STANDARDS-BASED GRADING: DOES IT GAUGE STUDENTS' UNDERSTANDING BETTER THAN THE TRADITIONAL REPORT CARD?

A Dissertation Submitted
to the Center for Leadership and Learning
Arkansas Tech University

in partial fulfillment of requirements
for the
EdD Degree

Ruby Earlene Dean

ACKNOWLEDGEMENTS

To my loving husband, Rodney - you have been my faithful supporter and partner in this journey. I cannot begin to count the times you managed all the responsibilities of our home while I managed my career and furthered my education. You are an amazing person and a wonderful example of a loving and committed man. Your encouragement, strength, and support have carried me through this process. You are the love of my life, and I would not be the woman I am today without you.

To my dad – At a very young age you taught me the importance of an education. You were my biggest fan and wanted a better life for me and my family. The foundation you laid for me has supported me, encouraged me, and made me a stronger person. For this, I am truly grateful. I hope that with all my education endeavors that I have made you proud. I wish you were here with me to celebrate this accolade.

To my dissertation committee – Dr. Steve Bounds, Dr. John Freeman, and – I gratefully acknowledge your hard work and dedication in supporting your students. I have been challenged and pushed to the limits I never even knew I had. Thank you for pushing me and encouraging me on this journey. I will be forever grateful.

To my good friends and classmates, I sincerely thank you for encouraging me and having faith in me when I did not. Your efforts to encourage me, sustain, and nurture me to grow professionally and personally are greatly appreciated.

Abstract

STANDARDS-BASED GRADING: DOES IT GAUGE STUDENTS' UNDERSTANDING BETTER THAN THE TRADITIONAL REPORT CARD?

Ruby Earlene Dean

Table of Contents

Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

Chapter I: Introduction

The standards-based reform movement began more than 20 years ago with the Elementary and Secondary Education Act (ESEA), which focused on setting challenging and rigorous content standards for all students. Along with content area standards are performance standards which give "concrete examples and explicit definitions of what students have to know and be able to do" to show mastery of the content standard (Goals, 2000). December 2015, President Barack Obama signed Every Student Succeeds Act (ESSA) into law. The ESSA is a reauthorization of the Elementary and Secondary Education Act (ESEA), which was signed into law by President Lyndon Johnson in 1965. ESEA was a civil rights law focused on providing grant funding to school districts serving low-income students as well as scholarships for low-income students. Every Student Succeeds Act also centers on special education and improving the quality of elementary and secondary education (The U.S. Department of Education, 2016).

President George Walker Bush signed the No Child Left Behind Act (NCLB) into law on January 8, 2002. It called for an increased role of the federal government in holding school personnel responsible for the academic growth of students, with a focus on certain subpopulations such as English-language learners, special education students, low-socioeconomic students, and minority students (The U.S. Department of Education, 2016). The caveat was that state leaders could have complied with the new requirements; however, if the states chose not to comply, they risked losing federal Title I money (Klein, 2015).

Under NCLB, students in grades three through eight and once in high school were tested in reading and math annually. District leaders then reported these test results and

included data on the identified subpopulations. Each year, district leaders must show they have reached a goal of adequate yearly progress or AYP. The NCLB required that all students reach proficiency by the 2013-2014 school year. In addition, the law states school leaders not meeting their AYP be subject to expanding sanctions that could include the state taking over the school (Klein, 2016).

According to Vatterott, with the passing of these educational laws, school leaders can no longer fail students who do not learn and move on. All students must now be proficient in the areas tested on the state achievement tests. Teachers are held accountable for student achievement, and grades must be more reflective of learning. No Child Left Behind initiatives have exposed traditional grading practices that may no longer be effective in measuring student progress in the classroom because they do not equate or correlate with performance on standardized tests (Vatterott, 2015). Additionally, academic performance has been used as a mechanism for managing behavior (Guskey, 2009; Vatterott, 2015). "If your grading system does not guide students toward excellence, it's time for something completely different" (Scriffiny, 2008, p. 70).

Background of the Problem

Public school leaders around the country have historically used the traditional grading scale using numbers, letters, or percentages. Developed in the 1940s, grades were mainly invented to make it easy for institutional leaders to communicate with one another about a student's academic readiness (Lassahn, n.d.). When there was no standard letter grading system, most teachers relied on a 100-point scale system (Effects of Grading System on Students: Pros and Cons, n.d.). With the traditional grading scale, everyone can understand what a simple grade means and is familiar with the concept (Hobden,

2019). Ideally, a grade is used to communicate to varying stakeholders some information about student learning (Marzano, 2006). It also allows a clear path for failing or passing a grade. The object of grading is to ensure that a student knows and has mastered the standards being taught. These grades are expected to reflect the learning that has taken place during the reporting period. The traditional grading scale is also very subjective and can be manipulated while teachers do attempt to assign fair, accurate, and meaningful grades. While minute details of traditional grading vary from teacher to teacher, the basic principle is that students receive a numeric grade for each assignment/test assigned and completed. Students may also earn points for cooperative learning in group projects or other types of activities.

According to Guskey and Brookhart (2019), “Grades are the symbols assigned to individual pieces of student work or to composite measures of student performance created for report cards and other summative documents” (p. 1). These grades accumulate during the grading period. The points are added together at the end of the term, and the teacher assigns a grade based on the total number of points earned (Marzano & Heflebower, 2011). A singular grade does not explain what a student(s) is learning, struggling with, where they excelled, or show weaknesses and/or strengths. There is no explanation behind the assessment/assignment, and the student may have a limited understanding of the content presented during the term or may even have learning gaps. Grades are a way of communicating feedback to students. However, this feedback does not give students the knowledge to clarify their learning (Bowman, 2020). "We take grading on last and always with some reluctance, because changing grading policies and practices means challenging some of education's longest-held traditions" (Guskey &

Brookhart, 2019, p. 1). Because traditional grading may not accurately reflect what students know or are learning at any given time, more research needs to be done to understand if alternative measures, such as standards-based grading system, more accurately reflect what the students are learning and how this is communicated to the parents, students, and teachers.

Problem Statement

The purpose of assigning grades to student learning varies from educator to educator. Leading researchers such as Doug Reeves, Thomas Guskey, and Ken O'Connor have demonstrated that grading practices fluctuate significantly from school to school and educator to educator based on a myriad of factors, including the lack of unified assessment and grading policies (Guskey, 2001; O'Connor, 2018; Reeves, 2004). Wormeli (2018) explains that allowing educators to decide how to assign grades to student learning may not be an accurate picture of what students know, understand, and be able to do. Teachers' beliefs in what to grade and to record are significant influences on what a student earns as a grade (Brookhart, 2017; Guskey, 2015; O'Connor 2011). Allowing teachers to assign grades or what work needs to be graded can be very subjective. Traditional grading systems utilize an A, B, C, D, or F or similar scales to denote student understanding of all content standards. In contrast, standards-based grading systems utilize a reporting system based on individual content standards (Guskey, 2015; O'Connor, 2011; Heflebower, Hoegh, Warrick, & et al, 2014). "More and more educators are beginning to question traditional grading practices that were developed to sort students into learners and non-learners, not to support learning for all" (Brookhart, 2011, p. 10).

The traditional grading practice tends to lump content with effort and behavior into one letter grade (Brookhart, 2011). All too often, traditional grading practices in the United States are based on instructional and motivational principles that cause some students to give up in hopelessness and accept failure rather than driving them toward academic success (Stiggins, 2014). More recently, the Common Core State Standards emphasized conceptual understanding, procedural skills, and increased content rigor (Briars & Foster, 2012). Content standards and common assessments improved consistency and coherency in curriculum and instruction, but grades and grading remained in the hands of the individual teacher (Wormeli, 2018) and can be seen as subjective. To make grades more meaningful, issues related to purpose and reporting format must also change (Guskey, 2015).

Using an SBG model, students have clear learning targets, a distinction exists between formative and summative work, grading is based on rubrics, and feedback is reported according to the learning targets (O'Connor, 2018; Reeves, 2011; Vatterott, 2015). A final grade determination is made based on evidence of student achievement according to proficiency targets tied to educational standards instead of percentages or letter grades (Welsh & D'Agostino, 2009).

Purpose of the Study

The purpose of this study is to determine whether the type of grading system (traditional or standards-based) used to assess student progress makes a difference on ACT Aspire results. School leaders can support policies and practices concerning classroom grading by determining the relationship between grading system and achievement. This research may bring validity to district leaders using standards-based

grading systems. Results from this study can be used to inform the decision-making process at the researcher's own school within the district as leaders transition from traditional grading to standards-based grading per the Professional Learning Community (PLC) at the elementary level.

Rationale

This study will be conducted to provide school district leaders with information that might enable teachers and administrators to align grading systems, classroom grading, and standardized assessment to meet the learning needs of students. While classroom grading systems and grades are intended to reflect student learning, researchers have long held that neither are measures of academic knowledge alone. The use of other factors, like behavior, attendance, and attitude, often does not give a clear understanding of grades. Standardized assessments are seen, by most researchers, as an effective means to determine students' understanding of concepts related to state standards. This study will provide quantitative data that reflect the relationship between grading systems and grades and the results produced through a standardized assessment. One aspect of the research will focus on the relative difference in the relationship between standards-based grading systems and traditional letter-based systems to the standardized assessment. The results from this study should inform district leaders of the possible benefits of one grading system over the other.

Research Questions/Hypotheses

RQ1: Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems?

RQ2: Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of socioeconomic status?

RQ3: Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of ethnicity?

RQ4: Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of gender?

Theoretical Framework/Conceptual Framework

This quantitative study will explore the differences in student achievement as measured by the ACT Aspire test in the state of Arkansas by comparing student results in districts where school leaders use a traditional grading system compared with districts where leaders use a standards-based grading system for 5^{th}-grade math. This study will review public documents reporting the school's demographics, standardized test scores, and Arkansas ESSA School Report Card rating.

The conceptual framework for this study was developed from theories on the Zone of Proximal Development (ZPD) and Bloom's mastery learning. Effective feedback utilizing SBG through a mastery-based learning instructional and assessment model would improve student motivation and self-efficacy, providing more opportunities for students to learn in the ZPD and resulting in increased student growth. The use of SBG pushes an instructional model that focuses on feedback according to curricular standards

throughout an entire instructional and assessment cycle or unit. The use of SBG is supported by both.

Lev Vygotsky developed theory of Zone of Proximal Development (ZPD) in the early 20th century. He asserted cognitive development in early childhood is advanced through social interaction with other people, particularly those who are more skilled (McLeod, 2020). Vygotsky proposed that learning comes before cognitive development in children and that children construct knowledge actively (McLeod, 2020). He learned that children who are in their zone of proximal development can almost perform the given task, but not without the help of others. Getting instructions from another person is important to cognitive development in early childhood (McLeod, 2020).

"As children are given instructions or shown how to perform certain tasks, they organize the new information received in their existing mental schemas. They use this information as guides on how to perform these tasks and eventually learn to perform them independently" (McLeod, 2020). Children learn through social interaction that includes collaborative and cooperative dialogue with someone more skilled in the tasks they are trying to learn (McLeod, 2020). As a child becomes more familiar with the task, less guidance can be provided, and the child should be able to perform the task by themselves.

Mastery learning or learning for mastery is an instructional strategy and educational philosophy first proposed by Bejamin Bloom in 1968 (Guskey, 2010). A student must perform a level of mastery at 90% in prerequisite knowledge before moving forward to learn subsequent material. If mastery is not achieved on the test, students are given additional support, review the information, and test again. The cycle continues until the

student accomplishes mastery. Once mastery is achieved, the student moves forward to the next stage. The theory of mastery learning stems from the work of Benjamin S. Bloom . His work looked at how teachers might adapt the most powerful aspects of tutoring and individualized instruction to improve student learning in general education classrooms. Bloom suggested that although students vary widely in their learning rates and modalities, if teachers could provide the necessary time and appropriate learning conditions, nearly all students could reach a high level of achievement (Guskey, 2010).

A major contribution of Vygotsky's theory of cognitive development in children is the acknowledgment of the social component in both cognitive and psychosocial development (McLeod, 2020) is especially helpful in forming the research questions for this study. Teachers giving students instructions on how to perform tasks and how the students learn these tasks, and how well they perform the tasks is a direct link to Vygotsky's theory of how a child learns. ZPD is a good fit for this study because the subjects of the study are teachers, students, and parents who depend on a standards-based grading system to inform them of how well the student is performing and what tasks the students need to improve in. ZPD and Bloom's mastery learning theory inform this research because the purpose is to examine how well a standards-based grading system informs teachers, students, and parents where they meet the tasks and where they need to focus more. Finally, the research question specifically explores whether standards-based report cards convey information regarding student learning more effectively than the current report card to parents, students, and teachers.

Significance of the Study

Grades are how teachers and school leaders communicate students' progress to parents and students. The research shows grades are not strictly the product of student achievement performance. Teachers may include criteria such as effort, attendance, and behavior (Deddeh, Main, & Fulkerson, 2010). On the other hand, school leaders are being held accountable for student performance on standardized achievement tests. This agglomerate can confuse stakeholders and others involved with the school system when students have high report card grades and low achievement test scores. Standards-based grading could be the approach to align report card grades and achievement test scores more closely. Stiggins (2004), said that when information about what is being learned and how to learn it is shared with the student, the grade becomes a powerful tool in enhancing student achievement. This also helps the student to be able to provide feedback to the teacher about what they have learned and what they do not know. More importantly, this allows the teacher to know how to help the student learn the material necessary and be prepared for the next level of learning. Some conflicts with instituting standards-based grading include teacher and parental attitudes and expectations and the creation of new reporting structures (Guskey, 2001). As district leaders search for ways to increase test scores, the two grading systems – standards-based versus traditional – need to be evaluated for effectiveness and accuracy in reporting what students are learning. This information will provide school leaders and teachers with information to make more informed grading decisions to enhance student achievement possibly.

Nature of the Study or Research Design

This study used a quantitative approach. The research design in this study included the collection and assessment of quantitative data from school districts in

Arkansas which changed from the traditional grading method to a standards-based report card. Since using numerical values demonstrates grades and student achievement data, using a quantitative approach to determine if a correlation exists between the two variables is possible. While grades often reflect the use of different scales, the overall purpose of grading systems is to communicate to parents, students, and stakeholders how well the student has performed over the course of the grading or marking period. Grades often reflect language that refers to a student's performance with the criteria established for the class and often employ descriptors like proficient, advanced, partially proficient, and unsatisfactory. Often models that are letter based refer to students meeting expectations, exceeding expectations, or failing to meet expectations. Either model can be used to compare levels of achievement with the ACT Aspire assessment that uses the in need of support, close, ready, or exceeding language to translate numerical values to common achievement language.

Definition of Terms

The following definitions were used to ensure consistency and understanding throughout the study:

Diagnostic Assessments: These are assessments used before the new learning begins (Reeves, 2011).

Formative Assessments: This assessment is used by teachers and students during instruction to provide feedback and adjust ongoing teaching and learning to improve students' achievement of intended instructional outcomes (Popham, 2008).

Free and Reduced Student: This student meets the criteria for free or reduced school breakfast or lunch based on their family need, according to the Child Nutrition Program Arkansas Department of Elementary and Secondary Education (DESE, 2021).

Low SES Household: This household meets the criteria for free or reduced school breakfast or lunch based on family needs, according to the Child Nutrition Program Arkansas Department of Elementary and Secondary Education (DESE, 2021).

High SES Household: This household meets the criteria for free or reduced school breakfast or lunch based on family needs, according to the Child Nutrition Program Arkansas Department of Elementary and Secondary Education (DESE, 2021).

Learning Goals: These are goals used in the day-to-day execution of classroom activities (O'Connor, 2009).

Mastery learning: Students have mastered content when they demonstrate a thorough understanding, as evidenced by doing something substantive with the content beyond merely echoing it (Wormeli, 2006).

ACT Aspire: This is a series of assessments for English language literacy, mathematics, and science in Grades 3-8. These assessments are designed to see if students in Arkansas are meeting the state standards (DESE, 2021). The assessments are vertically scaled, standards-based assessments that monitor student growth and progress toward college and career readiness (ACT, 2021),

Arkansas Department of Elementary and Secondary Education (DESE): This is the governing body for the public educational system in Arkansas (DESE, 2021).

Rubric: This is a coherent set of criteria for students' work that includes descriptions of levels of performance quality on the criteria (Brookhart, 2013).

Special Education (SPED): This is the special education department or classification of students based on federal guidelines with the Individuals Development Education Act (DESE, 2021).

Standards-Based Grading: This grading system references student achievement to specific topics within each subject area. In this system, the student only moves to the next level once they display competence at the current level (Marzano, 2011).

Summative Assessment: This is the information gathered and reported for use in judging the outcome of that development (Marzano, 2011).

Growth Mindset: Describes a way of viewing challenges and setbacks (Morin, n.d.)

Assumptions

As in any study, the researcher based the questions on certain assumptions. This researcher based the study on the following:

1. Students will do their best on the mathematics assessment on the ACT Aspire test.
2. Professional development was adequate to create change and was used with fidelity.
3. Teachers utilized standards-based grading with reasonable fidelity.

The teachers developed a four-tier level of student success. Following are the four tiers: Exceeds, Meets, Progressing, and Inconsistent. The mathematics teachers then developed the essential standards from the state standards at the fifth-grade level to be assessed. In addition, teachers developed rubrics to measure student success for the grade-level standards to be assessed.

Limitations

The following limitations were present in this study:

1. While many school leaders use similar grading systems, each system can be unique because of the individualized nature of teachers grading student work and assigning grades. This uniqueness would also be true of the difference between individual classrooms within a school district.

2. Another limitation was the accuracy of the ACT Aspire assessment for measuring student achievement. This assessment only allows for a one-time measure of a student's performance. This one-time "snap-shot" limits the ability of the ACT Aspire assessment to gauge student achievement if, indeed, the student had a poor day while taking the test, or a myriad of other factors that could play a role in poor performance, took place to inhibit an accurate assessment. Again, using a large population sample should help overcome individual variability.

Summary and Organization of the Remainder of the Study

This study is organized into five chapters. Chapter 1 contains an introduction to the study, context, and history of the issue, a statement of the problem, the significance of the study, a definition of terms, and limitations and delimitations. Chapter 2 includes a review of the literature that is organized according to the topic. Chapter 3 includes the research methodology, research questions, research design, and population of the study. Chapter 4 provides the results of the study, while Chapter 5 includes a summary of the findings, conclusions, and recommendations for future research and practice.

Chapter II: Literature Review

Introduction

The purpose of education has been debated and this debate continues today. Politicians, educators, researchers, and the public cannot seem to find common ground in public education. According to Kelly (2019), imbuing students with the knowledge they need to be functional adults in their day-to-day lives is the purpose most see in educating students. Educational reforms come and go as new elections begin and end. One topic with which most agree is the necessity of doing what is best for students. Politicians see the need for a systematic approach to standards and curricula reform throughout the United States. The movement to a focus on curriculum standards in public education can be traced to A Nation at Risk (1983) published by the United States Department of Education in 1983. This publication caused a dramatic shift in education reform by implying that the quality of the American education system was mediocre (Marzano, 1998). By 1998, standards were developed for most of the content areas in public schools. Researchers have argued that the educational system created by the federalist system was fragmented and did not create quality teacher education programs or opportunities for professional development. This was especially evident in the lack of a relationship between information that was assessed on standardized tests and the curriculum being taught in the classroom (Polikoff, 2015).

School leaders and teachers began to feel increased accountability as a result of the No Child Left Behind Act of 2001 and the Race to the Top Initiative of 2009 (United States Department of Education, 2009). These laws caused school leaders to examine pedagogical practices within the classrooms. No longer are teachers teaching to the test

but teaching to state standards aligned with the weighted categories released by the testing centers has become the priority. With all the changes in state and federal laws, school leaders have shown an increase in weighing more emphasis on student mastery of state standards (Brookhart, 2017).

Teachers' decisions on grades and student achievement can have a long-lasting impact on students. "Grading is the process of summing up student achievement with marks or symbols. Teachers grade individual assessments and assignments, and they also combine those grades into one summary mark or symbol for a report period" (Brookhart, 2012, p. 257). There are many different purposes for grading. Researchers have categorized teacher and school leader responses into six major categories (Airasian, 2001; Feldmesser, 1971; Frisbie & Waltman, 1992; Guskey & Bailey, 2001; Linn, 1983).

1. Communicate information about student achievement to parents and others.
2. Provide information and feedback to students for self-evaluation.
3. Select, identify, or group students for specific educational paths and programs. This could include higher-level AP and honors classes or possibly remedial and special education classes.
4. Provide incentives and motivation for students to learn.
5. Evaluate the effectiveness of instructional programs. Are students learning and is the content valuable?
6. Provide evidence of students' lack of effort or inappropriate responsibility.

Educators generally agree each of the six purposes is important and legitimate, but they seldom agree on which one is most important. Educators grading approach is

different, however, the goal is the same to transform grades to more accurate reflections of student learning (Streifer & Palmer, 2020).

History of Grading

Grading is an integral part of a student's learning experience (Schneider & Hutt, 2014). Understanding the history of grading systems to find the most effective grading system for the public school system is important. Scholars from Oxford and Cambridge brought their educational traditions to the United States between 1630 and 1641 (Kunnath, 2016). The ancient Greeks used assessments as formative and not evaluative learning tools. Harvard required exit exams in 1646 to attain a degree. In 1785, Yale president Ezra Stiles implemented the first grading scale in the United States based on four descriptions: Optimi, Second Optimi, Inferiores, and Periores (Lee, 2020). In the early 1900s, grading became more standardized and scale-based as education expanded and grades became widespread. By the 1940s, the dominant grading system was the A-F grading system. Along with this came the 4.0 scale and the 100 percent system (Schneider & Hutt, 2014).

Norm-referenced grading, which compares students to one another and uses class standing to assign grades, was advocated in the early 1900s (Brookhart et al., 2016). This type of grading was based on the normal distribution, or Gauss's curve, and is also known as the bell shape curve (Finkelstein, 1913). Beliefs that conforming grades to the curve would increase grading consistency in the classrooms (Meyer, 1908). The 100 percent grading system was the most common in high schools from 1890-1910 (Brookhart, 2009). During the early 1900s, the inaccuracies of the percent grading system defined by many scholars brought on the adoption of the letter grading system in the

1920s (Brookhart, 2009; Starch & Elliott, 1913). During the 1930s, standards or absolute standards grading was adopted. The most common grading system currently being used is percent grading which is used as a way to arrive at letter grades (Brookhart, 2009).

Purpose of Grading

Grading and reporting of grades have been part of the educational system in the United States since the mid-1800s. They play a key role in schools across America. Munoz and Guskey (2015) discussed this when they said, "Grading and reporting are foundational elements in nearly every educational system. Grading represents teachers' evaluations both formative and summative of students' performance. Reporting is how the results of those evaluations are communicated to students, parents, or others" (p. 64).

The purpose of grading and reporting grades has been defined in several ways. Munoz and Guskey (2015) state, "The purpose of grading is to describe how well students have achieved the learning objectives or goals established for a class or course of study" (p. 65). Rubrics used to design assessments covering standards lead to increased student performance by clearly showing/telling the student how their work will be evaluated and what is expected to determine if students have learned the goal or academic learning target (Mueller, 2018).

Heflebower, Hoegh, Warrick, Hoback, McInteer, Clemen and Marzano (2014) report, "Grading systems should ensure that students' grades are valid, fair, and consistent and clearly reflect what students know and are able to do" (p. 2). Tomlinson and Moon (2013) state that the purpose of grades "should support learning" (p. 128). The foundation of education is learning, and learning is shown by the use of grades. Marzano (2000) said, "Academic achievement should be the primary factor in grades" (p. 40).

Grades are also a means of communication from the teachers to the students and parents, can be used to sort students in groups, and work as a motivational tool that encourages students to learn (Boston, 2003). Tomlinson and Moon (2013) report, “A grade should communicate as clearly as possible to students, parents, and other stakeholders what a student knows, understands, and can do at a given point in time” (p. 132).

As pointed out earlier, grading is used for feedback to students and parents. Shute (2008), defined feedback as “information communicated to the learner that is intended to modify his or her thinking or behavior for the purpose of improving learning” (p. 154). Providing feedback to students and parents may be the most important function of grading (Marzano, 2000). Reeves (2008) added to this by saying, “The most effective grading practices provide accurate, specific, and timely feedback designed to improve student performance. In the best classrooms, grades are only one of the many types of feedback provided to students” (p. 85). Glasgow and Hicks (2003) reported, “Feedback is important for students in several ways: it helps them assess their mastery of course material, helps them assess their use of thinking and learning strategies, and helps them connect their efforts and strategies to their academic outcomes” (p. 95). Research has shown that systemic feedback can improve student performance by approximately 33% (Marzano, 2000). This is especially the case when feedback is linked to specific learning objectives, which gives students direction and a clear understanding of what they have and have not mastered (Beatty, 2013).

Use of Zeroes

When a student is absent for a day or forgets to turn in an assignment, the teacher will often assign a grade of zero for that assignment. Reeves (2010) indicated on a four-

point scale, where A = 4, B = 3, and so on, zero is accurate because the difference between the A, B, C, D, and F are all equal - one point. However, assigning a zero on a 100-point scale is a math error; it implies a 60-point difference between the D and F, while the other differences are typically about 10 points. “It makes missing a single assignment the ‘academic death penalty.’ It is not just unfair - it is not mathematically accurate” (Reeves, p.78). To improve the accuracy of grades when using a 100-point scale, instead of giving scores of zeroes for work that is not turned in, educators should consider using incompletes and interventions to motivate the students to complete the work. Giving a student a zero will skew the grade and make it inaccurate (Dueck, 2014). When a zero is factored into a grade it factors into a numerical value assigned to an activity that cannot be assessed. This zero averaged with other grades gives a false sense of the student’s academic performance. Conversely, grades inflated with nonacademic factors such as classroom behavior and extra credit can weaken the link between what is reported by the classroom teacher to parents and students and academic performance on standardized assessments (Welse, 2013).

Assessment researcher Marzano (2006) has proposed alternative scales that are more balanced and fairer. Marzano uses a 0- to 4-point scale, ensuring the scale is balanced. A score of zero shows a student has shown no understanding or demonstration of the learning standard even when provided help. A 1 on the Marzano scale shows that, with help, a student demonstrates partial understanding of the simplest details and processes of the learning standard. A score of 2 shows no major errors or omissions on the simpler details but major errors on the more advanced details and processes. A 3 shows no major errors regarding the learning standard. Last, a 4 shows a student has

earned a 3 but can also provide in-depth inferences and apply what was taught. Although zero is used in the Marzano scale, it does not skew the grade in a way that prevents the student from recovering.

Grade Inflation/Classroom Management

According to researchers, teachers tend to increase or decrease student's academic grade due to nonacademic behaviors such as effort, homework completion, attendance, punctuality, class participation, academic integrity, or extra credit (Fleenor, Lamb, Anton, Stinson, & Donen, 2011; O'Connor & Wormelli, 2011; Vatterott, 2015). By separating academic and behavior factors, proponents of this practice assert that reports become more valid, providing information that parents, students, and teachers can use to improve learning and behavior (Guskey & Jung, 2013; O'Connor, 2011; Schimmer, 2016; Vattorott, 2015). By doing this, studies show improving reliability of grades, since teachers using traditional grading practices vary considerably in which behaviors they include in a grade and to what extent (Brookhart et al., 2016; Guskey 2013; Guskey et al., 2011). Welsh and D'Agonstino (2009) discovered that teachers believe including behaviors such as effort motivates students to try harder. Beatty (2013) corroborated this finding with student participants who admitted they struggled to motivate themselves to do work that was not graded, or desired to receive some type of credit or reward for their hard work. Additionally, Guskey (2009) found that teachers use grades as a "vital component of classroom management and control" (p. 11). Zoekler (2007) determined teachers' grading practices are influenced by their ideas of acceptable character and behavior. Vatterott (2015) expands the idea of teachers drawing on their morals while grading, suggesting, "We've used academic grades for more than academics because we

believe our job is more than academics – our goals have always included shaping children into better people" (p. 12).

Grading Philosophies

McMillan and Nash (2000) conducted a study in Virginia on classroom teacher assessment and grading decision-making. They found common themes of teacher beliefs and values, classroom realities, external factors, rationale for decision-making, assessment practices, and grading practices. Even though the teachers did not state a philosophy on teaching, their explanations were indicative of a philosophy such as, "I weigh more on homework. The more you practice something, the more proficient you become in that skill" (p. 9).

According to Close (2009), fair grading is modeled on two fundamental principles. The first principle is that grading should not be partial, but consistent. The second principle is that a fair grade should be based on the student's competence in the curriculum they are learning. Common grades are A, B, C, D, and F. Sometimes these grades are followed by either a plus (+) or a minus (-). He goes on to make a distinction between grading and evaluating and that these two terms are separate. He assumes that a rough definition of "fair grade is that at a minimum, a summary mark of an accurate, expert evaluation of student academic work that (1) is normally made by the instructor, (2) appraises a student's knowledge and/or skills in the subject matter, and (3) is permanently recorded in the student's record (Close, 2009). He looks at three models of grading. The first model holds that grades should function as a means of rewarding and punishing students across a range of academic and institutional values. The second model is that grades are the goal in the classroom. The last model is that grading is an

information process concerning mastery of the content. This model focuses exclusively on the informational nature of grades.

Close (2009) also believes there are principles to be examined while assigning grades. First and foremost, grading should be impartial and consistent. There should be no differences made between students and the work. Grade components should have determinate weights expressible as a fraction of the final grade. Every student should be aware of the weights of the grades at the beginning of the year. He also believes that every set of scores should not be considered with a certain of A's, B's, C's, D's, and F's. This methodology is peer ranking and that there is a normal curve to the grades. This type of grading is unfair because of the grade distributions are not guaranteed to be normal +even when the sample size may be large. Sequencing of grades should all have equal weights to determine ending grade in a course. No one is exempt from a grade. This can be looked at as being impartial to a student. In conclusion, his belief is that fair and ethical grading is based on two fundamental principles. The first principle is that a fair grade is impartial. The second principle is that a fair grade is an expert evaluation of a student's mastery of course content (Close, 2009).

McMillan, Myran, and Workman (2002) conducted a study of over 900 3-5 grade teachers representing urban, suburban, and rural schools to determine the factors elementary teachers use in grading. "Many teachers seemed to have individual assessment policies that reflected their own individualistic values and beliefs about teaching" (p. 160). They found a variation of grading practices within schools was greater than the variation between schools. For example, they found student improvement was not used at all by 13% of the teachers, and 30% of the teacher participants used it either

"extensively" or "completely". Although the participant districts used percentage scale corresponding to letter grades, only 65% of the teachers reported using the scales "extensively" to "completely" (p. 208).

Use of Standards-Based Grading

The report, *A Nation at Risk,* which was published in 1983, is often given credit for initiating the modern standards movement (Marzano & Kendall, 1996; Shepard, 1993). The continued concerns about the nation's education system prompted the Education Summit of 1989.

Diane Ravitch, former Assistant Secretary of Education, is commonly recognized as one of the main designers of the modern standards movement (Marzano & Kendall, 1996). Ravitch (1995) set forth that standards would improve the effectiveness of American education in her book, *National Standards in American Education: A Citizen's Guide*. She sets forth the idea that all industries have standards because standards improve the quality of life (Ravitch, 1995).

The education system in the United States has seen numerous changes such as the federal government's increase in control on public schools and the increased pressure the teachers face to help all students meet federal guidelines (Marzano & Kendall, 1996). The standards movement delivered a revitalized emphasis on ensuring all students understand the essential information determined by the state (Wheelock, 1995). Researchers have determined that even though curriculum guides are in place, these guides frequently do not transfer into classroom procedure as teachers determine what should be emphasized, added, or even excluded from lessons in the best interest of students (Marzano, 1998). As a result, students receive different learning experiences in

different classrooms. The standards movement reduces the concern of students only being partially exposed to state standards since assessments are created based on the prescribed standards.

Standards-based grading is focused on the belief that educators provide students with specific, clearly defined goals, determine levels of mastery, and communicate with parents and students when those goals are achieved (Spencer, 2012; Guskey, 2011). According to surveys conducted by Guskey (2011) teachers agreed that while standards-based reporting took more time to create, it also gave a clearer understanding of learning to families. Parents also agreed that the information on standards-based report cards gave more meaningful information. Parents reported that knowing how their children were performing relative to required curriculum standards helped them better monitor student progress and offer more direct assistance to their children because the standards-based report card offers more detailed skills-based information than traditional grading (Guskey, 2011).

In Table 2.1 and Table 2.2, Scriffiny (2008) illustrates how a student's grade may not effectively demonstrate how well a student is performing academically in class. Given the same students with traditional grading and then comparing the numeric score to actual objective and the ability to meet these objectives, the standards-based report card provides more information and can be used to guide instruction and give parents a better understanding of their child's needs (Scriffiny, 2008).

Table 2.1

Comparing Traditional and Standards-Based Grading: Traditional Grading

Name	Homework Average	Quiz 1	Chapter 1 Test
John	90	65	70
Bill	50	75	78
Susan	110	50	62
Felicia	10	90	85

Amanda	95	100	90

Note. Adapted from "Seven Reasons for Standards-Based Grading," by P. L. Scriffiny, 2008, *Educational Leadership*, 66(2), pp. 70–74.

Table 2.2

Comparing Traditional and Standards-Based Grading: Standards-Based Grading

Name	Objective 1: Write am alternate ending for a story	Objective 2: Identify the elements of a story	Objective 3: Compare and contrast two stories
John	Partially Proficient	Proficient	Partially Proficient
Bill	Proficient	Proficient	Partially Proficient
Susan	Partially Proficient	Partially Proficient	Partially Proficient
Felicia	Advanced	Proficient	Proficient
Amanda	Partially Proficient	Advanced	Proficient

Note. Adapted from "Seven Reasons for Standards-Based Grading," by P. L. Scriffiny, 2008, *Educational Leadership*, 66(2), pp. 70–74.

The two tables above show how the same students are graded first using a traditional grading system and then comparing the numeric score to actual objective and the ability to meet these objectives. The standards-based report card provides more information and can be used to guide instruction and give parents a better understanding of their child's needs (Scriffiny, 2008).

The purpose of standards-based grading is to show how well students have mastered the content standards. These grades are based on what content the students have mastered as opposed to what grades the students have earned (Brookhart, 2011). Instead of a cumulative grade as in traditional grading, standards-based grading assesses final submissions of each standard to determine proficiency. SBG focuses on standards, isolating academic evidence, and allowing multiple opportunities to demonstrate proficiency. Each of these key components is designed to increase the validity, reliability, and equity of grades to ultimately improve student learning. In traditional grading, if a student does well at the beginning of a course and then is not as successful towards the end, the final average grade does not show a continuum of academic growth. When

providing students with a rating in standards-based grading, a focus on authentic assessments connected to specific learning objectives is utilized (O'Connor, 2009). Assessments combine product, progress, and progress into that final score and each criteria is reported separately as they each serve their own purpose (Guskey, 2015; Marzano, 2011; O'Connor, 2009).f\

With standards-based grading as students learn more in-depth content, the final assessment or project submission shows a current achievement level for a particular standard (Rundquist, 2012). This method shows the progress the students have made through their learning experiences. By reporting standard proficiency, each subject area is specified by learning targets which focus on exact content elements. The standard-based report card gives a scale 1-4. A score of 1 shows the student has little understanding, a 2 shows partial understanding, a 3 shows the student meets expectations, and 4 shows the student is advanced. Attitudes, efforts, and work habits are included in a separate area of the report card unlike traditional grading (Guskey & Jung, 2006).

Grading Practices Need Change

Grading is one of the teachers' greatest challenges and most important professional responsibilities. The traditional grading practice tends to lump content with effort and behavior into one letter grade (Brookhart, 2011). All too often, traditional grading practices in the United States are based on instructional and motivational principles that cause many students to give up in hopelessness and accept failure rather than driving them toward academic success (Stiggins, 2014). Wormeli (2018) explains that allowing educators to decide how to assign grades to student learning may not be an accurate picture of what students know, understand, and be able to do. Teachers' beliefs

in what to grade and to record are significant influences on what a student earns as a grade (Brookhart, 2017; Guskey, 2015; O'Connor 2011). O'Connor (2011) stated, "Teachers develop assessments based on their professional judgment of what is to be assessed and how–a subjective process" (p. 11). "Grading the work of others is a subjective experience even under the best of circumstances. As instructors, we try to be fair, unbiased, and objective, but the basic element of our humanity prevents our attaining a truly objective state" (Jae & Cowling, 2009, p.54).

Grading practices must change to "meet the learning needs and desires of future generations of young adolescents, the core middle school practices must continue to grow and thrive" (Schaefer, Malu, & Yoon, 2016, p. 18). To make grades more meaningful, issues related to both purpose and reporting format must change as well (Guskey, 2015). Grading experts (Guskey, 2015; Marzano, 2000; O'Connor, 2009; Reeves, 2008) agree teachers should update their grading practices to better align with the realities of how and what students are learning in schools.

Standards-based grading offers a clearer understanding of student performance on standardized assessments than traditional grading because of the emphasis of measuring student achievement against established curriculum learning targets that align to standardized assessments. The purpose of standards-based grading is to provide more consistency with grading and provide more reliable information to students and parents on academic achievement (Paeplow, 2011). In 2006, Guskey concluded that:

> At all levels of education, therefore, educators must strive to ensure that the procedures they use in assigning grades or marks to students' work are explicit, clear, and as objective as possible. They must work hard to guarantee that their personal opinions and unconscious biases do not influence their grading practices. Above all, teachers and professors must base their grading policies and

> practices on criteria that will be judged by all to be just, equitable, and unprejudiced. (p.13)

Teachers using a traditional grading system and being subjective, a student's grades may include work efforts, group work, and classroom behavior (Scriffiny, 2008). While some students are good at being students, they may not have learned the content. These students are the ones who turn in assignments, complete assignments for extra points, participate in class discussions but do not receive a proficient score on the ACT Aspire mathematics exam at the end of the year or on common assessments. These types of inconsistencies can be detrimental to the student as they move from grade to grade or state to state with the movement of Common Core and Essential Standards (Scriffiny, 2008).

Communicating Information Effectively

According to Guskey and Jung (2006), the most challenging aspect of reporting standards is communicating the information effectively to students and parents. "Letter grades make sense to parents too because that's what they received when they were in school" (Guskey et al, 2006, p. 16). When a district is developing standards-based grading and reporting, parents should be included in the process in order to gain buy in with all stakeholders (Cox, 2011). Standards-based grading and reporting more closely resembles job evaluations in the professional world. Adults are evaluated based on performance (Scriffiny, 2008). By receiving reports more closely related to real-world evaluations, students are better prepared for future employment.

Students with disabilities, gifted students, or English language learners can have difficulty with content standards at varying levels and teachers may struggle with assigning fair grades (Guskey, 2010). Teachers are at times compelled to assign extra

points based on effort or weighing assignments differently in order to compensate for a student's learning style or learning deficit. Determining whether or not the standards applied to diverse learners are appropriate for the learner can be a barrier for teachers implementing standards-based grading and reporting independently. Assessing a student who is performing below grade level on curriculum standards for the grade level because of learning difficulties would not provide accurate data on academic progress (Guskey, 2010). Students who need modified work or alternative assessments can still be assessed on how well they learned the content objective that will reflect in the reporting. This grading allows teachers to track exactly what students know, as well as their lack of understanding, in order to provide better instruction tailored to meet their needs (Spencer, 2012).

Graded Achievement Relationship Between Tested Achievement

An analysis of studies by Brookhart (2014) between graded achievement and tested achievement found correlations in traditional grading models ranging from .40-.70, with the variance attributed to different grading practices in the learning environment. Research on graded achievement in a standards-based grading environment, however, has been limited.

Studies of standards-based grading and standardized test performance at the elementary level is more abundant than studies at the high school level; however, studies are still generally small in number, and results vary across content areas (Craig, 2011; Hardegree, 2012; Norton, 2014; Prejean-Harris, 2013; Welsh, D'Agostino, & Kaniskan, 2013). Key studies by Craig (2011), Hardegree (2012), Prejean-Harris (2013), Norton

(2014), and Welsh et al. (2013) all found varying results comparing standards-based grading, traditional grading, and performance on standardized tests.

Prejean-Harris (2013) studied science and math state assessment scores of 823 elementary schools, comparing standards-based grading to traditional grading systems. While there was not statistical significance found in the study, scores were higher in both math and science on average in traditional grading schools (Prejean-Harris, 2013). Hardegree (2012) researched 550 Georgia fifth-grade student report cards in eight elementary standards-based schools. Her study concluded that standards-based report card grades correlated with the Georgia state assessment in math and reading, with stronger correlations in math (Hardegree, 2012). Welsh et al. (2013) examined the relationship between 125 third and fifth-grade students' grades to state test results. "A moderate degree of grade–test score convergence was observed using three agreement estimates (coefficient kappa, tau-b correlations, and classroom-level mean differences between grades and test scores)" (Welsh et al., 2013, p. 26). Additionally, there was little variation among teachers and convergence rates of students meeting the state standards indicating that standards-based practices help improve consistency in grading practices (Welsh et al., 2013). Finally, Norton (2014) studied six schools comparing fourth and fifth-grade student performance on the Kentucky K-PREP state assessment in standards-based grading environments compared to traditional grading environments. Kentucky leaders were early adopters of standards-based grading after receiving race to the top money from President Obama's 2009 initiative to challenge schools to improve standards and assessment practices (Guskey, Swan, & Jung, 2011). In this study, students scored

significantly higher on the K-PREP assessment in mathematics in standards-based systems (Norton, 2014).

Theoretical/Conceptual Framework

This quantitative study will explore the differences between results of the ACT Aspire test in the state of Arkansas between schools that use a traditional grading system compared with schools that use a standards-based grading system for 5th grade math. This study will include a review of public documents reporting the school's demographics, standardized test scores, and Arkansas ESSA School Report Card rating,

The theory of Zone of Proximal Development (ZPD) was developed in the early 20th century, and the main assertion of the theory is that cognitive development in early childhood is advanced through social interaction with other people, particularly those who are more skilled (McLeod, 2020). Vygotsky proposed that learning comes before cognitive development in children and that children construct knowledge actively (McLeod, 2020). The theory states that children who are in the zone of proximal development can almost perform the given task, but not without the help of others. Getting instructions from another person is important to cognitive development in early childhood (McLeod, 2020). Children are in the ZPD when a challenging task is presented that can be accomplished with guidance or encouragement. This level represents a stage where children have mastered some functions, while still fostering others. The third level is the level of potential development. This level is representative of tasks that children cannot do.

This study's conceptual framework is based on the theory that providing achievable targets (standards), removing designations of failure by utilizing a proficiency

scale through rubrics and SBG, which inform students of their development along a continuum of learning promotes motivation and self-efficacy (Stiggins, 2014) Dweck (2016) supported Vygotsky's theory when she stated that encouraging feedback can improve student outcomes. She explained that a growth mindset can develop student motivation, which in turn can boost student achievement. Ricci (2015) further detailed Dweck's growth mindset theory by stating that "educators with a growth mindset believe that all students can achieve at higher levels-with effort, perseverance, and resiliency. Learners with a growth mindset believe that they can grow their intelligence with hard work" (p. 1). Ricci also identified two main components of the growth mindset culture that supported the ZPD theory. One was that all students should have the opportunity to experience advanced learning opportunities, and the other emphasized how feedback and praise are essential. Vygotsky (1978) stated that the ZPD is where students are able to realize their potential. His theory provided the space in which students could advance their understanding with guidance from the teacher.

Teachers using SBG remove indications of failure and instead provide feedback on a continuum of learning. Consequently, students understand what they need to do to reach mastery. Creating opportunities for students to reach mastery is important to encouraging self-efficacy. Self-efficacy is a predictor of initiative, persistence, and achievement, and increases when students feel they are able to acquire a skill (Gist & Mitchell, 1992; Martochhio, 1994). Student self-efficacy is influenced by report card grades. In order for students to gain confidence, strengthen self-efficacy, and perceive themselves as capable, they need grading systems that provide meaningful feedback and

opportunities for improvement. The conceptual framework for this study is illustrated in Figure 3.

Figure 3. The Complete Beginner's Guide to Mastery Learning and Standards-Based Grading. Retrieved from The-Ultimate-Guide-to-SBG.pdf (otus.com)

"As children are given instructions or shown how to perform certain tasks, they organize the new information received in existing mental schemas. They use this information as guides on how to perform these tasks and eventually learn to perform them independently" (McLeod, 2020). Children learn through social interaction that includes collaborative and cooperative dialogue with someone more skilled in the tasks they are trying to learn (McLeod, 2020). As a child becomes more familiar with the task, less guidance can be provided, and the child should be able to perform the task by themselves.

Student academic growth and development of independent skills was mentioned by multiple researchers. Hardegree (2012) stated that teachers must understand student mastery levels in order to provide students with appropriate challenges that will foster growth. Guskey (2011) states, "when learning is assessed using a well-defined set of credible learning standards that include graduated levels of performance, then progress and growth criteria can be considered synonymous" (p. 41). Furthermore, O'Connor

(2009) described how teachers often "aim to develop students to be self-directed, independent, lifelong learners" (p.5). Schimmer (2016) identified SBG as a possible solution to these instructional goals. He analyzed that SBG helps teachers identify individual student mastery which can inform challenging instruction.

A major contribution of Vygotsky's theory of cognitive development in children is the acknowledgment of the social component in both cognitive and psychosocial development (McLeod, 2020) is especially helpful in forming the research questions for this study. Teachers giving students instructions on how to perform tasks and how the students learn these tasks and how well they perform the tasks is a direct link to Vygotsky's theory of how a child learns. ZPD is a good fit for this study because the subjects of the study are teachers, students, and parents who depend on a standards-based grading system to inform them of how well the student is performing and what tasks the students need to improve in. ZPD informs this research because this study looks at how well a standards-based grading system informs teachers, students, and parents where they meet the tasks and where they need to focus more and whether or not the student scores better on achievement tests than when a traditional type grading system is used within the classroom. The research question specifically explores whether standards-based report cards convey information regarding student learning more effectively than the current report card to parents, students, and teachers and if the standards-based type grading specifically conveys better scores on achievement tests than those schools who use traditional grading.

Chapter III: Methodology

Introduction

The goal of the No Child Left Behind (NCLB) law was to have every child make the grade on state-defined education standards by the end of the 2013-2014 school year (United States Department of Education, 2020). The law required testing for each student throughout elementary and high school, and students had to achieve specified performance levels each year to meet the requirements set forth by NCLB (2011). In 2011, the "Obama Administration invited states to apply for waivers of key requirements of the Every Student Succeeds Act (ESSA) as amended by NCLB" (Kober & Riddle, 2012, p. 2). The purpose of this study is to determine whether the type of grading system (traditional or standards-based) used to assess student progress makes a difference on ACT Aspire results. The significance of studying standards-based grading was to provide educators with the research necessary to make informed decisions when considering the transition to standards-based grading. Included in Chapter Three are the problem and purpose, research design, population and sample, the instrumentation used, and how data were collected and analyzed.

Research Questions/Hypotheses

RQ1: Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems?

RQ2: Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of socioeconomic status?

RQ3: Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of ethnicity?

RQ4: Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of gender?

Research Methodology

The enigma of understanding drove this study if standards-based grading impacted the scores of the ACT Aspire assessment as compared to traditional report card grading. Standardized assessments are seen, by most researchers, as an effective means to determine a student's understanding of concepts related to state standards. This study will provide quantitative data that reflects the relationship between grading systems and grades and the results produced through a standardized assessment. One aspect of the research focused on the difference between standards-based grading systems and traditional letter-based systems relative to standardized assessment. The results from this study should inform districts of the possible benefits of one grading system over the other.

Research Design

This study will use a quantitative research approach to determine the relationship between standards-based grading and the mathematics ACT Aspire test. This research design was used to collect and analyze the data because the data collected were in the form of numbers and statistics rather than individual perceptions (McCusker & Gunaydin, 2015). The hypothesis was that the mean mathematics ACT Aspire

achievement of the fifth graders taught utilizing standards-based grading is significantly different than the mean mathematics ACT Aspire achievement of the fifth graders being taught without standards-based grading or using the traditional report card. The specific research questions to be answered are:

1. Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems?
2. Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of socioeconomic status?
3. Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of ethnicity?
4. Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of gender?

The research will compare fifth-grade students' mathematics ACT Aspire scores in schools that use a traditional grading system transitioning to a standards-based one. According to Gay (1996), a relationship is conducted to gain insight into the factors or variables related to complex variables such as academic achievement. If two variables are correlated, this can predict from one variable to the other with a certain degree of accuracy (Jackson, 2009). For example, are mathematics ACT Aspire assessment scores and standards-based grading correlated?

The ACT Aspire scale score determines the student's achievement level. There are four levels: exceeding, ready, close, or in need of support. These levels are based on the student's scores based on their skill scores. The score report includes the percent and number of points earned out of the total number of points possible in each skill. Student performance in each skill is also compared to an ACT Readiness Range, which indicates whether the student appears to be on target or may need help. Test scores are estimates of a student's educational development. ACT Aspire scale scores begin with a possible 400 points and increase depending on the subject and grade. For fifth-grade math, the maximum possible score is 446. The benchmark score for fifth-grade math is 418 (ACT Aspire, 2020), which means the student is at the ready or exceeding level.

Deleted: ¶

Population and Sample Selection

The purpose of this study is to determine whether the type of grading system (traditional or standards-based) used to assess student progress has an impact on ACT Aspire results. The population for this study will consist of fifth-grade students in Arkansas schools. An email will be sent to principals in all Arkansas schools with a fifth grade explaining the purpose of the study and directing them to an electronic survey to determine if they use standards-based grading or traditional grading for fifth-grade math. They will also be asked whether they would be willing for their anonymized ACT Aspire data to be used in the study. From the population, the sample will consist of schools using SBG and a comparable number of schools with similar demographic characteristics using the traditional grading method.

Instrumentation

The purpose of this study is to determine whether the type of grading system (traditional or standards-based) used to assess student progress has an impact on ACT Aspire results. Subsequently, archived ACT Aspire fifth-grade math data will be used in the study. No personally identifiable information will be used.

Operational Definitions of Variables

Type of grading system = independent variable (IV)

Percent meeting Benchmark = dependent variable (DV)

Data Collection

An email will be sent to superintendents in all Arkansas schools with a fifth grade explaining the purpose of the study and directing them to an electronic survey to determine if they use standards-based grading or traditional grading for fifth-grade math. They will also be asked whether they would be willing for their anonymized ACT Aspire data to be used in the study.

A list of schools that are willing to participate in the study will be compiled. For each participating SBG school, a traditional-grading school with similar demographics will be selected. The schools will be asked to submit their anonymized fifth-grade math ACT Aspire raw scores for the specified year in a format that can be uploaded to Excel. No names or personally identifiable information will be requested. The archived data is available from the state's password-protected eSchool database.

Data Analysis

When the raw data is received from the participating schools, it will be entered into an Excel spreadsheet for coding. The student's gender, ethnicity, and socioeconomic status will be coded. The socioeconomic status will be determined based on eligibility for

free or reduced-priced meals. The math ACT Aspire score will be entered as raw data. Schools will be coded as a number with no name attached. Once the data has been properly coded, it will be uploaded into SPSS for statistical analysis.

Research question 1 asks if there is a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems. To answer this question, an independent samples t-test will be conducted comparing the ACT Aspire scores of schools using SBG with those using traditional grading.

Research question 2 asks if there is a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of socioeconomic status. To answer question 2, student scores from low SES households will be selected. An independent samples t-test will be conducted comparing the ACT Aspire scores of schools using SBG with those using traditional grading. Student scores from high SES households will be selected. An independent samples t-test will be conducted comparing the ACT Aspire scores of schools using SBG with those using traditional grading.

Research question 3 asks if there is a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of ethnicity. To answer question 3, an independent samples t-test will be conducted by comparing the ACT Aspire scores of schools using SBG with those using traditional grading.

Research question 4 asks if there is a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading

systems considering the factor of gender. To answer this question, an independent samples t-test will be conducted comparing the ACT Aspire scores of schools using SBG with those using traditional grading.

Internal and External Validity

This study will determine whether the type of grading system (traditional or standards-based) used to assess student progress impacts ACT Aspire results. Archived ACT Aspire fifth-grade math data will be used in the study. Consequently, validity is not an issue.

Ethical Assurances

Before data collection, the study will receive approval from Arkansas Tech University's Institutional Review Board (IRB). Permission to use the school's anonymized ACT Aspire 5^{th}-grade math data will be obtained formally from the school's superintendent. No personally identifiable information will be collected. Schools will be asked to remove student names, identification numbers, and any other information that might be used to identify the student before providing the researcher with the math score data. Individual schools participating in the study will not be identified in the study.

The researcher will receive data electronically by email and downloaded to the researcher's password-protected OneDrive storage medium. Once the data has been downloaded, the email containing the file will be deleted. All data and analysis results will be saved to the OneDrive medium. The data will remain on the OneDrive medium for one year after the acceptance of the dissertation by the committee.

Summary

In this chapter, the researcher presented information regarding the research design and methodology and discussed the population and sample to be studied. The study will examine whether there is a significant difference in 5th-grade ACT Aspire math scores between schools that use standards-based grading and traditional grading. Anonymized data will be obtained from schools and analyzed using a Chi-Square test. Results of the analysis should inform schools whether one method of grading might result in better scores on standardized tests indicating that more learning has occurred. In chapters four and five, the analysis results will be presented, followed by a discussion of the findings.

Chapter IV: Results

Introduction

The purpose of this study is to determine whether the type of grading system (traditional or standards-based) used to assess student progress makes a difference on ACT Aspire results. School leaders can support policies and practices concerning classroom grading by determining the relationship between grading system and achievement. This research may bring validity to district leaders using standards-based grading systems.

The study's design was quantitative and nonexperimental, featuring an associative/predictive research approach and methodology. Four distinct research questions with accompanying hypotheses were formally posed to address the study's topic and research problem. Descriptive, inferential, and associative/predictive statistical methods were utilized to address the study's research questions and hypotheses. Study data were initially taken from archived data from Arkansas Department of Education Data Center ((https://adedata.arkansas.gov/) and coded within an Excel Spreadsheet. Data analysis, interpretation of findings, and reporting of findings were subsequently conducted using the 29th version of IBM's *Statistical Package for the Social Sciences* (SPSS).

Results from this study can be used to inform the decision-making process at the researcher's own school within the district as leaders transition from traditional grading to standards-based grading per the Professional Learning Community (PLC) at the elementary level. This chapter contains a report of the study's findings, with specific focus on the four research questions and accompanying hypotheses.

1. Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems?
2. Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of socioeconomic status?
3. Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of ethnicity?
4. Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of gender?

Description of Sample

Six Arkansas schools that transitioned from using a traditional report card to a standards-based report card were selected for analysis. Six additional Arkansas schools similar in size, demographics, and ethnicity that used a traditional report card were selected for comparison. The researcher obtained permission from the superintendents of the school districts to retrieve fifth-grade archived ACT Aspire math scores from the 2020-2021 school year prior to data collection. The data was retrieved from archived data housed in the Arkansas Department of Education Data Center (https://adedata.arkansas.gov/) and coded within an Excel Spreadsheet. Data analyses were subsequently conducted using IBM's Statistical Package for the Social Sciences (SPSS).

Each district's descriptive data is presented to provide the reader with a more complete understanding of the student make-up for each district. Each district was anonymized and labeled with a letter of the alphabet from A-L. The purpose of this study is to determine whether the type of grading system (traditional or standards-based) used to assess student progress makes a difference on ACT Aspire results.

Data Analysis

The following research questions guided this study:

RQ1: Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems?

$H_0$1: There is not a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems.

H_a1: There is a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems.

RQ2: Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of socioeconomic status?

$H_0$2: There is no significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of socioeconomic status.

H_a2: There is a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of socioeconomic status.

RQ3: Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of ethnicity?

$H_0$3: There is no significant difference in fifth-grade math ACT Aspire scores that use traditional versus standards-based grading systems considering the factor of ethnicity.

H_a3: There is a significant difference in fifth-grade math ACT Aspire scores that use traditional versus standards-based grading systems considering the factor of ethnicity.

RQ4: Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of gender?

$H_0$4: There is no significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering gender.

H_a4: There is a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering gender.

Descriptive Data

Twelve Arkansas school districts agreed to participate in this research. All anonymized data was collected from the Arkansas Department of Education. Of these twelve school districts, six used the traditional grading and six used the standards-based grading. ACT Aspire scores for fifth grade math were examined to provide data related to

the research questions of the study. School districts A-F used traditional grading and school districts G-L used standards-based grading. Each district's descriptive data was presented as part of the data analysis to provide the reader with a more complete understanding of the make-up for each district.

School A

School A is a suburban school in Arkansas with an enrollment of 294 students. The school is ethnically diverse with 52% Caucasian, 20% Hispanic, 20% two or more races, 6% African American, 2% Asian, and 1% Native American. The percentage of students whose families earn low incomes is 70%. English language learners are 14%.

According to the Arkansas Department of Education, elementary schools have Every Child Succeeds Act (ESSA) public school rating score of 69.16, or "C" for 2022. Arkansas adopted the ACT Aspire as its statewide annual assessment until the year 2024 at which time, it will move to the ATLAS testing platform.

School B

School B is a suburban school in Arkansas with an enrollment of 430 students. The school is ethnically diverse with 54% Hispanic, 16% Caucasian, 13% African American, 10% two or more races, and 8% Asian. The percentage of students whose families earn low incomes is 90%. English language learners are 48%. According to the Arkansas Department of Education, elementary school has Every Child Succeeds Act (ESSA) public school rating score of 57.59, or "F" for 2022.

School C

School C is a rural school in Arkansas with an enrollment of 407 students. The school is not very ethnically diverse with 95.6% Caucasian, 2.7% Hispanic, 1.0% two or

more races, 0.5% African American, and 0.2% Native American. The percentage of students whose families earn low income is 82%. English language learners are 2%. According to the Arkansas Department of Education, elementary schools have ESSA public school rating score of 67.6, or "C" for 2022.

School D

School D is a rural school in Arkansas with an enrollment of 232 students of which 85.4% are Caucasian, 11.6% Hispanic, 2.6% two or more races, and 0.4% African American. The percentage of students whose families earn low income is 80%. English language learners are 6%. According to the Arkansas Department of Education, elementary schools have ESSA public school rating score of 63.78, or "D" for 2022.

School E

School E is a rural school in Arkansas with an enrollment of 347 students of which 58.4% are Caucasian, 29.7% African American, 7.5% Hispanic, 3.5% two or more races, 0.6% Native American, and 0.3% Asian. The percentage of students whose families earn low income is 77%. English language learners are 5%. According to the ADE, the elementary school has ESSA public school rating of 65.68, or "C" for 2022.

School F

School F is a rural elementary school with an enrollment of 395 students. Diversity consists of 73.4% Caucasian, 22.5% Hispanic, 1.8% two or more races, 1.3% Native American, and 1.0% African American. The percentage of students whose families earn low income is 77%. English language learners are 14%. According to the ADE, the elementary school has ESSA public school rating of 70.09, or "C" for 2022.

School G

School G is an urban middle school with an enrollment of 435 students. Diversity consists of 73.9% Caucasian, 2.8% two or more races, and 1.1% Hispanic. The percentage of students whose families earn low income is 72%. English language learners are 0%. According to the ADE, the middle school has ESSA public school rating of 73.24, or "B" for 2022.

School H

School H is a suburban elementary school with an enrollment of 490 students. The diversity of the school includes 57.2% Caucasian, 16.5% African American, 16.5% Hispanic, 5.5% Asian, 4.1% two or more races, and 0.2% Hawaiian/Pacific Islander. The percentage of students whose families earn low income is 43%. English language learners are 14%. According to the ADE, the elementary school has ESSA public school rating of 76.17, or "C" for 2022.

School I

School I is a suburban elementary school with an enrollment of 489 students. The diversity of the school includes 44.2% African American, 20.4% Caucasian, 20% Hispanic, 14.5% two or more races, 0.4% Asian, and 0.4% Native American. The percentage of students whose families earn low income is 96%. English language learners are 16%. According to the ADE, the elementary school has ESSA public school rating of 57.79, or "F" for 2022.

School J

School J is a suburban elementary school with an enrollment of 448 students. The diversity of the school includes 54.6% Hispanic, 21.2% Caucasian, 19.0% Hawaiian/Pacific Islander. 2.5% two or more races, 1.8% African American, 0.7% Asian,

and 0.2% Native American. The percentage of students whose families earn low income is 88%. English language learners are 53%. According to the ADE, the elementary school has ESSA public school rating of 64.29, or "D" for 2022.

School K

School K is a suburban elementary school with an enrollment of 540. The diversity of the school includes 43.3% Caucasian, 37.4% Hispanic, 12.8% Hawaiian/Pacific Islander, 3.0% two or more races, 1.5% African American, 1.3% Asian, and 0.7% Native American. The percentage of students whose families earn low income is 67%. English language learners are 37%. According to the ADE, the elementary school has ESSA public school rating of 70.81, or "C" for 2022.

School L

School L is a suburban elementary school with an enrollment of 593. The diversity of the school includes 52.5% Caucasian, 25.5% Hispanic, 19.7% African American, 1.7% two or more races, 0.3% Asian, and 0.3% Hawaiian/Pacific Islander. The percentage of students whose families earn low income is 64%. English language learners are 17%. According to the ADE, the elementary school has ESSA public school rating of 71.14, or "C".

The sample for the current study consisted of 666 fifth grade students from the twelve participating districts. Of these students, 358 students attend a school who uses standards-based grading, and 308 students attend a school who uses traditional grading.

Findings

The chi-square (χ^2) statistical method was used to compare the categorical values. The alpha level was set at .05 to determine significance. The data collected for this study

were transferred to IBM's statistical analysis software, IBM SPSS, Version 29, to be processed and analyzed. It organizes data from each categorical variable into a table and generates a *p*-value. A *p*-value lower than 0.05 suggests a significant difference among the categorical values. Conversely, if the *p*-value exceeds 0.05, there is no statistical difference among the categorical values (Knapp, 2018).

Research Question 1

The first research question explores whether there's a notable difference in the number of fifth-grade students meeting the math ACT Aspire readiness benchmark between schools employing traditional grading and those using standards-based grading systems.

A chi-square test of independence was performed to evaluate the relationship between grading system and readiness benchmark. The relationship between these variables was significant, $\chi^2(1, N = 666) = 19.29$, $p < .001$. Students graded using a standards-based method were more likely to meet the math readiness benchmark than were those graded using the traditional method. The count for standards-based grading method meeting benchmark (169) was higher than the expected count (141.4). The count for traditional grading method meeting benchmark (94) was lower than the expected count (121.6). Table 3 contains the total number of students who met the readiness benchmark compared to the ones who did not in both the standards-based grading and traditional based grading. Analysis of the results reveals that a higher percentage of students met the math benchmark in schools that used standards-based method of grading.

Table 3

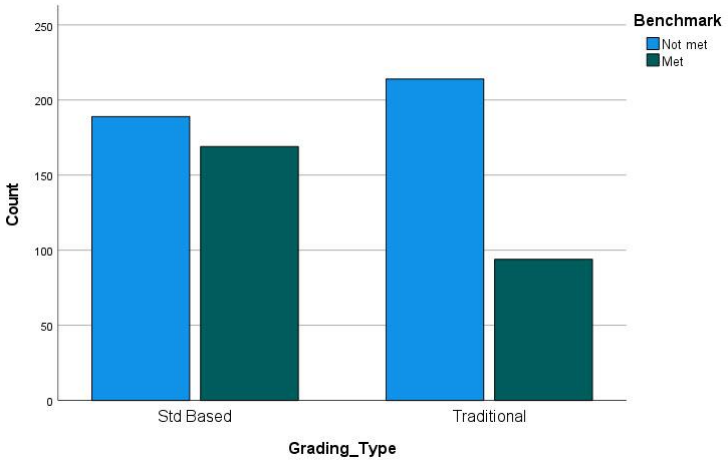
Number of Students Tested by Grading System

Grading System	N		Total
	Benchmark Not Met	Benchmark Met	
Standards-based	189 52.8%	169 47.2%	358
Traditional	214 69.5%	94 30.5%	308
Total	403 60.5%	263 39.5%	666

Figure 4 graphically illustrates the number of students meeting the math readiness benchmark for schools using the traditional and standards-based methods of grading.

Figure 4

Number of Students Tested by Grading System



Research Question 2

The second research question asks if there is a significant difference in the percentage of students meeting the math ACT Aspire readiness benchmark between schools that use traditional versus standards-based grading methods considering the factor of socioeconomic status.

A chi-square test of independence was performed to evaluate the relationship between grading system and readiness benchmark for students who were classified as socioeconomically disadvantaged. The relationship between these variables was significant, $\chi^2(1, N = 525) = 4.53, p < .033$. Students graded using a standards-based method were more likely to meet the math readiness benchmark than were those graded using the traditional method. The count for standards-based grading method meeting benchmark (93) was higher than the expected count (81.6). The count for traditional grading method meeting benchmark (79) was lower than the expected count (90.4). Table 4 contains the total number of students who met the readiness benchmark compared to the ones who did not in both the standards-based grading and traditional based grading. Analysis of the results reveals that a higher percentage of students met the math benchmark in schools that used standards-based method of grading.

Table 4

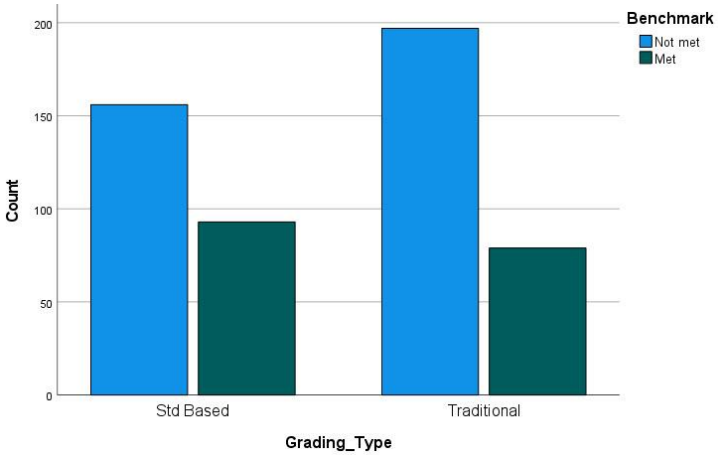
Number of Socioeconomically Disadvantaged Students Tested by Grading System

Grading System	N		Total
	Benchmark Not Met	Benchmark Met	
Standards-based	156 62.7%	93 37.3%	249
Traditional	197 71.4%	79 28.6%	276
Total	353 67.2%	172 32.8%	525

Figure 5 visually depicts the number of socioeconomically disadvantaged students who attained the math readiness benchmark across schools employing traditional and standards-based grading systems.

Figure 5

Number of Socioeconomically Disadvantaged Students Tested by Grading System



A chi-square test of independence was performed to evaluate the relationship between grading system and readiness benchmark for students who were not classified as

socioeconomically disadvantaged. The relationship between these variables was significant, $\chi^2(1, N = 141) = 5.64, p < .018$. Students graded using a standards-based method were more likely to meet the math readiness benchmark than were those graded using the traditional method. The count for standards-based grading method meeting benchmark (76) was higher than the expected count (70.3). The count for traditional grading method meeting benchmark (15) was lower than the expected count (20.7). Table 5 contains the total number of students who met the readiness benchmark compared to the ones who did not in both the standards-based grading and traditional based grading. Analysis of the results reveals that a higher percentage of students met the math benchmark in schools that used standards-based method of grading.

Table 5

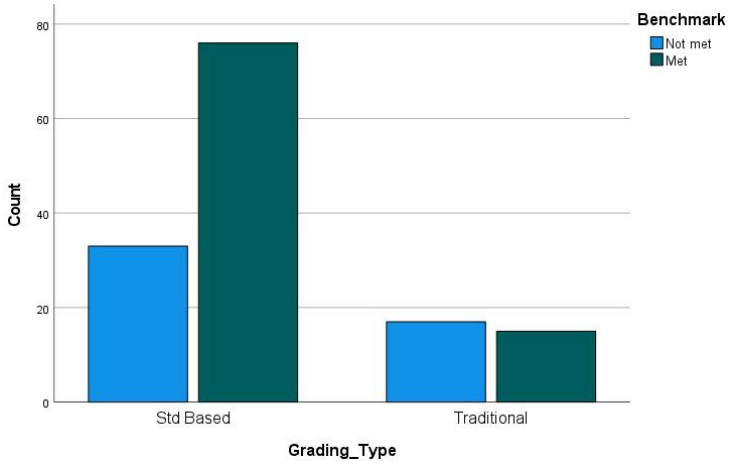
Number of Students Who Are Not Socioeconomically Disadvantaged Tested by Grading System

Grading System	N		Total
	Benchmark Not Met	Benchmark Met	
Standards-based	33 30.3%	76 69.7%	109
Traditional	17 53.1%	15 46.9%	32
Total	50 35.5%	91 64.5%	141

Figure 6 graphically illustrates the number of students who are not socioeconomically disadvantaged meeting the math readiness benchmark for schools using the traditional and standards-based methods of grading.

Figure 6

Number of Students Who Are Not Socioeconomically Disadvantaged Tested by Grading System



Research Question 3

The third research question asks if there is a significant difference in the percentage of students meeting the math ACT Aspire readiness benchmark between schools that use traditional versus standards-based grading methods considering the factor of ethnicity. Table 6 contains the total number of students who met the readiness benchmark by ethnicity compared to the ones who did not in both the standards-based grading and traditional based grading.

Table 6

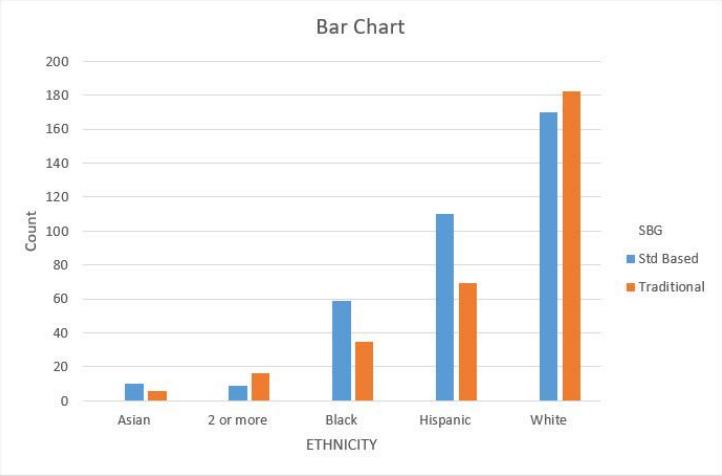
Number of Students by Ethnicity Tested by Grading System

Grading System	Ethnicity	N		Total
		Benchmark Not Met	Benchmark Met	
Standards-Based	Asian	2	8	10
	2 or More	5	4	9
	Black	36	23	59
	Hispanic	67	43	110
	White	59	88	147
	Native American	20	3	23
Traditional	Asian	5	1	6
	2 or More	8	8	16
	Black	31	4	35
	Hispanic	49	20	69
	White	120	61	181
	Native American	1	0	1
Total		402	264	666

Figure 7 graphically illustrates the number of students based on ethnicity who took the math readiness benchmark for schools using the traditional and standards-based method of grading.

Figure 7

Number of Students Considering Factor of Ethnicity Tested by Grading System



Students identified as Asian, Native American, or 2 or More, were excluded from individual statistical analysis due to their low numbers in each category. Individual statistical tests were performed for the ethnicities of Black, Hispanic, and White. Furthermore, students were classified as White and Non-white for analysis.

A chi-square test of independence was performed to evaluate the relationship between grading system and readiness benchmark for students who were classified as White. The relationship between these variables was significant, $\chi^2(1, N = 352) = 14.35$, $p < .001$. Students graded using a standards-based method were more likely to meet the math readiness benchmark than were those graded using the traditional method. The count for standards-based grading method meeting benchmark (91) was higher than the expected count (73.4). The count for traditional grading method meeting benchmark (61) was lower than the expected count (78.6). Table 7 contains the total number of students

who were classified as White and who made benchmark compared to the ones who did not meet benchmark in both the standards-based grading and traditional based grading. Analysis of the results reveals that a higher percentage of students met the math benchmark in schools that used standards-based method of grading.

Table 7

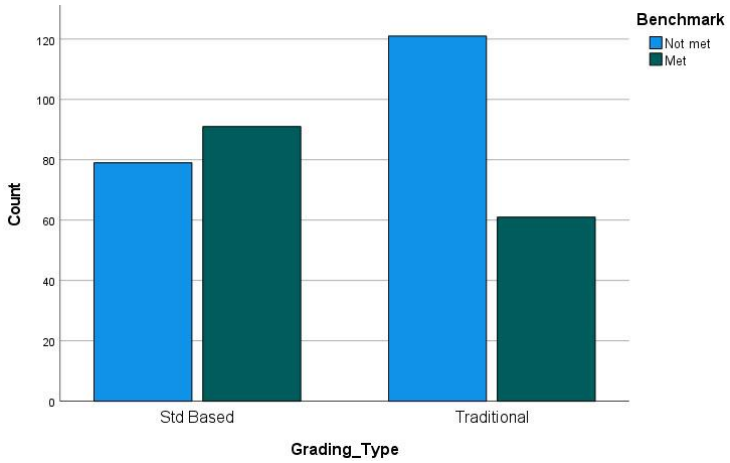
Number of White Students Tested by Grading System

Grading System	N		Total
	Benchmark Not Met	Benchmark Met	
Standards-based	79 46.5%	91 53.5%	170
Traditional	121 66.5%	61 33.5%	182
Total	200 56.8%	152 43.2%	352

Figure 8 graphically illustrates the number of White students meeting the math readiness benchmark for schools using the traditional and standards-based methods of grading.

Figure 8

Number of White Students Tested by Grading System



A chi-square test of independence was performed to evaluate the relationship between grading system and readiness benchmark for students who were classified as Non-White. The relationship between these variables was significant, $\chi^2(1, N = 314) = 7.73$, $p = .005$. Students graded using a standards-based method were more likely to meet the math readiness benchmark than were those graded using the traditional method. The count for standards-based grading method meeting benchmark (78) was higher than the expected count (66.5). The count for traditional grading method meeting benchmark (33) was lower than the expected count (44.5). Table 8 contains the total number of students who were classified as Non-White and who made benchmark compared to the ones who did not meet benchmark in both the standards-based grading and traditional based grading. Analysis of the results reveals that a higher number of students met the math benchmark in schools that used standards-based method of grading.

Table 8

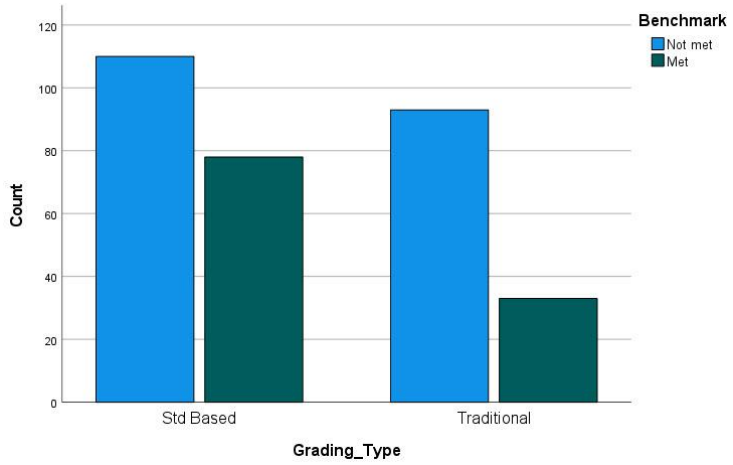
Number of Non-White Students Tested by Grading System

Grading System	N		Total
	Benchmark Not Met	Benchmark Met	
Standards-based	110 58.5%	78 41.5%	188
Traditional	93 73.8%	33 26.2%	126
Total	203 64.6%	111 35.4%	314

Figure 9 graphically illustrates the number of Non-White students meeting the math readiness benchmark for schools using the traditional and standards-based methods of grading.

Figure 9

Number of Non-White Students Tested by Grading System



A chi-square test of independence was performed to evaluate the relationship between grading system and readiness benchmark for students who were classified as Black. The relationship between these variables was significant, $\chi^2(1, N = 94) = 8.15$, $p = .004$. Students graded using a standards-based method were more likely to meet the math readiness benchmark than were those graded using the traditional method. The count for standards-based grading method meeting benchmark (23) was higher than the expected count (16.9). The count for traditional grading method meeting benchmark (4) was lower than the expected count (10.1). Table 9 contains the total number of students who identified as Black and who made benchmark compared to the ones who did not meet benchmark in both the standards-based grading and traditional based grading. Analysis of the results reveals that a higher number of students met the math benchmark in schools that used standards-based method of grading.

Table 9

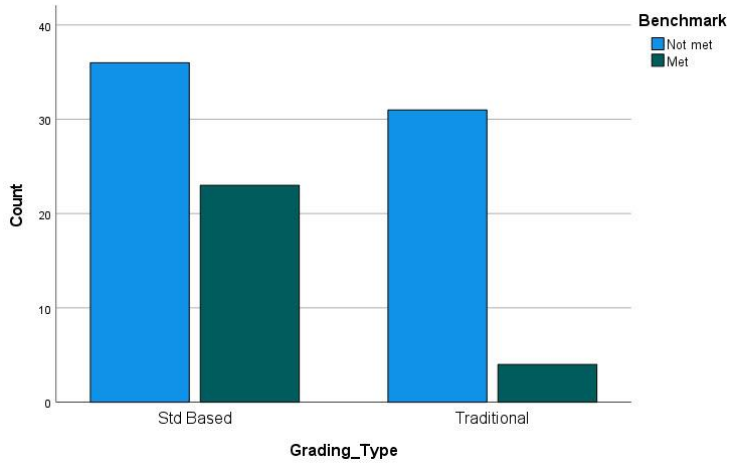
Number of Black Students Tested by Grading System

Grading System	N		Total
	Benchmark Not Met	Benchmark Met	
Standards-based	36 61.0%	23 39.0%	59
Traditional	31 88.6%	4 11.4%	35
Total	67 71.3%	27 28.7%	94

Figure 10 graphically illustrates the number of Black students meeting the math readiness benchmark for schools using the traditional and standards-based methods of grading.

Figure 10

Number of Black Students Tested by Grading System



A chi-square test of independence was performed to evaluate the relationship between grading system and readiness benchmark for students who were classified as Hispanic. The relationship between these variables was not significant, $\chi^2(1, N = 179) = 1.90$, $p = .168$. Students graded using a standards-based method were not more likely to meet the math readiness benchmark than were those graded using the traditional method. The count for standards-based grading method meeting benchmark (43) was higher than the expected count (38.7). The count for traditional grading method meeting benchmark (20) was lower than the expected count (24.3). Table 10 contains the total number of students who identified as Hispanic and who made benchmark compared to the ones who did not meet benchmark in both the standards-based grading and traditional based grading. Analysis of the results reveals that a higher percentage of students met the math

benchmark in schools that used standards-based method of grading even though it was not significant.

Table 10

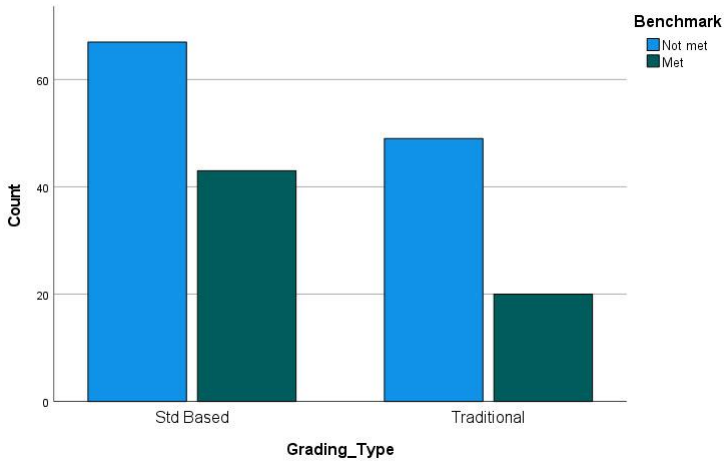
Number of Hispanic Students Tested by Grading System

Grading System	N		Total
	Benchmark Not Met	Benchmark Met	
Standards-based	67 60.9%	43 39.1%	110
Traditional	49 71.0%	20 29.0%	69
Total	116 64.8%	63 35.2%	179

Figure 11 graphically illustrates the number of Hispanic students meeting the math readiness benchmark for schools using the traditional and standards-based methods of grading.

Figure 11

Number of Hispanic Students Tested by Grading System



Research Question 4

The fourth research question asks if there is a significant difference in the percentage of students meeting the math ACT Aspire readiness benchmark between schools that use traditional versus standards-based grading methods considering the factor of gender.

A chi-square test of independence was performed to evaluate the relationship between grading system and readiness benchmark for students who were classified as males. The relationship between these variables was significant, $\chi^2(1, N = 327) = 16.29$, $p < .001$. Students graded using a standards-based method were more likely to meet the math readiness benchmark than were those graded using the traditional method. The count for standards-based grading method meeting benchmark (98) was higher than the expected count (80.0). The count for traditional grading method meeting benchmark (45)

was lower than the expected count (63.0). Table 11 contains the total number of students who identified as male and who made benchmark compared to the ones who did not meet benchmark in both the standards-based grading and traditional based grading. Analysis of the results reveals that a higher number of students met the math benchmark in schools that used standards-based method of grading.

Table 11

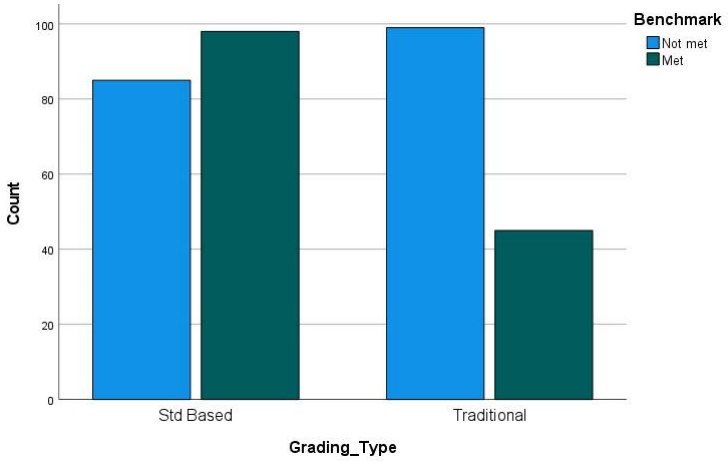
Number of Male Students Tested by Grading System

Grading System	N		Total
	Benchmark Not Met	Benchmark Met	
Standards-based	85 46.4%	98 53.6%	183
Traditional	99 68.8%	45 31.3%	144
Total	184 56.3%	143 43.7%	327

Figure 12 graphically illustrates the number of male students meeting the math readiness benchmark for schools using the traditional and standards-based methods of grading.

Figure 12

Number of Male Students Tested by Grading System



A chi-square test of independence was performed to evaluate the relationship between grading system and readiness benchmark for students who were classified as females. The relationship between these variables was significant, $\chi^2(1, N = 339) = 4.23$, $p = .040$. Students graded using a standards-based method were more likely to meet the math readiness benchmark than were those graded using the traditional method. The count for standards-based grading method meeting benchmark (71) was higher than the expected count (61.9). The count for traditional grading method meeting benchmark (49) was lower than the expected count (58.1). Table 12 contains the total number of students who identified as female and who made benchmark compared to the ones who did not meet benchmark in both the standards-based grading and traditional based grading. Analysis of the results reveals that a higher number of students met the math benchmark in schools that used standards-based method of grading.

Table 12

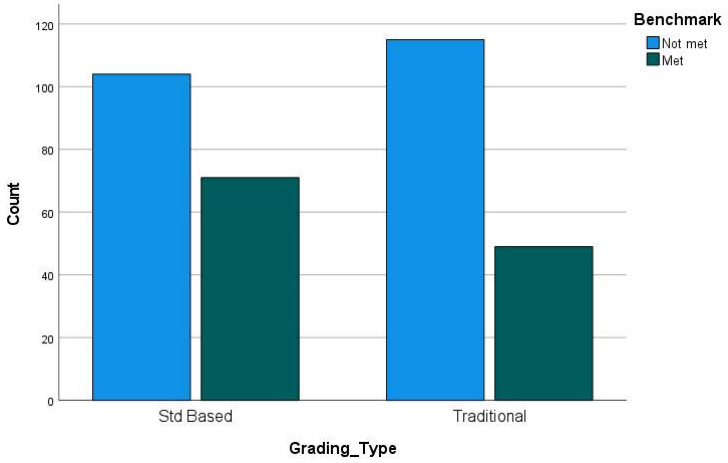
Number of Female Students Tested by Grading System

Grading System	N		Total
	Benchmark Not Met	Benchmark Met	
Standards-based	104 59.4%	71 40.6%	175
Traditional	115 70.1%	49 29.9%	164
Total	219 64.6%	120 35.4%	339

Figure 13 graphically illustrates the number of female students meeting the math readiness benchmark for schools using the traditional and standards-based methods of grading.

Figure 13

Number of Female Students Tested by Grading System



Summary

Chi-square tests of independence were conducted to examine the connection between the grading system type and the number of students meeting the 5th-grade math readiness benchmark. In every case except for Hispanic students, a notable relationship was found between these factors. Schools employing a standards-based grading approach had a greater proportion of students meeting the math readiness benchmark compared to those using the traditional method. This trend was consistent across various socioeconomic statuses, ethnicities, and genders.

Chapter 5 presents a comprehensive overview of the current study along with an analysis of its findings. It will discuss the implications and constraints of the study, and propose suggestions for future research endeavors.

Chapter V: Conclusions

The purpose of this study was to determine whether type of grading system (traditional or standards-based) used to assess student progress makes a difference on the ACT Aspire results. The results of this study provide information to educational leaders to assist in guiding decisions regarding grading practices. Chapter 5 is organized into five sections. The first section provides a summary of the study which includes an overview of the problem, statement of the purpose, restatement of the research questions, methodology review, and summary of the major findings. The second section provides an analysis and interpretation of the findings. The third section details the implications of the study for practitioners. The fourth section is a discussion of limitations of the study. Lastly, the fifth section offers recommendations for future study.

Introduction

Teachers are held accountable for student achievement, and grades must be reflective of teaching (Vatterott, 2015). The ACT Aspire math test was considered a high-stakes assessment system because the results a student earns have many implications for the student and the teacher (Arkansas Department of Education, 2023).

The traditional model of the school system was developed in the nineteenth century during the Industrial Revolution. During this time period, high-stakes assessments were not utilized. The 21st-century education has seen many changes in classroom layout with increased use of teaching with technology. Robinson (2011) believes the system of education utilized in the United States is not designed to meet

challenges current educators face. According to Robinson (2011), educators must encourage transformation rather than reformation.

Many educators view the report *A Nation at Risk: The Imperative for Educational Reform* (National Commission on Excellence in Education, 1983) as the inspiration for the modern standards movement. This publication caused a dramatic shift in education reform by implying that the quality of the American education system was mediocre (Marzano, 1998). By 1998, standards were developed for most content areas in public schools.

School leaders and teachers began to feel increased accountability as a result of the No Child Left Behind Act of 2001 and the Race to the Top Initiative of 2009 (United States Department of Education, 2009). These laws caused school leaders to examine pedagogical practices within the classroom. No longer are teachers teaching to the test but teaching to state standards aligned with the weighted categories released by the testing centers. School leaders have shown an increase in weighing more emphasis on student mastery of state standards (Brookhart, 2017).

The main objective of standards-based grading is to ensure that teachers establish clear learning objectives, effectively communicate these objectives, and accurately assess students' mastery of the content standards. These grades are determined by the extent to which students have mastered the content, rather than simply reflecting the grades they have earned (Brookhart, 2011). Moss and Brookhart (2012) claimed, “The most effective teaching and the most meaningful student learning happens when teachers design the right learning target for the day’s lesson and use it along with their students to aim for and assess understanding” (p. 2). Standards-based grading promotes the use of a rubric to

analyze student mastery. Rubrics allow for student growth measures to be determined and in turn allow teachers to differentiate instruction on the students that have the opportunity to demonstrate true mastery of standards (Brookhart, 2013). In standards-based grading, as students delve deeper into the content, their final assessment or project submission reflects their current level of achievement for a specific standard (Rundquist, 2012). This method shows the progress the student has made through their learning experiences. By reporting standard proficiency, each subject area is specified by the learning target which focus on exact content elements (Guskey & Jung, 2006).

This study focused on twelve schools located across Arkansas, with six schools utilizing standards-based grading and the other six employing traditional grading systems. Among the schools implementing standards-based grading, the duration of its implementation ranged from 2 to 6 years.

Summary of Study

In 2022, a report from the National Assessment of Educational Progress stated, "math scores fell in nearly every state. This represented the steepest decline ever recorded." The students in the bottom 25th percentile lost the most ground compared to their classmates (Mervosh & Wu, 2022).

In 2023, Arkansas Governor, Sarah Huckabee Sanders passed the LEARNS Act. Within this new law, it states that by the 2023-24 school year, schools must develop a math intervention plan for 3rd-8th graders not performing at grade level. The law also states that by the 2024-25 school year, each district must report the type of math intervention they are using and the number of students receiving them (Arkansas Department of Education, 2023).

At the end of the 2022-23 school year, Arkansas had replaced the ACT Aspire exam to the Arkansas Teaching and Learning Assessment System (ATLAS) along with rewriting Arkansas' core subject standards. The ATLAS is a standardized test that measures the reading, writing, science, and math skills of students in grades 3-8 in Arkansas. It is aligned to the Arkansas Curriculum Standards and used to track student progress (Arkansas Department of Education, 2023).

Standards-based grading helps ensure grading is directly correlated to the mastery of the learning targets. The correlation of grades to the mastery of learning targets gives grades meaning and communicates students' progress towards mastery. Quality information provided through student progress measures allows teachers to adjust instruction and differentiate based on specific student need (Scriffiny, 2008).

The primary objective of all teachers is for their students to master the academic content in a particular subject area. For teachers to determine if students are mastering the content, teachers must assess students on content covered in the instruction. Grades serve as the communication tool to inform teachers, parents, and students about the student's progress toward mastery of the content covered during instruction (Allen, 2005).

Deddeh, et al. (2010) claimed traditional forms of grading did not provide stakeholders with adequate predictors of student success on standardized assessments due to the subjectivity of evaluations and the vastly different approach to grading. Wormeli (2006) stated that the inclusion of nonacademic achievement factors in traditional grading systems further compounds the relationship between grades and student performance on standardized assessments. By comparing differences in scores on ACT Aspire tests of

schools who use standards-based and traditional grading methods, school district leaders can implement policies and procedures regarding classroom grading practices.

In this study, the researcher investigated the relationship between two categorical variables. The key aspects of the study include: a) comparing two categorical variables, b) collecting data at a single time point, c) gathering at least two scores for each categorical variable, d) reporting Chi-Square percentages along with the coefficient in the data analysis, and e) interpreting the results of the statistical test, as outlined by Creswell (2012).

The research questions provided the framework and direction for the current research study:

Research Question 1 (RQ1): Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems?

The hypothesis was tested by using the Chi-Square test to compare two categorical variables: the type of grading method utilized and whether the benchmark was achieved on the ACT Aspire math test.

Research Question 2 (RQ2): Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of socioeconomic status?

The hypothesis was tested by using the Chi-Square test to compare two categorical variables: the type of grading method utilized and whether the benchmark was achieved on the ACT Aspire math test. Two subgroups were examined: individuals classified as socioeconomically disadvantaged and those who were not.

Research Question 3 (RQ3): Is there is a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of ethnicity?

The hypothesis was tested by using the Chi-Square test to compare two categorical variables: the type of grading method utilized and whether the benchmark was achieved on the ACT Aspire math test. Four subgroups were investigated: White, Black, Hispanic, and Non-white students, with the latter encompassing all minority students.

Research Question 4 (RQ4): Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of gender?

The hypothesis was tested by using the Chi-Square test to compare two categorical variables: the type of grading method utilized and whether the benchmark was achieved on the ACT Aspire math test. Two subgroups were examined: males and females.

Summary of Findings and Interpretation of Results

The results of the four research questions indicated a statistically significant higher percentage of students meeting the math benchmark in schools that employed a standards-based grading method. In this section, we will assess and review each research question and hypothesis. Additionally, we will analyze the results to ascertain their relevance to the research discussed in the literature review in Chapter 2.

Research Question 1 (RQ1): *Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems?*

In the current study, the researcher found results of the Chi-Square test reveal that a significant relationship exists between the two variables. Analysis of the results reveals that a higher percentage of students met the math benchmark in schools that used the standards-based method of grading. Therefore, the researcher concludes there is a significant relationship between the variables: type of grading method used (standards-based or traditional) and whether benchmark was met or not on the ACT Aspire math test.

This question encompassed all students, regardless of their socioeconomic status, ethnicity, or gender. Although a significant relationship was found between the type of grading and the number of students meeting the math readiness benchmark, the researcher decided to delve deeper into the data to explore whether this finding applied across different subgroups.

Research Question 2 (RQ2): *Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of socioeconomic status?*

Students were identified as socioeconomically disadvantaged if they qualified for free or reduced-price (FR) meals at school. A Chi-square test of independence was conducted to examine the connection between the grading system type and the number of students meeting the 5th-grade math readiness benchmark considering only students identified as socioeconomically disadvantaged. Another test was conducted involving only students who were not identified as socioeconomically disadvantaged.

The researcher found results of both Chi-Square test outcomes indicated a significant relationship between the two variables. Further examination of the findings suggests that a greater percentage of students achieved the math benchmark in schools utilizing the standards-based grading approach regardless of socioeconomic status.

Research Question 3 (RQ3): *Is there is a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of ethnicity?*

Students self-identify with various ethnicities such as White, Black, Hispanic, Asian, Native American, or belonging to two or more races. The researcher chose to further analyze the data to investigate whether student ethnicity could lead to different outcomes. A series of Chi-square tests of independence were performed to explore the relationship between the type of grading system and the number of students meeting the 5th-grade math readiness benchmark within each ethnic group.

Students identified as Asian, Native American, or belonging to two or more races were excluded from individual statistical analysis due to their small numbers in each category. Instead, individual statistical tests were conducted for the ethnicities of White, Black, Hispanic, and Non-white, which comprised all students not identified as White.

The results of the Chi-Square tests reveal that a significant relationship exists between the two categories in all cases except the Hispanic students. Even with that exception, analysis of the results reveals that a higher percentage of students met the math benchmark in schools that used standards-based method of grading.

Research Question 4 (RQ4): *Is there a significant difference in fifth-grade math ACT Aspire scores between schools that use traditional versus standards-based grading systems considering the factor of gender?*

Two Chi-square tests of independence were performed to explore the relationship between the type of grading system and the number of students meeting the 5th-grade math readiness benchmark, one involving only male students and the other involving only female students.

The researcher found results of both Chi-Square test outcomes indicated a significant relationship between the two variables. Further examination of the findings suggests that a greater percentage of students achieved the math benchmark in schools utilizing the standards-based grading approach

Conclusions

Mueller (2018) suggested that rubrics used to design assessments covering standards lead to increased student performance by clearly showing the student how their work will be evaluated and what is expected to determine if students have learned the goal or academic learning target. Beatty (2013) argued feedback linked to specific learning objectives gives students direction and a clear understanding of what they have and have not mastered. Close (2009) contended that grades should consist of an expert evaluation of a student's mastery of course content. Spencer (2012) and Guskey (2011) stated that when educators provide students with specific, clearly defined goals, determine levels of mastery, and communicate with parents and students when those goals are achieved, student performance improves.

The purpose of standards-based grading is to show how well students have mastered the content standards. The findings in this study suggest that when teachers use standards-based grading, student performance improves. Standards-based grading appears to provide teachers with a more accurate way to assess students' mastery of the content being taught which translates into improved learning.

Limitations

This study was constrained by the availability of an exam administered once a year under the direction of the state. Students might have performed poorly on the exam due to potential lack of exposure to the testing format.

Only twelve schools, evenly divided between those employing traditional grading methods and those utilizing standards-based grading, were included in the study. At the outset of this study, there were 33 schools employing standards-based grading. A different outcome might have arisen if a random sample of six different schools had been selected.

This study did not take into account other variables such as the effectiveness of classroom teachers, student motivation, or the provision of accommodations aimed at enhancing performance on the ACT Aspire math test, which could potentially influence student performance.

This study solely focused on one subject at a single elementary grade level. Different outcomes might have emerged if the study had included other subjects or grade levels.

Recommendations for Further Research

Another study could investigate various subjects and grade levels to offer additional evidence supporting standards-based grading over traditional grading, or vise versa. Data from other states could be examined.

A mixed methods or qualitative study could provide valuable insights into the effectiveness of standards-based grading from the viewpoint of teachers.

This study provided a snapshot view by only analyzing data from one year. A similar study could explore results from multiple years of administration of the exam for a more comprehensive understanding.

As Arkansas transitions to a new testing platform aligned with Arkansas standards, comparing the results from this test with other archived data from the ACT Aspire could provide more comprehensive insights by correlating the data from the tests. This analysis could help inform district decisions regarding grading methods.

Summary

The aim of the current study was to inform the standards-based grading practices of schools in Arkansas. Despite its limitations, the research findings offer valuable insights to assist educational leaders in making crucial decisions regarding grading practices. The researcher investigated the differences between standards-based and traditional grading methods in relation to student performance on the ACT Aspire 5th grade math test. Based on the study's findings, schools are encouraged to continue supporting the adoption of standards-based grading. As accountability for student achievement increases, educators must strive to identify the most effective methods for measuring student growth and mastery. The findings of this study, combined with

existing literature and prior research on standards-based grading, strongly suggest that it is a viable and sustainable option for school leaders.

References

ACT (2021). Retrieved from ACT Aspire - Products and Services | ACT.

Airasian, P. (2001). Classroom Assessment: Concepts and Applications (4th Ed). Blacklick, OH: McGraw-Hill.

Author unknown (n.d.) Common Goal Systems, Inc., What is Standards-Based Grading? Retrieved from https://www.teacherease.com/standards-based-grading.aspx

Author, Unknown (2022). Lev Vygotsky's Sociocultural Theory of Cognitive Development. *The* Psychology Notes HQ, Online Resources for Psychology Students. Retrieved from https://www.psychologynoteshq.com/vygotsky-theory/

Author, Unknown (2023). The Complete Beginner's Guide to Mastery Learning and Standards-Based Grading. Retrieved from The-Ultimate-Guide-to-SBG.pdf (otus.com)

Bowman, J. D. (2020). Edutopia. Getting Students to Engage with Feedback. Retrieved from https://www.edutopia.org/article/getting-students-engage-feedback.

Briars, D. & Foster, D. (2012). Common Core Mathematics in a PLC at Work, Grades 6-8. Bloomington, IN: Solution Tree Press.

Brookhart, S. M. (2017). How to Use Grading to Improve Learning. ASCD. Retrieved from https://files.ascd.org/staticfiles/ascd/pdf/siteASCD/publications/books/HowToUseGradingToImproveLearning.pdf.

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., Smith, L. F., ... Welsh, M. E. (2016, December). A century of grading research: meaning

and value in the most common educational measure. *Review of Educational Research*, *86*, 803-848. http://dx.doi.org/doi: 10.3102/0034654316672069.

Brookhart, S. M. (2014). Teachers' grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279-301.

Brookhart, S. M. (2013). *How to create and use rubrics for formative assessment and grading.* Alexandria, VA: ASCD.

Brookhart, S. M. (2012). How to Give Effective Feedback to Your Students. Alexandria, VA: Association for Supervision and Curriculum Development.

Brookhart, S. M. (2012). Action Tool A: Understanding Learning Targets. Retrieved from Moss & Brookhart-Learning Targets pages.indd (ascd.org)

Brookhart, S. M. (2011). Starting the conversation about grading. Educational Leadership, 69(3), 10-14.

Brookhart, S. M. (2009). *Grading* (2nd ed.). Upper Saddle River, NJ: Pearson.

Close, D. (2009). Fair Grades. *Teaching Philosophy*, 32(4), 361-398. Retrieved from teachphil_2009_0032_0004_0035_0072.pdf (pdcnet.org)

Cox, K. B. (2011). Putting classroom grading on the table: A reform in progress. *American Secondary Education*, *40*(1), 67–87. Abstract retrieved from https://eric.ed.gov/?id=EJ951229

Crowe, S., Cresswell, K., Robertson, A., Huby, G., Avery, A., & Sheikh, A. (2011). The case study approach. BMC Medical Research Methodology, 11(1), 100-108. doi:10.1186/1471-2288-11-100

Deddah, H., Main, E., & Fulkerson, S. (2010). Eight steps to meaningful grading. *The Phi Delta Kappan*, *91*(7), 53-58.

Dweck, C. S. (2006). *Mindset: The new psychology of success.* New York, NY: Ballantine Books.

Effects of Grading System on Students: Pros and Cons. (n.d.). Retrieved from https://www.uopeople.edu/blog/effects-of-grading-system-on-students/

Farrell, S. (2016). 28 Tips for Creating Great Qualitative Surveys. Retrieved from https://www.nngroup.com/articles/qualitative-surveys/.

Feldmesser, R. (1971). The Positive Function of Grades. Retrieved from ED049704.tif.pdf.

Finkelstein, I. E. (1913). *The marking system in theory and practice*. Baltimore, MD: Warwick & York.

Fleenor, A., Lamb, S., Stinson, T., Donen, T. (2011). The grades game. *Principal Leadership*, 11(6), 48-52.

Frisbie, D. A. & Waltman, K. K. (1992). Developing a Personal Grading Plan. Retrieved from Developing a Personal Grading Plan - Frisbie - 1992 - Educational Measurement: Issues and Practice - Wiley Online Library.

Gist, M. & Mitchell, T. (1992). Self-efficacy: A theoretical analysis of its determinants and malleability. *Academy of Management Review*, 17(2), 183-211.

Goals 2000: Educate America Act (1994). (P.L. 103-227). Retrieved from Goals 2000: Educate America Act – US Metric Association (usma.org)

Guskey, T. R. (2015). On your mark: Challenging the conventions of grading and reporting. Bloomington, IN: Solution Tree Press.

Guskey, T. R. (2013). The case against percentage grades. *Educational Leadership*, 71(1), 68-72.

Guskey, T. R. (2010). Lessons of Mastery Learning. *ASCD*, October 1, 2010, 68(2).

Guskey, T. R. (2009). *Bound by tradition: Teachers' views of crucial grading and reporting issues.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Guskey, T. R. (2010). Lessons of Mastery of Learning. *ASCD, 68(2).*

Guskey, T. R. (2001). Fixing grading policies that undermine standards. *Education Digest,*

66(7), 16-21.

Guskey, T. R., & Bailey, J. M. (2001). Developing Grading and Reporting Systems for Student Learning. Experts in Assessment. Thousand Oaks, CA: Corwing Press, Inc.

Guskey, T. R., & Brookhart, S. M. (2019). What we know about grading What works, what doesn't and what's next. Alexandria, VA: ASCD.

Guskey, T. R., Swan, G., & Jung, L. (2011). Grades that mean something. Kappan, 93(2), 52-57.

Guskey, T. R. & Jung, L. (2008). Retrieved from http://www.naesp.org/ContentLoad.do?contentId=2127&action=print.

Guskey, T. R. & Jung, L. (2013). *Answers to essential questions about standards, assessments, grading, and reporting.* Thousand Oaks, CA: Corwin Press.

Harding, J. (2013). Qualitative data analysis from start to finish (1st ed.). Los Angeles, CA: Sage Publications

Heflebower, T., Hoegh, J. K., Warrick, P. B., McInteer, M., Hoback, M., & Clemens, B. (2014). *A school leader's guide to standards-based grading.* Marzano Research Laboratory.

Hesse-Biber, S. N, & Leavy, P. (2010). Handbook of emergent methods. New York, NY: The Guilford Press

Hobden, H. (2019). Traditional Grading Scale: Pros, Cons, & Alternatives. Retrieved from https://kid.io/traditional-grading-scale-pros-cons-alternatives/

Kelly, Melissa. (2020, August 26). The Many Purposes of Education. Retrieved from https://www.thoughtco.com/what-is-the-aim-of-education-8417

Klein, A. (2016). No Child Left Behind: An overview. *Education Week*, 34(27). http://www.edweek.org/ew/section/multimedia/no-child-left-behind-overview-definitionsummary.html.

Knapp, H. (2018). Intermediate Statistics Using SPSS. Thousand Oaks, CA: SAGE Publications, Inc.

Kunnath, J.P. (2017). Creating Meaningful Grades. Retrieved from https://files.eric.ed.gov/fulltext/EJ1158167.pdf

Kunnath, J.P. (2016). A critical pedagogy perspective of the impact of school poverty level on the teacher grading decision-making process. (Doctoral dissertation). Retrieved from http://scholarworks.calstate.edu/bitstream/handle/10211.3/179568/JoshKUNNATH.pdf?sequence=1

Lassahn, N. (n.d.) History of Grading Systems. Retrieved from https://classroom.synonym.com/history-grading-systems-5103640.html.

Lee, C. (2020). What is the History of Grading. Retrieved from What Is the History of Grading? | Turnitin.

Linn, R. (1983). Testing and Instruction: Links and Distinctions. *Journal of Educational Measurement*, 59(3), p 179-189. Retrieved from TESTING AND INSTRUCTION: LINKS AND DISTINCTIONS - LINN - 1983 - Journal of Educational Measurement - Wiley Online Library.

Marzano, R. (2011). Formative Assessment & Standards-Based Grading. Solution Tree. Press.

Marzano, R. (2006). Classroom assessment & grading that work. Alexandria, VA: ASCD. 79

Marzano, R. (1998). A Theory-Based Meta-Analysis of Research on Instruction. Retrieved ED427087.pdf.

Marzano, R., & Heflebower, T. (2011). Grades that show what students know. Educational Leadership, 69(3), 34-39.

Marzano, R., & Kendall, J.S. (1996). *The fall and rise of standards-based education.* Arlington, VA: NASBE.

McCusker, K., & Gunaydin, S. (2015). Research using qualitative, quantitative or mixed methods and choice based on the research. Perfusion, 30(7), 537-542. doi:10.1177/0267659114559116

McLeod, S. A. (2020, July 22). *Lev Vygotsky.* Simply Psychology. https://www.simplypsychology.org/vygotsky.html.

McMillan, J. & Nash S. (2000). Teacher classroom assessment and grading practices decision making.

McMillan, J. H., Myran, S., & Workman, D. (2002). Elementary Teachers' Classroom Assessment and Grading Practices. *The Journal of Educational Research*, 95(4), 203-213. http://www.jstor.org/stable/27542381

Merriam, S. B., & Tisdell, E. J. (2016). Qualitative research a guide to design and implementation (4th ed.). [Kindle]. Retrieved from https://read.amazon.com/?asin=B01176PX48

Meyer, M. (1908, August 21). The grading of students. *Science*, *28*, 243-250. Retrieved from https://www.jstor.org/stable/pdf/1635696.pdf.

Morin, A. (n.d.). What is growth mindset?. Retrieved from Growth Mindset: How to Develop Growth Mindset | Understood.

Moss, C. M., & Brookhart, S. M. (2012). *Learning targets: Helping student's aim understanding in today's lesson.* Alexandria, VA: ASCD.

Mueller, J. (2018). Authentic Assessment Toolbox http://jfmueller.faculty.noctrl.edu./toolbox/whatisit.htm.

National Commission on Excellence in Education (1983). *A nation at risk: The imperative for educational reform. A report to the nation and the Secretary of Education, United States Department of Education*. Washington, DC: The Commission.

National Governor's Association Center for Best Practices, Council of Chief State School Officers. (2012). Common Standards. Retrieved from www.corestandards.org.

O'Connor, K. (2018). How to grade for learning: Linking Grades to Standards *(*4th ed.). Corwin.

O'Connor, K. (2011). A repair kit for grading: 15 fixes for broken grades. Boston, MA: Pearson Education.

O'Connor, K. (2009). *How to Grade for Learning.* Thousand Oaks, CA: Corwin.

O'Connor, K. & Wormelli, R. (2011). Reporting student learning. *Educational Leadership*, 69(3), 40-44.

Paeplow, C. G. (2011). Easy as 1, 2, 3: Exploring the implementation of Standards-Based Grading in Wake County Elementary Schools. Retrieved from EASY AS 1, 2, 3: (ncsu.edu).

Patton, M. Q. (2015). Qualitative research and evaluation methods (4th ed.). Thousand Oaks, California: SAGE Publications.

Polikoff, M. S. (2015). How Well Aligned Are Textbooks to the Common Core Standards in Mathematics? *American Educational Research Journal*, *52*(6), 1185–1211. https://doi.org/10.3102/0002831215584435

Polikoff, M. S., Porter, A. C., & Smithson, J. (2011). How Well Aligned Are State Assessments of Student Achievement With State Content Standards? American Educational Research Journal, 48(4), 965–995. https://doi.org/10.3102/0002831211410684

Popham, W. J. (1987). The merits of measurement-driven instruction. *Phi Delta Kappan Magazine*, 68, 679.

Prejean-Harris, R. M. (2013). The relationship between standards-based reporting systems and third grade mathematics and science achievement (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses. (UMI No. 3573420).

Ravitch. D. (1995). *National standards in American education: A citizen's guide.* Washington, DC: Brooking Institute.

Reeves, D. B. (2011). Taking the grading conversation public. *Educational Leadership,* 69 (3), 76-79.

Reeves, D. B. (2004). *Accountability for learning: How teachers and school leaders can take charge.* ASCD.

Ricci, M. K. (2015). Mindsets in the Classroom: Building a Growth Mindset Learning Community. Routledge, NY: Taylor & Francis Group.

Robinson, S. K. (2011). *Out of our minds: Learning to be creative.* Westford, MA: Courier Westford, Inc.

Schimmer, T. (2016). *Grading from the inside out: Bringing accuracy to student assessment through a standards-based mindset.* Bloomington, IN: Solution Tree Press.

Schneider, J., & Hutt, E. L. (2014). Making the grade: A history of the A-F marking scheme. *Journal of Curriculum Studies*, 46, 201-224.

Scriffiny, P. L. (2008). Seven Reasons for Standards-Based Grading. *Educational Leadership*, (66), 2, 70-74.

Shepard, L. (1993). *Setting performance standards for student achievement.* Stanford, CA: National Academy of Education, Stanford University.

Spencer, K. (2012). Standards-Based Grading: New Report Cards Aim to Make Mastery Clear.

Starch, D., & Elliott, E. C. (1913). Reliability of grading work in mathematics. *School Review*, 254-259.

Stiggins, R. (2014). Revolutionize assessment. Thousand Oaks, CA: Corwin Press.

Streifer, A. & Palmer, M. (2020). Alternative Grading: Practices to Support Both Equity and Learning. Retrieved from Alternative Grading: Practices to Support Both Equity Learning | Center for Teaching Excellence (virginia.edu).

The Glossary of Education Reform. (2019). Retrieved from http:edglossary.org/standards-based/.

Townsley, M. & Wear, N. L. (2020). Making Grades Matter: Standards-Based Grading in a Secondary PLC at Work. Solution Tree Press.

U.S. Department of Education (2016). Every Student Succeeds Act of 2015. Retrieved from https://www.ed.gov/essa?src%3Drn.

U.S. Department of Education (2001). The No Child Left Behind Act of 2001. Retrieved from https://search.usa.gov/search?utf8=%E2%9C%93&affiliate=ed.gov&query=no+child+left+behind+act.

U.S. Department of Education (2012). Race To The Top. Retrieved from Race To The Top - Office of Elementary and Secondary Education.

Vatterott, C. (2015). *Rethinking grading: Meaningful assessment for standards-based.* ASCD.

Vygotsky, L. (1978). Social Learning Theory. Retrieved from Social Development Theory (Lev Vygotsky) - InstructionalDesign.org

Welsh, M. E. & D'Agostino, J. V. (2009). Fostering consistency between standards-based grades and large-scale assessment results. In T. R. Guskey (Ed.), *Practical*

*solutions for serious problems in standards-based grading (*p. 75-104). Thousand Oaks, CA: Corwin Press.

Wheelock, A. (1995). Standards-based reform: What does it mean for the middle grades? *Edna McConnell Clar Foundation,* Retrieved from http://files.eric.ed.gov/fulltext/ED389882.pdf

Wormeli, R. (2018). Fair isn't always equal: Assessing & grading in the differentiated classroom. Portland, ME: Stenhouse Publishers.

Zoeckler, L. G. (2007). Moral aspects of grading: A study of high school English teachers' perceptions. *American Secondary Education,* 35(2) 83-102.