the control group received the traditional grade feedback normally provided by their teacher.

Analysis of Data

According to Decoster (2006), there are six steps in hypothesis testing:

1. Establish null and alternative hypotheses

2. Obtain a sample from the population of interest
3. Gather data that can be used to discriminate the null and alternative hypotheses.

4. Compute a test statistic based on the data collected.

5. Determine the p-value of your test statistic.

6. Make the statement about whether you reject or fail to reject the null hypothesis.

Accordingly, the null hypothesis stipulated that there is no statistically significant difference between the post-test gains of the group of 7th grade math students who received standards-based feedback and those who received traditional grade feedback. Next, the purposeful sample was determined from the two elite-, two advanced-, and two standard-level 7th grade math classes. A total of three classes were chosen for the experimental group and three classes were chosen for the control group. Student grades from the previous nine weeks were examined to ensure similarity among the groups used in both the experimental and control groups.

The data collected came from the pre- and post-tests administered through MasteryConnect. The difference was found between the two tests and this difference, or “post-test gain,” was used in the t-test for two independent samples. From these data points, a t- and p-value were determined. Based on these values, the researcher was able to reject or fail to reject the null hypothesis.

After the initial t-test was conducted examining the control group and the experiment group, additional t-tests were implemented to determine significance between different groups within the experiment. Additional questions that were answered included: Did the level of student impact the amount of gain? For example, did honors students demonstrate more gains than a student in a standard-level 7th grade math class? Was there a difference found when examining gender? These questions were answered as well.
Mastery trackers were used during the experimentation phase. The researcher used these to ensure fidelity to implementation and explanatory power. These mastery trackers were used for qualitative support after the quantitative analyses were completed. These minor qualitative findings were also reported in the findings section.
CHAPTER 4: RESULTS

Introduction

The purpose of this study was to determine if standards-based feedback provided students more useful information to identify and correct areas in need of refinement compared to traditional feedback of a single numerical grade with no reflection prompted by the teacher for the student. Three different teachers of three different levels of students participated in the experimental part of this study. Those three levels were standard, advanced and elite. The teachers who participated did not have previous experience with standards-based grading but used the MasteryConnect platform as their main source of assessment. This platform better enabled them to provide the necessary standards-based feedback needed for the study.

Each teacher had an experimental group and a control group; therefore, six classes were utilized for this study. Students in both groups took a pre-test at the beginning of the unit. Students in the experimental group were asked to use a mastery tracker as a control for fidelity of implementation. After each assessment, teachers had their students in the experimental group complete a reflection on their results. The students plotted their progress, and answered three questions: Where am I going? How am I going there? Where to next? Then, when students completed the unit, both the experimental and the control group took a post-test.

Teachers determined that their students had difficulty answering these questions on their own. Therefore, an adjustment was made. The teachers provided prompts for the students when these students were completing their reflections.

During the study, the teachers also fell behind in their pacing, causing them to start the unit prior to a two-week break from school. This caused the unit to be completed after the break, which resulted in a slight decline in post-test performance.
Upon completion of pre-test and post-test data, post-test gains were calculated for the experimental group and the control group. Post-test gains were also calculated by the three individual standards assessed. Multiple two-tailed, two sample unequal variance t-tests were performed to answer each research question.

**Research Question 1**

Research Question 1: Is there a significant difference in post-test gains from students who receive individual standards-based feedback and students who receive traditional evaluation feedback on formative assessments in the 7th grade math classroom?

Hypothesis:

$H_0: \mu_t \neq \mu_s$

There is a significant difference in the post-test gains from students who receive individual standards-based feedback ($\mu_s$) and students who receive traditional evaluation feedback on formative assessments ($\mu_t$).

Null Hypothesis:

$H_0: \mu_t = \mu_s$ Post-test gains are equal in the experimental (standard-based feedback) group and the control (traditional evaluation feedback) group.

The two groups analyzed in this study consisted of three different leveled 7th grade classes (standard, advanced, and elite). Each level of class was taught by a single teacher. Teacher A taught the elite classes, Teacher B taught the advanced classes and Teacher C taught the standard classes. For the first research question, all three levels were compared together, and only the control group and the tested group were compared. The tested group consisted of 55 students, while the control consisted of 58 students.
Due to the delay beginning the unit, the pre-test was administered prior to a two-week break, and the post-test was administered one month after the students had returned from break.

Results

A two-tailed, two sample unequal variance t-test was performed with the two groups post-test gains means. The t-value found from the test was 0.43 and the probability of that t-value occurring was 0.67. Because this value was greater than the α of 0.05, there was no statistical significance to reject the null hypothesis. Therefore, based upon the evidence, the standards-based feedback to the experimental group did not show a statistical difference in post-test gains from the control group. These results were similar to Winton (2015), which showed no statistical difference between the means of a traditional and a standards-based graded group of students on end of course exams.

The results of this test can be found in Table 4.1 below:

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>T-Value</th>
<th>Probability of T-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tested</strong></td>
<td>55</td>
<td>-0.49</td>
<td>2.37</td>
<td>0.43</td>
<td>0.67</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>58</td>
<td>-0.67</td>
<td>2.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>α = 0.05</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Because \( p(t) > \alpha \), there was not significant evidence to reject the null hypothesis that stated the means of the two groups’ gains were equal.
Research Question 2

Is there a significant difference in post-test gains from female students who receive individual standards-based feedback and male students who receive individual standards-based feedback on formative assessments in the 7th grade math classroom?

Hypothesis:

$H_a$: $\mu_f \neq \mu_m$

There is a significant difference in the post-test gains from female students ($\mu_f$) who receive individual standards-based feedback and male students ($\mu_m$) who receive individual standards-based feedback on formative assessments.

Null Hypothesis:

$H_0$: $\mu_f = \mu_m$

Post-test gains are equal in the female group and the male group.

The two groups analyzed for this question were both in the tested group. There were 29 females and 26 males in the three different classes that comprised the tested group. The male and female students in this tested sample were taught by three different teachers but were all utilized for this t-test.

Results

The post-test gain mean for the female tested group was −0.66 and −0.31 for the male group. After running a two-tailed, two sample unequal variance t-test, the t-value found was 0.54. The probability of that t-value was 0.59. Since the $\alpha$ was 0.05, there was no statistically significant evidence to reject the null hypothesis. Therefore, the post-test gains from the male and female group were not found to be statistically different.
The results for this analysis can be found below in Table 4.2.

### Table 4.2: T-Test Summary for Female vs. Male Post-test Gains

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>T-Value</th>
<th>Probability of T-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>29</td>
<td>-0.66</td>
<td>2.53</td>
<td>0.54</td>
<td>0.59</td>
</tr>
<tr>
<td>Male</td>
<td>26</td>
<td>-0.31</td>
<td>2.22</td>
<td>0.54</td>
<td>0.59</td>
</tr>
</tbody>
</table>

*α = 0.05

Because p (t) > α*, there was not significant evidence to reject the null hypothesis that stated the means of the two groups’ gains were equal.

### Research Question 3

Is there a significant difference in post-test gains from students in standard level classes who receive individual standards-based feedback and students in honors level classes who receive individual standards-based feedback on formative assessments in the 7th grade math classroom?

Hypothesis:

Hₐ: There is a significant difference in the post-test gains from students in standard level classes who receive individual standards-based feedback and students in honors level classes who receive individual standards-based feedback on formative assessments.

Null Hypothesis:

H₀: Post-test gains are equal in the standard group and the honors group.

This research question was analyzed via three separate t-tests. Research Question 3a was answered by comparing the standard class to the advanced class. Research Question 3b was answered comparing the standard class to the elite class. Research Question 3c was answered by
comparing the standard class to the advanced and elite classes, combined. These two classes combined were referred to as “Upper”.

Research Question 3a

Hypothesis:

\[ H_a: \mu_s \neq \mu_a \]

There is a significant difference in the post-test gains from students in standard level classes who receive individual standards-based feedback and students in advanced level classes who receive individual standards-based feedback on formative assessments.

Null Hypothesis:

\[ H_0: \mu_s = \mu_a \]

Post-test gains are equal in the standard group and the advanced group.

Results

The first t-test compared the standard level class and the advanced level class. Both classes were part of the tested group, and thus, had smaller samples. There were 19 students in the standard class and 12 students in the advanced class. The standard level class had a mean post-test gain of 0.47, while the advanced class had a mean of –1.17. The t-value calculated was 1.74 and the probability of that occurring was 0.09. While this value is not less than 0.05, for purposes of this study, the \( \alpha \) value was made to be 0.10. Therefore, at a 90% confidence level, the null hypothesis the means of the two different level classes were equal could be rejected.

The summary for this test is shown below in Table 4.3.
Table 4.3: T-Test Summary for Standard vs. Advanced Level Classes Post-test Gains

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>T-Value</th>
<th>Probability of T-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>19</td>
<td>0.47</td>
<td>2.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td>12</td>
<td>-1.17</td>
<td>2.37</td>
<td>1.74</td>
<td>0.09</td>
</tr>
</tbody>
</table>

*α = 0.10

Two-Tailed Test: Because p (t) < α*, there was significant evidence to reject the null hypothesis that stated the means of the two groups’ gains were equal.

Research Question 3b

Hypothesis:

Hₐ: µₛ ≠ µₑ

There is a significant difference in the post-test gains from students in standard level classes who receive individual standards-based feedback and students in elite level classes who receive individual standards-based feedback on formative assessments.

Null Hypothesis:

H₀: µₛ = µₑ

Post-test gains are equal in the standard group and the elite group.

Results

The standard class results were the same as reported from the previous t-test, but for this analysis, they were compared to the elite class. There were 24 students in the elite class who received standards-based feedback, and their post-test gains average was –0.92. The t-value calculated for this two-tailed, two sample unequal variance t-test was 1.88. The probability of
this t-value occurring was 0.07. While this value is not less than the typical $\alpha$ value of 0.05, the probability is less than 0.10. Therefore, for this t-test, the $\alpha$ value was 0.10, and based upon the results, the null hypothesis was able to be rejected. Thus, this demonstrated there was statistically significant evidence that the two groups analyzed had different means.

The summary for the t-test completed on these two groups are summarized below in Table 4.4.

**Table 4.4: T-Test Summary for Standard vs. Elite Level Classes Post-test Gains**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>T-Value</th>
<th>Probability of T-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Standard</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>0.47</td>
<td>2.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Elite</strong></td>
<td>24</td>
<td>-0.92</td>
<td>1.74</td>
<td>1.88</td>
<td>0.07</td>
</tr>
</tbody>
</table>

*$\alpha = 0.10$*

*Two-Tailed Test: Because $p(t) < \alpha^*$, there was significant evidence to reject the null hypothesis that stated the means of the two groups’ gains were equal.*

**Research Question 3c**

Hypothesis:

$H_a: \mu_s \neq \mu_u$

There is a significant difference in the post-test gains from students in standard level classes who receive individual standards-based feedback and students in upper level classes (advanced and elite) who receive individual standards-based feedback on formative assessments.

Null Hypothesis:

$H_0: \mu_s = \mu_u$

Post-test gains are equal in the standard group and the upper group.
Results

The final t-test conducted on the tested groups post-test gains scrutinized the standard level class compared to the advanced and the elite classes combined, designated as the “Upper” level classes. The mean for the standard group remained at 0.47, while the mean for the upper level group was –1.00. The t-value found from the two-tail, two sample unequal variance t-test was 2.03, and the probability of that t-value occurring was 0.052. While this value is not less than the typical α value of 0.05, for purposes of this t-test, the α value was changed to 0.10. Therefore, the null hypothesis, stating post-tests means of the two groups were equal, was able to be rejected.

The summary of results for the t-test can be found in Table 4.5.

Table 4.5: T-Test Summary for Standard vs. Upper (Advanced & Elite) Level Classes

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>T-Value</th>
<th>Probability of T-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>19</td>
<td>0.47</td>
<td>2.84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td>36</td>
<td>-1.00</td>
<td>1.94</td>
<td>2.03</td>
<td>0.052</td>
</tr>
</tbody>
</table>

*α = 0.10
Two-Tailed Test: Because p (t) < α*, there was significant evidence to reject the null hypothesis that stated the means of the two groups’ gains were equal.

Additional Results

Upon further investigation into the results from the post-test, there were also some significant findings based upon individual standard. Different levels of classes seemed to have higher post-test gains on individual standards. The three standards assessed included: 7.NS.A.3,
7.RP.A.2, and 7.EE.A.2. The two standards that yielded statistically significant results were 7.RP.A.2 and 7.EE.A.2. As mentioned previously, the 7.RP.A.2 standard detailed recognizing and representing proportional relationships between quantities, while 7.EE.A.2 involved rewriting expressions in different forms to solve.

Hypothesis:

\( H_a: \mu_s \neq \mu_a \)

There is a significant difference in the post-test gains under the 7.RP.A.2 standard from students in standard level classes who receive individual standards-based feedback and students in advanced level classes who receive individual standards-based feedback on formative assessments.

Null Hypothesis:

\( H_0: \mu_s = \mu_a \)

Post-test gains under the 7.RP.A.2 standard are equal in the standard group and the advanced group.

The standard level class had 19 students, with an average of 0.26 post-test gain under the 7.RP.A.2 standard. The advanced class had 12 students with –0.67 as their average post-test gain for the ratios and proportion standard. The two-tailed, two sample unequal variance t-test used found a t-value of 1.86 and a probability of that t-value occurring to be 0.08. Similar to previous results, the \( \alpha \) value used for this test was 0.10. Therefore, there was statistically significant evidence to reject the null hypothesis which stipulated that the average gains for the two groups would be equal. The summary for the t-test can be found in Table 4.6.
Table 4.6: T-Test Summary for Standard 7.RP.A.2 Standard vs. Advanced Post-test Gains

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>T-Value</th>
<th>Probability of T-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>19</td>
<td>0.26</td>
<td>1.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Advanced</td>
<td>12</td>
<td>-0.67</td>
<td>1.37</td>
<td>1.86</td>
<td>0.08</td>
</tr>
</tbody>
</table>

*α = 0.10 Because p(t) < α*, there was significant evidence to reject the null hypothesis that stated the means of the two groups’ gains were equal.

Hypothesis:

Hₐ: μₛ ≠ μᵤ

There is a significant difference in the post-test gains under the 7.RP.A.2 standard from students in standard level classes who receive individual standards-based feedback and students in upper level classes who receive individual standards-based feedback on formative assessments.

Null Hypothesis:

H₀: μₛ = μᵤ

Post-test gains under the 7.RP.A.2 standard are equal in the standard group and the upper group.

The standard level class had 19 students and an average of 0.26 gain on their post-test.

The upper level classes, which include both the advanced and elite classes, had a total of 36 students with an average –0.42 drop in post-test score. A two-tailed, two sample unequal variance t-test was performed, and the t-value found from this test was 1.81. The probability of that occurring was 0.08. The α value used for this t-test was 0.10, therefore, the probability of the t-value occurring was less than the α value, allowing the null hypothesis to be rejected. This
means that the post-test gains under the standard 7.RP.A.2 for the standard versus the upper level classes were significantly different, thus not resulting from chance.

The summary of the test can be found in Table 4.7.

**Table 4.7: T-Test Summary for Standard 7.RP.A.2 Standard vs. Upper Level Post-test Gains**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>T-Value</th>
<th>Probability of T-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>19</td>
<td>0.26</td>
<td>1.33</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td>36</td>
<td>-0.42</td>
<td>1.32</td>
<td>1.81</td>
<td>0.08</td>
</tr>
</tbody>
</table>

*α = 0.10

Because \( p(t) < \alpha \), there was significant evidence to reject the null hypothesis that stated the means of the two groups’ gains were equal.

Differences were also noted in another standard, 7.EE.A.2. For this standard, the two groups of focus included: standard versus elite and advanced versus elite.

Hypothesis:

\( H_a: \mu_s \neq \mu_e \)

There is a significant difference in the post-test gains under the 7.EE.A.2 standard from students in standard level classes who receive individual standards-based feedback and students in elite level classes who receive individual standards-based feedback on formative assessments.

Null Hypothesis:

\( H_0: \mu_s = \mu_e \)

Post-test gains under the 7.EE.A.2 standard are equal in the standard group and the elite group.

The standard group had 19 students with an average of 0.63 post-test gains in the 7.EE.A.2 category. The elite group had 24 students and they had an average 0.04 loss on their
post-tests under the 7.EE.A.2 category. The two-sample t-test resulted in a t-value of 1.82 and the probability of that occurring was 0.08. Using an α value of 0.05, the null hypothesis can be rejected. Therefore, there was a significant difference between the standard and elite group in the 7.EE.A.2 category.

The summary for the t-test can be found in Table 4.8.

**Table 4.8: T-Test Summary for Standard 7.EE.A.2 Standard vs. Elite Level Post-test**

<table>
<thead>
<tr>
<th>Gains</th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>T-Value</th>
<th>Probability of T-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard</td>
<td>19</td>
<td>0.63</td>
<td>1.21</td>
<td>1.82</td>
<td>0.08</td>
</tr>
<tr>
<td>Elite</td>
<td>24</td>
<td>-0.04</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*α = 0.10  
Because p (t) < α*, there was significant evidence to reject the null hypothesis that stated the means of the two groups’ gains were equal.

The final analysis performed was for the advanced class compared to the elite class for the 7.EE.A.2 standard. The advanced class had 12 students with an average of 0.75 gain on their post-test, while the elite class had 24 students with an average of 0.04 loss on their post-test. The t-value found from the two-tailed, two sample unequal variance t-test was 2.14. The probability of this occurring was 0.04. With α = 0.05, then p(t) < α, which proved significant evidence that the advanced and the elite classes had significantly different means for their post-test gains under the 7.EE.A.2 standard.

**Hypothesis:**

\[ H_a: \mu_a \neq \mu_e \]
The post-test gains under the 7.EE.A.2 standard from students in advanced level classes who receive individual standards-based feedback was not equal to the students in elite level classes who receive individual standards-based feedback on formative assessments.

Null Hypothesis:

\[ H_0: \mu_a = \mu_e \]

Post-test gains under the 7.EE.A.2 standard in the advanced group were equal to the post-test gains in the elite group.

The summary of the results from these tests can be found in Table 4.9.

**Table 4.9: T-Test Summary for Standard 7.EE.A.2 Advanced vs. Elite Level Post-test Gains**

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>T-Value</th>
<th>Probability of T-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced</td>
<td>12</td>
<td>0.75</td>
<td>0.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Elite</td>
<td>24</td>
<td>-0.04</td>
<td>1.2</td>
<td>2.14</td>
<td>0.04</td>
</tr>
</tbody>
</table>

*α = 0.05

Because \( p(t) < \alpha \), there was significant evidence to reject the null hypothesis that stated the means of the two groups’ gains were equal.

**Summary**

In summation, there were nine two sample t-tests performed and reported. While there were many other t-tests performed, the tests reported were those that showed any significant evidence. Research Question 1 and Research Question 2 failed to produce enough evidence to reject the null hypothesis.
Research Question 3 provided some significant evidence. This evidence indicated the standard level group had higher post-test gains than the other two groups. Findings also resulted from scrutinizing each standard individually.
CHAPTER 5: CONCLUSIONS, IMPLICATIONS, & RECOMMENDATIONS

Assessments provide the student and teacher with evidence of the student understanding of the content covered. Students may need better, more precise information from their assessment results. The type of feedback provided, and subsequent student reflection are sources of this information. Marzano (2010b) discussed the benefits of students tracking their own progress on standards. On average, this resulted in a 32 percentile point gain when used. For students to clearly track their progress in understanding a specific topic, students should be provided with a standards-based feedback to better identify areas of focus. Hattie (2012) focused on the path feedback should provide to a student. Feedback should prompt the students to answer three questions: “Where am I going? How am I going there? Where to next?” (pg. 131-135).

Using this as the foundation, this study focused upon standards-based feedback provided in the 7th grade math classroom. This study sought to determine if students who received traditional, single numerical grade feedback would progress in their learning in a similar fashion to those students who received standards-based feedback and had time for reflection upon that feedback.

Summary

There were three levels of classes used for this study: standard, advanced, and elite. Each level had one class in the tested group (receiving standards-based feedback) and one class in the control group (receiving traditional feedback). Each level of class was taught by a different teacher. Teacher A taught the elite class. Teacher B taught the advanced class. Teacher C taught the standard level class. All three levels of classes were located at Middle School A, a rural middle school located in East Tennessee with a population of approximately 825 students.

Each research question in this quantitative study was analyzed using a two-tailed, two-sample of unequal variance t-test. The three research questions can be found below:
1. Is there a statistically significant difference in the post-test gains from students who receive individual standards-based feedback and students who receive traditional evaluation feedback on formative assessments in the 7th grade math classroom?
   a. The null hypothesis stipulated that the post-test gains for the control group and the experimental group would not be significantly different.

2. Is there a significant difference in post-test gains from female students who receive individual standards-based feedback and male students who receive individual standards-based feedback on formative assessments in the 7th grade math classroom?
   a. The null hypothesis stated that the post-test gains in the female group and the male group would not be significantly different.

3. Is there a significant difference in post-test gains from students in standard level classes who receive individual standards-based feedback and students in honors level classes who receive individual standards-based feedback on formative assessments in the 7th grade math classroom?
   a. The null hypothesis specified that the post-test gains in the standard and honors group would not be significantly different.

Question 1 focused on the comparison of the tested versus the control group, but the last two questions focused on subgroups within the tested group. There were also additional findings focused on the tested subgroups and their success with specific standards.
Conclusions

Research Question 1

Based upon the results of the t-test comparing the results of the tested group and the control group, there was not enough evidence to reject the null hypothesis. The probability of the two groups resulting in the same post-test gains was very high, resulting in a failure to reject the null hypothesis.

Research Question 2

Upon investigation of the tested male and female subgroups, there was no significant evidence to indicate one subgroup would respond better to standards-based feedback. The results of the t-test failed to reject the null hypothesis.

There could have been one possible cause for the failure to reject the null hypothesis for Research Question 1 and Research Question 2. As mentioned previously, an unpredicted obstacle in this study was the delay in completing the unit. It is feasible that the first two research questions may have had different results if the students did not have a two-week break in the middle of the unit. However, retention could also be introduced. Using this method of standards-based feedback and reflection did not provide evidence to support this method as a means for furthering retention of the topics covered.

Research Question 3

The third research question provided evidence to reject the null hypothesis. The post-test gains in the tested standard level class showed a statistically significant difference between the advanced and elite level classes. One t-test was used to compare the standard to the advanced, and another was used to compare the standard to the elite. Then, a third t-test was used to compare the standard to the “upper” level classes, which included both the advanced and the elite
level classes. All three t-tests showed evidence to reject the null hypothesis that stated the post-test gains of two groups would be equal.

There are multiple possible causes for equal gains. The standard group’s pre-test average (6.21) was initially lower than the advanced (8.42) and the elite (8.92), which would allow for more room for growth. The other cause could be the fidelity of implementation. When the study was initiated, all three teachers had the same lack of experience with standards-based grading. Teacher C, the standard level teacher, had her students complete the mastery tracker the most.

**Additional Findings**

Additional t-tests were used to identify whether specific standards covered during the unit evidenced more post-test gains for each of the tested subgroups. There were three standards covered during the unit: 7.NS.A.3, 7.RP.A.2, and 7.EE.A.2. Only two of the three standards showed statistical significance for the groups who received standards-based feedback. The post-test gains for each standard were analyzed for the standard, advanced, and elite classes. The 7.NS.A.3 standard did not show any statistically significant difference between any two groups tested. The other two standards, however, demonstrated evidence of a difference in their post-test gains in the 7.RP.A.2 and the 7.EE.A.2 standard.

Upon investigation, the standard and advanced classes had a significant difference in post-test gains under the 7.RP.A.2 standard, as did the standard versus the “upper” level classes. The standard and elite classes had a significant difference under the 7.EE.A.2 standard, as did the advanced and elite classes.

The evidence detailed 7.RP.A.2 and 7.EE.A.2 as the two standards more likely for improvement. One reason for 7.NS.A.3 not showing evidence similar to the other two standards was that this is a number system standard. This standard focuses more on the computation of
complex fractions. Students may have had a stronger foundation on this topic, which was addressed in 6th grade. This standard was also covered only in the time prior to the two-week break. The other two standards that showed more improvement were deeper and more conceptual standards. Understanding ratios and proportions (7.RP.A.2) was covered both before and after the two-week break, while rewriting expressions (7.EE.A.2) was covered solely after the two-week break.

Implications

There were three main findings during this study. The two-week break and the drop in some post-test scores highlighted the issue of retention of the content. There was a significant difference between the tested standard group and the tested “upper” level groups. Finally, the last main finding from this study came from the discovery of the student’s difficulty in self-reflection.

The first finding resulted from facing the unforeseen obstacle of a two-week break in the middle of the unit. Did the test scores fall from this? This is feasible, but there is another implication. The standards-based feedback provided to the students did not increase their retention of the content. This could be a case for future study, focusing on retention over a longer time frame.

The second finding was the significant difference in post-test gains from the standard group in comparison with the other two tested groups. At the start of this study, this result could only be attributable to the level of the students, because all three teachers had similar, minimal experience with standards-based grading. However, during the study, Teacher C, the standard level teacher, began to show more of an interest in standards-based grading. She would have her students fill out their mastery tracker on a more frequent basis and her assessments were more
standards-based in structure. She would group questions for one standard together, so the students would be familiar with the standard when assessed. Another case study could focus on one teacher who teaches multiple levels of students to compare only the level of student.

Finally, the third finding was the discovery of how students found it difficult to self-reflect and self-regulate. After students would mark their scores on their mastery trackers, they would answer three questions: Where am I? How am I going there? Where to next? Upon the first session of plotting and reflection, the teachers discovered their students could not answer these questions independently. Prompts had to be provided for each question. An example of one of the prompts is shown in Table 5.1.
<table>
<thead>
<tr>
<th>Where Am I Going?</th>
<th>How Am I Going There?</th>
<th>Where To Next?</th>
</tr>
</thead>
<tbody>
<tr>
<td>If I missed these questions….</td>
<td>Then I need to work on this topic…</td>
<td>Pick one: To get better at my weakness, on my own, I will…</td>
</tr>
<tr>
<td>• 1, 14, 15</td>
<td>• Identifying &amp; setting up an equation for a proportional relationship</td>
<td>• Rework incorrect problems</td>
</tr>
<tr>
<td>• 3</td>
<td>• Applying ratios</td>
<td>• Explain my mistakes to a peer</td>
</tr>
<tr>
<td>• 7, 10, 11</td>
<td>• Unit rate/constant of variation with a graph</td>
<td>• Work on ALEKS path for extra problems (looking at the fraction or ratio &amp; proportion piece of the pie)</td>
</tr>
</tbody>
</table>

Because self-reflection and self-regulation were problematic for students, it was more difficult for them to implement their action plan under the “How am I going there?” question. A possible future study could include focusing more on the preparation of students in their own efforts to self-regulate. The teacher could provide more structured options that allowed students to have choices, but necessitated follow-up with these choices.
**Recommendations**

This study prompted three main ideas for future study. The first included the focus on retention of content when students receive standards-based feedback. The second focused on the level of study, ensuring there would be a control for the teacher. It is recommended that only a single teacher, who has multiple level classes, be involved in the study. It is also recommended that a study focused on the preparation of the student for self-reflection and self-regulation be conducted. This study did not focus on enabling the students to guide their own learning progression; another study could focus on empowering students to do so.

**Conclusion**

This study investigated standards-based feedback in the 7th grade math classroom. Instead of receiving the traditional single, numerical grade on an assessment, students received standards-based feedback on each standard assessed. Students also tracked their progress in each standard and answered three reflection questions for every assessment. It was determined there was no significant difference between the tested group and control group, as evidenced from a t-test, on post-test gains. There was also no significant difference between the tested male and female subgroups. Significant differences were found when comparing the three levels of students in the tested group. Multiple t-tests showed the standard level group had a significant difference in post-test gains when compared with the advanced and elite level classes.
References


Ainsworth, L. (2013). Prioritizing the Common Core: Identifying specific standards to emphasize the most. Englewood, CO: The Leadership and Learning Center


Finch, M. A. (2017). Tennessee to the top: One state’s pursuit to win Race to the Top. 


Kebede, L. (2017). In final years of Common Core, Tennessee teachers can use practice test questions from PARCC. *Chalkbeat.* Retrieved from


Appendix A

NS.A.3 Mastery Tracker
Mastery Tracker: 7.NS.A.3
Finding Unit Rate With Complex Fractions

Solve real-world and mathematical problems involving the four operations with rational numbers. (Computations with rational numbers extend the rules for manipulating fractions to complex fractions.)

<table>
<thead>
<tr>
<th>Assessment #</th>
<th>Assessment Name</th>
<th>Assessment Score</th>
<th>Mastery Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Reflection: 7.NS.A.3

Finding Unit Rate with Complex Fractions

Solve real-world and mathematical problems involving the four operations with rational numbers. (Computations with rational numbers extend the rules for manipulating fractions to complex fractions.)

<table>
<thead>
<tr>
<th>How Will I Know I Am Successful?</th>
<th>Where Am I Going?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I can find simple unit rates</td>
<td></td>
</tr>
<tr>
<td>2. I can find unit rate from complex fractions</td>
<td></td>
</tr>
<tr>
<td>3. I can convert unit rates</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How Am I Going There?</th>
<th>Where To Next?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix B

RP.A.2 Mastery Tracker
# Mastery Tracker: 7.RP.A.2

Recognizing & Representing Proportional Relationships

Recognize and represent proportional relationships between quantities.

<table>
<thead>
<tr>
<th>Assessment #</th>
<th>Assessment Name</th>
<th>Assessment Score</th>
<th>Mastery Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**RECOGNIZING & REPRESENTING PROPORTIONAL RELATIONSHIPS**

Recognize and represent proportional relationships between quantities.

<table>
<thead>
<tr>
<th>How Will I Know I Am Successful?</th>
<th>Where Am I Going?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. I can identify proportional relationships</td>
<td>Where To Next?</td>
</tr>
<tr>
<td>2. I can use graphs to determine if a relationship is proportional</td>
<td></td>
</tr>
<tr>
<td>3. I can solve proportional relationships</td>
<td></td>
</tr>
<tr>
<td>4. I can find the constant rate of change</td>
<td></td>
</tr>
<tr>
<td>5. I can describe the rate of change as slope and what it represents</td>
<td></td>
</tr>
<tr>
<td>6. I can determine if a linear relationship is a direct variation</td>
<td></td>
</tr>
</tbody>
</table>
Appendix C

EE.A.2 Mastery Tracker
Front of EE.A.2 Mastery Tracker

MASTERY TRACKER: 7.EE.A.2
REWITING EQUATIONS

Understand that rewriting an expression in different forms in a contextual problem can provide multiple ways of interpreting the problem and how the quantities in it are related.
REFLECTION: 7.EE.A.2
REWITING EQUATIONS
Understand that rewriting an expression in different forms in a contextual problem can provide multiple ways of interpreting the problem and how the quantities in it are related.

<table>
<thead>
<tr>
<th>How Will I Know I Am Successful? What are my success criteria?</th>
<th>Where Am I Going?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How Am I Going There?</th>
<th>Where To Next?</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix D

Permission Form from Director of Schools
Permission Form from Director of Schools

I, [Name], Director of Blount County Schools, grant Karoline Stache with permission to use Heritage Middle School 7th Grade as part of her case study for her dissertation, “Standards Based Feedback in the 7th Grade Math Classroom.” She can collect and report pre- and post-test data obtained from the study in her dissertation.

Director of Schools

[Signature]
Appendix E

Permission Form from Teachers
Permission Form from Teachers

Teacher A
I, __________________, agree to participate in the study, Standards Based Feedback in the 7th Grade Math Classroom. I will administer pre- and post-test assessments and use the mastery tracker throughout the unit of study. I agree to let Karoline Stache use the data from this study in her dissertation.

______________________________
______________________________

Teacher B
I, __________________, agree to participate in the study, Standards Based Feedback in the 7th Grade Math Classroom. I will administer pre- and post-test assessments and use the mastery tracker throughout the unit of study. I agree to let Karoline Stache use the data from this study in her dissertation.

______________________________
______________________________

Teacher C
I, __________________, agree to participate in the study, Standards Based Feedback in the 7th Grade Math Classroom. I will administer pre- and post-test assessments and use the mastery tracker throughout the unit of study. I agree to let Karoline Stache use the data from this study in her dissertation.

______________________________
______________________________