The Difference Between 11th-Grade ACT Scores Before and After the Implementation of Standards-Based Grading

Molly Smith

BA, William Jewell College, 2007

MS, Northwest Missouri State University, 2012

Submitted to the Faculty of the School of Education of Baker University

in partial fulfillment of the requirements for the degree of

Doctor of Education in Educational Leadership

Susan K Rogers
Susan K. Rogers, PhD
Major Advisor

Brian Henry
Brian Henry, EdD

Mark Maus
Mark Maus, EdD

Date Defended: December 15, 2022

Abstract

Leaders in District A had data to show that as students grew older, the less engaged they became in school, the lower the student achievement; the district was not meeting their goal number of students being college ready or persisting to college. District leaders noticed that the traditional grading system was putting barriers in place for students (District A deputy superintendent, personal communication, April 5, 2020). When the COVID-19 pandemic hit, District A leaders decided to re-evaluate the grading system at the secondary level. Standards-based grading aims to add more meaning to a student's grade by making grades a better representation of what a student knows or does not know (Heflebower et al., 2014). The first purpose of this study was to determine the extent there is a difference in ACT composite, English, math, reading, and science scores of students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022). The second purpose of this study was to determine the extent the difference in ACT composite, English, math, reading, and science scores of first-time 11th-grade ACT test takers taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022) are affected by student gender and ethnicity. Results of this study showed a significant difference by grading type with a decrease in ACT composite, English, math, reading, and science scores in the standards-based system. Results by gender and ethnicity showed no effect on the difference in composite scores or subscores. The implications for action include continuing the study for three more years to not include the COVID-19 year and investing in professional

development in standards-based learning. Recommendations for future research include using grade point average in place of ACT scores and disaggregating the data by school.

Dedication

This dissertation is dedicated to my husband, Wade, and our three kids, Murphy, Thorsten, and Nelle. I love you most.

Acknowledgements

First, I would like to acknowledge my husband and children. You all gave up so much time with me, and I am excited to spend more time playing games, going to parks, and having sleepovers instead of writing.

Thank you also to Dr. Susan Rogers, who was there pushing me even when I did not want to be pushed. Thank you also for telling me I would not finish when I wanted and giving me the motivation I needed to prove you wrong. I mean this in the best way possible.

Thank you to Dr. Peg Waterman for always taking the time to help, run data, and provide feedback, but mostly thank you for sharing my love of data and spreadsheets.

Thank you to Dr. Brian Henry for agreeing to serve on my committee and giving me feedback throughout this process.

Thank you to Dr. Mark Maus for not only agreeing to serve on my committee but also serving as my mentor and friend for so many years. Your opinion means so much to me, but please do not judge me on this dissertation.

Finally, thank you to my dear friend Jackie Gosney who was beside me on this journey and made it much more bearable and fun. Also, thank you to all the DoorDash drivers that delivered us food every Thursday. I could not have made it without you.

Table of Contents

List of Tables

Chapter 1

Introduction

The traditional 100-point grading system dates to the early 19th century (Durm, 1993). The purpose of the 100-point system was to advance students in grade levels (Spencer, 2012). In 1913, Finkelstein wrote about how uncalibrated the traditional 100-point grading system was, yet in 2021, American schools continued to use it at the time this research was conducted.

According to Mahr (2020), the reauthorization of the Elementary and Secondary Education Act in 1994 and the signing of the No Child Left Behind Act in 2001 opened the door to standards-based instruction (SBI) and standards-based grading (SBG). These were followed by the 2009 adoption of the Common Core State Standards (Mahr, 2020). The intent behind all these reforms was to help bring consistency to our education system that was lacking in the traditional system. Consistency makes students more successful because expectations are clear (Mahr, 2020). Standards-based education is the idea that teachers might have better-defined goals and an outline of what those goals look like (Spencer, 2012). A clearer vision of what the student is attempting to achieve allows for easier communication between teachers and parents about the learning progress toward academic goals (Spencer, 2012). This clarity potentially keeps the focus on learning instead of behavioral compliance.

Background

This study was conducted in school District A, a district of over 20,000 kindergarteners through 12th-grade students located in the Midwest (Department of

Elementary and Secondary Education, 2022). Demographics for District A can be seen in Table 1.

Table 1

District A Demographics 2021

Demographics	% of Students
Ethnicity	
American Indian/Alaska Native	0.30
Asian	3.40
Black	14.80
Hawaiian/Pacific Islander	1.30
Hispanic	14.80
Multi-Race	10.80
White	54.80
Socioeconomic Status	
Free or Reduced Lunch	38.80
Full Pay Lunch	61.20

Note. Percentages were rounded to the nearest tenth and, based on rounding error, do not sum exactly to 100 for ethnicity. Adapted from *District Demographic Data*, by Department of Elementary and Secondary Education, 2022. https://apps.dese.mo.gov/MCDS/Reports/SSRS_Print.aspx?Reportid=6c5b805c-5af7-4c33-be41-dc2b83ded4aa

District leaders analyzed student survey data and found that as students grew older, the less engaged they became in school and the lower the student achievement. Also, the district was not meeting its goal number of students being college ready or persisting to college. The district had a strong focus on equity for all students and noticed that the traditional grading system was putting barriers in place for students (District A deputy superintendent, personal communication, April 5, 2020). When the COVID-19 pandemic hit, District A leaders decided to re-evaluate its grading system at the secondary level. The district wanted something that truly showed what a student knew and did not consider their behaviors or circumstances (District A deputy superintendent, personal communication, April 5, 2020). The district decided on standards-based grading.

There are four high schools in District A; three of the four piloted a version of standards-based grading. Heflebower et al. (2014) outlined a process to shift to standards-based grading, beginning with identifying priority standards and developing proficiency scales. Priority standards were identified by the state of Missouri and District A. Teachers in each building developed their own proficiency scales. As schools make the shift to standards-based grading, using a standards-based report card is usually toward the end of implementation (Heflebower, 2014). At first, many schools still lived in a percentage system, which was the case in District A. Examples of grading scales used can be seen in Appendix A.

Statement of the Problem

The traditional grading system dates to the early 19th century and, like many things in education, has not changed in the last 200 years (Durm, 1993). The drawback of this grading system is that it does not represent what a student knows but how well

students play the game of school. Grades are not always a reflection of student knowledge (Muñoz & Guskey, 2015). Standards-based grading is a new practice implemented at the high school level in District A (District A deputy superintendent, personal communication, April 5, 2020). Because of the lack of data surrounding SBG, there is a need to explore the effect grading practices, specifically traditional versus standards-based, have on a student's academic achievement. Because it is new, we do not know if standards-based grading practices used in the classroom have impacted 11th-grade student's performance on their first attempt at taking the ACT. The role grading plays in the lives of our students makes it an important topic to study.

Purpose of the Study

The overall purpose of this study was to determine how first-time 11th-grade ACT test-takers score on the ACT. The first purpose was to determine the extent there is a difference in ACT composite, English, math, reading, and science scores of students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022). The second purpose of this study was to determine the extent the difference in ACT composite, English, math, reading, and science scores of first-time 11th-grade ACT test takers taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022) are affected by student gender and ethnicity.

Significance of the Study

Pattison et al. (as cited in Brookhart et al., 2016) purported that grades are important, especially at the high school level, because they could predict a student's chance of finishing high school and their future after high school. Some researchers have

aimed to show how SBG impacts student achievement at the elementary level or in higher education (Lee et al., 2018; Reese, 2015; Thompson, 2009; Weinhold, 2015; Welsh et al., 2013). Buttrey (2014) recommended that more studies on standards-based grading at the high school level were needed. Other researchers have conducted studies at the high school level and compared state standardized test scores. ACT is commonly used by post-secondary institutions to show academic achievement and college readiness. This study varies from others because the effects of standards-based grading compared to a traditional grading system and the effect on students' academic achievement on the ACT were analyzed to determine if SBG plays a role in students' post-secondary readiness. If SBG is found to have a significant, positive impact on student achievement on the ACT, this allows for improved post-secondary opportunities for students taught using SBG. This study contributes to the body of literature because SBG is more often used and studied at the elementary level, while this research analyzes the effects of SBG at the high school level. Because of this, other districts, including the district in this study, that are considering the implementation of SBG at the high school level can use this research to make a more informed decision as to if it could positively impact student academic achievement.

Delimitations

Delimitations are perimeters, confines, and choices set by the researcher for purpose of the study (Lunenburg & Irby, 2008). The delimitations set by the researcher for this study are as follows:

1. The study was conducted in a suburban school district in the Midwest with over 20,000 students, of whom approximately 6,000 are high school students.

2. This study is limited to 11th-grade students enrolled in one of the district's high schools who took the ACT for the first time.
3. Data for this study was collected from the 2018-2019 and 2021-2022 school years to lessen the effect of the COVID-19 pandemic.
4. Composite ACT scores and subscores in reading, math, English, and science were analyzed.

Assumptions

Assumptions are generally accepted as true for the purpose of research (Lunenburg & Irby, 2008). This study was mindful of the following assumptions:

1. Handling of the ACT materials by test proctors was completed ethically.
2. The administration of the ACT was completed in a standardized manner.
3. The content and standards tested were comparable from year to year.
4. Students gave their best effort on the ACT at the time the test was administered.

Research Questions

In this study, the researcher examined the following questions in an effort to determine the effects of standards-based grading compared to a traditional grading system on students' academic achievement on the ACT to determine if SBG plays a role in their post-secondary readiness.

RQ1

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT composite scores for students taught using a traditional grading scale (school year

2018-2019) and students taught using a standards-based grading scale (school year 2021-2022)?

RQ2

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT composite scores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

RQ3

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT English subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022)?

RQ4

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT English subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

RQ5

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT math subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022)?

RQ6

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT math subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

RQ7

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT reading subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022)?

RQ8

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT reading subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

RQ9

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT science subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022)?

RQ10

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT science subscores for students taught using a traditional grading scale (school year 2018-

2019) and students taught using a standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

Definition of Terms

To avoid confusion, it is important to define key terms central to the study (Lunenburg & Irby, 2008). Key terms were chosen based on the research questions and hypothesis of the study.

Ethnicity

For this study, ethnicity is defined as a population group of people who share a common cultural background, including language, heritage, religion, or customs (Washington University, 2019.).

Standards-Based Grading

For this study, standards-based grading is defined as a system of reporting student progress toward defined standards. (Heflebower et al., 2014).

Traditional Grading

Bouchrika (2021) defined a traditional grading system as one where students are awarded points for quizzes, tests, projects, or homework. These scores are recorded in a gradebook and then averaged to determine a student's overall percentage. This percentage is then translated to a letter grade representing their passing or failure (Bouchrika, 2021).

Organization of the Study

This study is organized into five chapters. Chapter 1 included the background, statement of the problem, purpose of the study, significance of the study, delimitations, assumptions, research questions, and the definition of terms. Chapter 2 is a review of

literature pertaining to the history of grades, legislation, traditional grading practices, standards-based practices, and standards-based versus traditional and their effect on student achievement. Chapter 3 includes a description of the methods used to conduct the study. In Chapter 4, the results are presented. Chapter 5 provides a study summary, findings related to the literature, and the conclusions.

Chapter 2

Review of the Literature

Grading is the process in education in which teachers use formative and summative assessments to place a value on a student's performance. Grades are used as a means of communication between the teacher and the student and the parent about the student's performance on a particular task (Muñoz & Guskey, 2015). Standards-based grading aims to add more meaning to a student's grade by making grades a better representation of what a student knows or does not know (Heflebower et al., 2014). Pattison et al. (as cited in Brookhart et al., 2016) stated that grades play a large role in determining a student's chance of finishing high school as well as their post-secondary path. ACT is commonly used by post-secondary institutions to see academic achievement. The following topics are included in this chapter: the history of grades, legislation leading to the standards-based movement, traditional grading practices, standards-based grading practices, Traditional practices versus SBG practices, and differences between traditional and SBG practices on student achievement

History of Grades

Grading is a fundamental element of education, but the practices of grading have evolved over time. Grades were first seen at the college level at Yale in 1785 when 58 students were administered an examination (Durm, 1993). William and Mary University followed a similar grading system as Yale in 1817 (Lee, 2020). While schools were implementing the beginnings of the grading system, they were not always sharing grades with the students. Schinske and Tanner (2014) purported that not showing grades, might discourage competition and take the focus from learning. In 1846, Horace Mann wrote,

"if superior rank at recitation be the object, then, as soon as the superiority is obtained, the spring of desire and effort for that occasion relaxes" (as cited in Mann, 1872, p. 504). He continued that students may begin focusing on exams "as to incur moral hazards and delinquencies" (as cited in Mann, 1846, pp. 504-505). The purpose of the 100-point system was to advance students in grade levels (Spencer, 2012). In the early 1900s, education was becoming compulsory for children from kindergarten through high school. With this expansion, the need for a consistent unified approach became apparent (Lee, 2020). In 1913, Finkelstein wrote about how uncalibrated the traditional 100-point grading system was, yet in 2021, American schools continue to use it. Even though the 100-point system is not new, the use of letter grades is relatively new. It was not until the 1940s that letter grades were commonly used (Lee, 2020). Only 67% of primary and secondary schools used letter grades as of 1971 (Lee, 2020). Eventually, the 100-point system was combined with the 4-point scale giving birth to the current traditional grading system (Schneider & Hutt, 2013). Table 2 shows an example of a conversion between the 100-point scale, letter grades, and standard grade point average (GPA) on the 4-point scale.

Table 2

Unweighted GPA Conversion Chart

Letter Grade	%	Standard GPA
A	90-100	4.0
B	80-89	3.0
C	70-79	2.0
D	60-69	1.0
F	0-59	0.0

Note. Adapted from Arches Academy Blog. https://www.archesacademy.com/single-post/2019/06/10/how-to-read-a-report-card

When determining a student's grade, traditionally teachers record and average test scores, quizzes, projects, essays, etc. Some teachers weight their grades, putting more value on any of the aforementioned (Muñoz & Guskey 2015). Advantages of the traditional 100-point grading system include that it is universally recognized, simple for teachers, students, and parents to understand, and allows for comparison to other students (Meador, 2019). This system was founded on the idea of streamlining communication between educational institutions, not to improve or facilitate learning (Lee, 2020).

According to Brookhart and Nitko and Cross and Frary (as cited in Muñoz & Guskey, 2015), the averaging and weighting of assignments often results in grades that are hard to interpret meaningfully. This line of thinking often raises the argument that grades do not always represent what a student knows or can do (Brookhart et al., 2016). Meador (2019) wrote that because the 100-point grading system does not allow for any

reasoning as to how a grade was determined, it is limited in its capacity for consistency. Another issue with the 100-point grading scale is that it is subjective based on classroom decisions, such as whether work is required to be shown and what counts for points (Meader, 2019). An advantage to this system is that it is easy to understand, but that does not mean it is easy for classroom teachers to use. This system can lead to hours of subjective grading for teachers which can lead to simpler assignments to score such as tests or exams (Meador, 2019). In 2015, Munoz and Guskey wrote that the purpose of grades was to represent what students have achieved on specific learning targets. This raises the concern of grades sending an accurate representation of student knowledge to parents.

Legislation

In 1981, T. H. Bell, Secretary of Education, created the National Commission on Excellence in Education (NCEE). This committee was tasked to detail the quality of education in America by the end of 1983 (NCEE, 1983). Their purpose was not only to outline problems with the American education system, but also to outline solutions (NCEE, 1983). In their report, A Nation at Risk, NCEE (1983) stated,

> We report to the American people that while we can take justifiable pride in what our schools and colleges have historically accomplished and contributed to the United States and the well-being of its people, the educational foundations of our society are presently being eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people. (p. 7)

One of their findings related to the cause of this decline in American Education was that there are different expectations for grades that show to what degree content is mastered

(NCEE, 1983). Another finding was that minimum competency exams tended to lower the educational standards for all students (NCEE, 1983). These findings show the cause of the emergence of the standards movement in the United States. As a result of these findings, it was recommended that for grades to be relied on to represent a student's readiness to move on in a specific content, that they should be indicators of a student's academic achievement (NCEE, 1983, p. 24).

As the emphasis on grades being more consistent and reliable grew, so did the emphasis on holding schools accountable for student academic achievement through periodic testing based on outlined standards. The Elementary and Secondary Education Act (ESEA), passed in 1965 by President Lyndon B. Johnson's administration, was written to provide funding for lower-income families and the schools serving them (Edwards, n.d.). In return for government funding, schools must show that they provide a quality education (EdPost Staff, 2015). This is one of the first pieces of legislation that held schools accountable for academic achievement. In 1990, President George H. W. Bush began the push for national goals for schools; however, it was not until 1994, when President Bill Clinton signed the Goals 2000: Educate America Act and the Improving America's Schools Act that the idea began to come to fruition (Edwards, n.d.). This act set standards-based education reform into motion. The purpose of Goals 2000 was to set goals or standards to be met by the year 2000 and required schools to have students make yearly progress towards those goals or lose federal funding (Edwards, n.d.). In 2002, ESEA was revamped and renamed the No Child Left Behind Act (NCLB) by President George W. Bush (Klein, 2015). NCLB expanded federal funding allocated to schools but also laid out standards schools had to meet to receive the funding. Under NCLB all states

were to have all students reach proficiency by 2013-2014 on their state assessments (Klein, 2015). What the act did not define was what proficiency was or dictate the test to use (Klein, 2015). This led to some states creating easier state assessments and lowering their definition of proficiency in order to try to meet the goals set by NCLB (EdPost Staff, 2015).

The actions of decreasing standards and modifying assessment, along with the failure of all states to meet NCLB standards led to the development of the Every Student Succeeds Act (ESSA), which was passed in 2015. ESSA was President Barack Obama's overhaul of NCLB. The difference between the two pieces of legislation is that NCLB only accounted for students who met the standard. It did not give any credit for students that made progress throughout a school year even though they might still not be at the proficient level (EdPost Staff, 2015).

Another by-product of the NCLB Act was the movement toward increased accountability for schools, and the lack of standardization was the production of the Common Core State Standards (CCSS). In 2009, the CCSS resulted from a state-led initiative from the National Governors Association Center that aimed to create clear, consistent expectations and methods of assessment for states (Gewertz, 2015). CCSS are not directly tied to NCLB; however, NCLB required states to have challenging standards in math, reading, and science (EdPost Staff, 2015). By the end of 2011, 46 states had adopted the CCSS, but due to large amounts of backlash, several had backed out by 2015 (Gewertz, 2015). Even though many states backed out of adopting the CCSS, the standards movement in the United States was in motion.

Even with states backing out of the adoption of the CCSS, schools were not backing out of standards-based education. State assessments were still written around standards. Schools and students are annually given assessment data and report cards based on their performance on each standard. In the interim, schools, to help communicate to parents, guardians, and students their progress or achievement on standards, have been on their own to develop a report card that shows this by standard (Swan et al., 2014).

Traditional Practices

In traditional grading systems, students are awarded points for quizzes, tests, projects, or homework. These scores go in a gradebook and are then averaged to determine a student's overall percentage. This percentage is then translated to a letter grade representing their passing or failure (Bouchrika, 2021). This system of grading is usually done in a 100-point system. The purpose of the 100-point system was to advance students in grade levels (Spencer, 2012). At upper-grade levels, the 100-point system is combined with the 4-point scale in the current traditional grading system to give students a grade point average (GPA) (Schneider & Hutt, 2013). A GPA represents the average score a student receives at the end of all of their courses combined (Moody, 2018). Table 3 shows an example of how GPA is calculated.

Table 3

Sample GPA Calculation

									Sum
Grade Received	A	A	C	F	B	C	A	D	8 courses
4.0 Scale Conversion	4	4	2	0	3	3	4	1	20 points
GPA Calculation	20 points/8 courses = 2.5 GPA								

Note. Adapted from "How Do You Calculate Your GPA? Step-by-Step Instructions," by D. A. Wulick, 2020, Prep Scholar. https://blog.prepscholar.com/how-do-you-calculate-gpa https://blog.prepscholar.com/how-do-you-calculate-gpa

In the example found in Table 3, the GPA is 2.5, which on the conversation chart in Table 2 means the student is, on average, a B or C student. The simplistic nature of the traditional grading system makes it easy for teachers to use and parents, guardians, and students to understand (Meador, 2019). Even with variations, the traditional grading system is recognized almost universally (Meador, 2019). Almost anyone in the United States knows that earning an A is good, while earning an F is an indicator of failure (Meader, 2019). This easy recognition of the traditional grading scale allows it to be a clear method of communication to parents, guardians, and students' successes as determined by the teacher. The traditional system also allows for comparisons among students. For high school students, GPA determines class rank, which could affect college admissions and scholarship eligibility (Moody, 2018). For colleges, this can be an indicator of success at the next level (Moody, 2018). Because of this, grades play a prominent role in determining a student's post-secondary path.

Standards-Based Grading Practices

Standards-based grading aims to add more meaning to a student's grade by making grades better represent what a student knows or does not know. It is a system of reporting student progress about defined standards (Heflebower et al., 2014, pp. 3-4). Spencer (2012) tells the story of a physics teacher who noticed their students were able to attain an A grade without mastering some of the more advanced essential concepts. In this grading system, students had no incentive to aim for high-level concepts. There was no differentiation between basic knowledge and more advanced, only an emphasis on earning a certain number of points. Also, there is no communication about why a grade was received, which can frustrate parents and students (Bouchrika, 2021). Scriffiny (2008) offered reasons for standards-based grading. One reason is that grades should have meaning and provide feedback as to where a student is at in progressing towards mastery of a standard. Timely, quality feedback has one of the strongest influences on student success (Knight & Cooper, 2019). On Hattie's ranking of 252 influences and effect sizes related to student achievement, feedback ranks 30th. Standards-based grading effectively communicates and gives feedback on a student's progress based on clearly defined criteria (Heflebower et al., 2014). This clear feedback allows students to improve their performance in certain areas (Heflebower et al., 2014).

Tierney et al. (2011) synthesized multiple sources and determined essential principles for SBG. The first is, "When the purpose of grading is to report on student achievement, grades should be referenced to the curriculum objectives or learning expectations (criterion referenced)" (Tierney et al., 2011, p. 212). In SBG, students receive feedback solely on the learning targets and their progress towards mastery of

them. Therefore, the first step in implementing SBG is to identify what students need to know and receive feedback to succeed (Heflebower et al., 2014). The criteria that can be used to determine priority standards were described by Heflebower et al. (2014). One of the criteria is assessment, including state assessments, which further supports that the standards movement resulted from legislation like *A Nation at Risk* that led to current state reporting systems and created the urgency for schools to change theirs.

Tierney et al. (2011) states that another principle for SBG is "A grade should be an accurate representation of achievement, so non-achievement factors should be reported separately to permit valid interpretation by stakeholders" (p. 212). Because of the focus on academic achievement, non-academic factors are not included and are communicated separately (Tierney et al., 2011). This separation also helps to remove any bias that may be inadvertently in a teacher's grading system. Guskey (2006) stated,

> 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)

SBG aims to remove subjectivity, effort rewarding, and behavior from the equation and focus on grading based more on performance. Grades determined on performance more closely resemble that of a real-world professional job evaluation (Scriffiny, 2008). This replication of real-world evaluations will also help better prepare students for future employment (Reese, 2015).

Another part of SBG is creating assessments of all types that match the standards and allow students various methods and attempts to show mastery (Heflebower et al., 2014). Tierney et al. (2011) state that the third principle indicates that to summarize student achievement accurately, results from multiple assessments must be carefully combined with weighting that reflects the learning. Heflebower et al. (2014) would add that not only do students need multiple assessments, but they also need multiple opportunities on assessments to show mastery.

The final principle for SBG from Tierney et al. (2011) is that clear communication with parents, students, and teachers is necessary to create clarity around the system and meaning of the grades. All of the other principles lose their meaning if this is not done. As stated earlier, the purpose of SBG is to add meaning to grades by representing what is learned and what is still a work in progress. If this is not communicated to parents, guardians, and students, the meaning will not exist.

A common method of schools communicating with parents is through their report cards. As schools make the shift to standards-based grading, using a standards-based report card is usually toward the end of implementation (Heflebower et al., 2014). At first, many schools still live in a percentage system. Heflebower et al. (2014) outlined a process to shift to standards-based grading, beginning with curriculum and communication. Included in this phase is the identification of priority standards and the development of proficiency scales. Proficiency scales are the center of a standards-based system and show a progression of knowledge that is then translated to a numeric score (Heflebower et al., 2014). Table 4 shows a generic proficiency scale.

Table 4

Generic Proficiency Scale

Score	Description
4.0	Complex content – a performance beyond what a standard requires
3.5	In addition to score 3.0 performance, partial success with score 4.0 content
3.0	Target content – the level of learning required for all students
2.5	No major errors or omissions regarding score 2.0 content and partial success with score 3.0 content
2.0	Simple content – basic knowledge or skill necessary for mastering the target content
1.5	Partial success with score 2.0 content and major errors or omissions regarding score 3.0 content
1.0	With help, partial success with score 2.0 content and score 3.0 content
0.5	With help, partial success with score 2.0 content but not with score 3.0 content
0.0	Even with help, no success

Note. Adapted from "A school leader's guide to standards-based grading" by Heflebower et al., 2021, p. 29.

Table 5 below shows how Marzano (2010) recommended translating scores from the 4.0 system to the traditional percentage or letter grade system.

Table 5

Conversion to Letter Grade from 4-point Scale

4.0 Scale Score	Traditional Letter Grade
3.75-4.00	A+
3.26-3.74	A
3.00-3.25	A-
2.84-2.99	B+
2.67-2.83	B
2.5-2.66	B-
2.34-2.49	C+
2.17-2.33	C
2.00-2.16	C-
1.76-1.99	D+
1.26-1.75	D
1.00-1.25	D-
Below 1.00	F

Note. Adapted from *Formative Assessment & Standards Based Grading*, by R. Marzano, 2010 (p. 106).

The conversion in Table 5 helps bridge the gap between traditional grading practices and a standards-based scale.

Traditional Practices vs. SBG Practices

The traditional grading system first appeared at Yale in 1785 (Durm, 1993). This longevity has proven to give the system many benefits. One of the benefits of the

traditional grading system is its simplicity. This simplicity makes it easier for teachers to use and students and parents to understand (Hobden, 2019). It also makes it a very recognizable scale. Few in the United States are unfamiliar with letter grades and what they represent (Hobden, 2019). The familiarity with the scale also makes it easy to use to make comparisons between students (Moody, 2018).

While the traditional grading system is easily understood and universally known, it also has drawbacks. In the traditional grading system, it is not clear what a student knows and is able to do because there are no clear criteria on which they are graded (Mink, 2015). It is limited in this regard because it lacks an explanation of where a student struggles and where their successes are (Meador, 2019). Other factors like extra credit, behaviors, or grade inflation can also play a role in traditional grading practices. The traditional system has inconsistencies because of its subjective nature. Meader (2019) gives the example of two different math classes. One teacher requires work to be shown, while the other only requires answers. A student in the first teacher's math class could be earning a C, but that same student in a different teacher's math class could be making an A even though the work is identical, making it hard to compare (Hobden, 2019).

While its simple nature makes it easy for parents, students, and teachers to understand, it is still time-consuming for teachers. Meador (2019) stated,

> The traditional grading scale leads to hours of subjective grading and fosters a testing culture. While it may be simple for teachers to understand, it takes a lot of time to create and grade the assessments that drive the traditional grading system ("Cons of Traditional Grading Scale" section).

Furthermore, Meador stated that because assessments in the traditional system are easier to grade than those in a standards-based system, it promotes that culture of testing even more.

Many of the advantages of the traditional grading system are the disadvantages of the standards-based system and vice versa. One of the advantages of SBG is that it promotes a growth mindset in students by allowing multiple attempts to master content without the fear of a penalty (Khan, 2019). Conversely, one of the drawbacks of standards-based grading is that because students have multiple attempts, they often do not give their best effort on their first attempt (Khan, 2019). One of the advantages of the traditional system was the simplicity and easy recognition for parents and students (Hobden, 2019). While SBG may not be as recognizable, it aims to bring clarity and consistency in communication to parents and students. In SBG, parents and students are able to identify where a student is at in their progression of learning and what standards they need to improve upon, thus making students focus on mastering topics instead of achieving a specific grade (Khan, 2019). Iamarino (2014) stated that SBG is a way of providing feedback with the expectation of improvement, while the traditional grading system is a way of judging what a student has already achieved. This ability to give clearer feedback does not come without a cost. Bouchrika (2021) stated that in a standards-based system, all lessons, activities, and assessments must be tightly aligned with standards to give clear feedback. This tight alignment could mean teachers would need to enter into the time-consuming process of overhauling their previous materials to ensure they have both formative and summative assessments as well as multiple ways for students to show mastery (Bouchrika, 2021).

Differences Between Traditional and SBG Practices on Student Achievement

While one of the advantages of the traditional system is its recognition and ease for patrons to understand, the standards-based system could be anything but that. Parents and students may be uncomfortable with standards-based grading and question its effect on student achievement (Townsley, 2019). Various studies have been conducted to show if there is an effect on academic achievement.

Kemp (2007) conducted a quasi-experimental study in Mississippi with 65 seventh-grade mathematics classes to determine if students taught in a standards-based or a traditional system would show higher academic achievement in mathematics as measured by the PLATO eduTest. Pre- and post-tests were administered to 65 students. Kemp found no statistically significant difference in student achievement based on the instructional method. However, students in the standards-based system showed a higher mean score and greater gains from pre- to post-test.

Haptonstall (2010) conducted a quantitative study with five school districts in the state of Colorado. The purpose of the study was to determine if there was a correlation between grades and student achievement measured by Colorado Student Assessment Program (CSAP) tests in reading, writing, math, and science for sixth through 10th-grade students. Haptonstall determined that there was a statistically significant correlation between grades and CSAP scores for all schools. However, the school using the standards-based model showed higher correlations and higher mean scores for all grades and subgroups.

Hardegree (2012) conducted a non-experimental causal-comparative study in rural Georgia with approximately 550 fifth-grade students from eight elementary schools.

The purpose of the study was to determine if standards-based grades were a predictor of student achievement as measured by the Criterion Referenced Competency Test (CRCT) in mathematics and reading. The results of the study indicated a significant correlation between students' standards-based grades and their scores on the CRCT. Hardegree found that students on free or reduced lunch status tended to score lower in mathematics and reading. Also, in reading, those with limited English proficiency scored lower. When looking at students with the same grades, females tended to score higher on the CRCT than their male peers.

Sieling (2013) conducted a quantitative study comparing math scores on the Minnesota Comprehensive Assessment (MCA) of 44 seventh graders, 64 eighth graders, and 41 ninth graders from a rural public school to determine if a correlation exists. Scores were used from 2010-2012. Students from 2010-2011 were taught in a traditional grading system, while students from 2011-2012 were taught in a standards-based system for math. Grades were also collected to determine the predictability of MCA mathematics scores. Sieling (2013) found that student achievement for students from both systems was similar on the MCA mathematics assessment; however, by conducting a Pearson product-moment correlation, Sieling determined that the grades from the standards-based system were a better predictor of student scores.

Buttrey (2014) conducted a mixed-methods study in a rural Kentucky district with 674 fourth- and fifth-grade students' math and reading/language arts scores as measured by the K-PREP state assessment. The district in the study was comprised of six elementary schools and reported 12% of families living below the poverty line. Five elementary schools used a traditional grading system, and one used a standards-based

system. The goal of the study was to determine if there was a difference in students' scores between those graded using a traditional system and a standards-based system. The results of the study showed a significant increase in the student's mean scores in math for the students from the standards-based system. Reading/language arts mean scores were also raised but not significantly. An analysis was conducted to determine if there was a connection between grading system type and K-Prep math and reading/language arts scores. Buttrey (2014) also looked at the socioeconomic status of students to determine if there was a correlation between grading and academic achievement. All groups except the low socioeconomic status standards-based math group showed a significant positive correlation. Those students coming from a traditional grading system showed a slightly higher correlation except for the low socioeconomic reading/language arts scores on the K-PREP assessment.

Norton (2014) also conducted a mixed methods study comparing fourth and fifth grade students from a standards-based system to K-PREP state assessment scores to determine if there was a correlation. Like Buttrey (2014), Norton's (2014) study was conducted in a rural Kentucky district with six elementary schools, one of which had implemented standards-based grading. Norton's findings were consistent with Buttrey's and showed that scores for students in math were significantly higher for students attending a school with a standards-based system compared to a traditional system but that there was no significant difference in reading/language arts scores. Because of the significant increase in math scores, Norton (2014) recommended that the district continue with standards-based grading.

Yoakum (2014) conducted a quantitative study at a Midwest high school of 819 students from 2008 to 2013, omitting the data from the 2010-2011 school year because that was the transitional year for standards-based grading. The researcher studied communication arts and mathematics scores as measured by the Missouri Assessment Program End of Course Exams (EOC) to determine if they were affected by the implementation of standards-based grading at the school. Yoakum compared the difference in communication arts and mathematics EOC scores before and after the implementation of SBG. Yoakum found a statistical difference in communications arts scores; they declined after the implementation. The *t*-test results used to compare mathematics EOC scores before and after the implementation showed that there was also a statistical difference in Math EOC scores. Conversely to communication arts scores, mathematics scores inclined after the implementation of SBG instead of declining.

Tyree-Hamby (2015) conducted a correlational study in a rural Missouri elementary school to determine if there was a relationship between traditional grading methods and student achievement or standards-based grading and student achievement. Student achievement was measured by the MAP assessment for mathematics and reading. Tyree-Hamby compared 120 third graders in 2012-2013 that transitioned to fourth graders in 2013-2014, which was the first year of implementing standards-based grading at this elementary school. Tyree-Hamby found a statistically significant positive correlation between teacher-assigned standards-based grades and teacher-assigned traditional grades and English scores on the MAP assessment. A stronger correlation was found between teacher-assigned standards-based grades and English scores on the MAP

assessment. The same results were shown for the traditional and standards-based grades and the mathematics scores on the MAP assessment.

Graves (2016) conducted a mixed-methods study to identify high schools in Missouri that have fully implemented standards-based grading to determine the effect on student achievement as measured by state assessments. Included in this study were eight schools of varying sizes from the state, seven public schools and one charter school. Graves found that four of the eight schools improved English scores after implementation. The other four schools showed no noticeable increasing trend. In algebra 1, five schools showed an increasing trend in test scores, 2 showed a declining trend, and one did not have enough data points to analyze a trend.

McCarthy and Sharp (2016) conducted a study to determine if there was a relationship between grades and scores on Missouri state science assessments at the end of Grade 5, Grade 8, and after biology 1. Furthermore, to determine if there was a difference in science scores, students taught in a traditional grading system were compared to students taught in a standards-based system. This study, conducted in Northwestern Missouri, included 294 students who were juniors during the 2016-2017 school year. Students were divided into two groups based on the middle school they attended. One middle school used the traditional grading system, while the other implemented standards-based grading in 2012. Historical state assessment data was collected for the group of students from Grade 5, Grade 8, and after biology 1. McCarthy and Sharp found a relationship between grades and state science assessment scores and a positive, statistically significant difference in scores for students taught in a standards-based system.

Poll (2019) conducted a correlational study to determine if academic achievement was impacted by being in a standards-based setting versus a traditional grading setting for secondary students in mathematics, science and language arts as measured by end-of-level students in academically gifted education (SAGE) test scores. End-of-term grades were also used to determine if there was a correlation between grades, as measured by student GPA, and end-of-level SAGE test scores in a standards-based setting. For this study, 45 teachers were using standards-based grading, and 45 teachers were using traditional grading, ranging from Grade 7 to Grade 12, from 24 different schools, eight high schools and 16 junior high schools, all in the western United States. Students in the standards-based setting achieved a higher mean GPA than their peers from traditional settings. Poll noted that while there is a statistical difference, it translates to only half a grade higher than their peers of traditional grading. However, results provided evidence of a statistical difference between the average SAGE scaled score and students in a standards-based setting compared to those from a traditional grading setting. Students from a standards-based setting were shown to outscore their traditionally graded peers by 0.293 standard deviations. Additionally, Poll found that neither system was better than the other. Scores for the standards-based students were under-predicted by 0.107, and scores for the traditionally graded students were over-predicted by 0.109. When looking at mathematics, science, and language arts, all three contents showed a statistically significant difference in grades of those in each grading system compared to content-specific SAGE scale scores. Grades from the traditional setting over-estimated SAGE scales in all three contents. In contrast, grades from standards-based settings under-estimated SAGE scale scores in mathematics and English, while science was over-

estimated. English was shown to have the closest correlation, while mathematics had the lowest correlation.

Rainey (2016) conducted a quantitative study using an explanatory research design at one elementary school in a large suburban district in north Texas. The purpose of the study was to determine if there was a correlation between grades and scores on State of Texas Assessments of Academic Readiness (STAAR) in reading. Rainey (2016) also examined the scores of those students with limited English proficiency in a standards-based system compared to students with non-limited English proficiency in a standards-based system. Rainey also compared students from a lower socioeconomic status to those of a higher socioeconomic status in a standards-based system. Rainey used STAAR reading achievement scores as a measurement for academic achievement. Scores for 218 third-grade students from 2013-2015 were used. Rainey concluded that there was a statistically significant correlation between the grades of third-grade students in a standards-based system as measured by the STAAR assessment in reading. Rainey also found a strong correlation between English proficiency and reading scores on STAAR in the standards-based system. Finally, Rainey's results showed a statistically significant relationship between socioeconomic status and STAAR reading scores in a standards-based system.

Townsley (2017) conducted a quasi-experimental study to determine if a high school's grading system, standards-based or traditional, affected ACT scores, specifically mathematics and English, and GPA. Data were collected at two comparable midwestern high schools, one using each type of grading. There were 327 students involved in the study, 159 that experienced traditional grading systems and 168 that experienced the

standards-based system over two school years, 2015 and 2016 graduating classes. Townsley's results indicated that math GPAs, English GPAs, and cumulative GPAs were not significantly different in either grading system, which indicates that the grading system did not impact student GPAs. There was a statistically significant difference in mean mathematics, English, and composite ACTs scores, with students in the standards-based system scoring lower than those in the traditional grading system in all three areas. This finding suggests that high school students in the traditional grading system outscored students in the standards-based system. This study also showed no difference in using GPA as a predictor of ACT scores for traditional or standards-based systems.

Decker et al. (2018) conducted a mixed-methods study at a private school in Tennessee using middle school English and mathematics mid-terms and high school advanced placement (AP) psychology scores to determine the impact of SBG. The researchers compared a group of students over a two-year period, 2016-2017 fifth-grade students and 2017-2018 sixth-grade students. Decker et al. analyzed mathematics and English midterm scores each year. The researchers found that in mathematics, after three semesters of standards-based grading, the mean scores on midterms decreased; however, the difference was not significant when analyzed by an independent-samples *t*-test. However, when the data was further analyzed using the Wilcoxon signed-rank test, a significant difference was found. For the comparison of AP psychology scores, two different groups of student scores were compared, scores from 2016 and 2017. Scores from 2016 were from students in a traditional grading system, and the scores from 2017 were from students in a standards-based system. The results of the data analysis showed

there was no statistical difference between the scores of the students in a standards-based system and a traditional system.

Kelly (2018) sought to identify the relationship standards-based grading had on student achievement at the secondary level. Also, Kelly sought to discover the relationship between standards-based grading and student achievement based on gender, ethnicity, special education, and free and reduced lunch. Finally, Kelly sought to determine the relationship between a standards-based grading system and attendance rates, dropout rates, and ACT scores. This study was conducted in Missouri with 13 middle and eight high schools with varying demographics, enrollment, and socioeconomic status. Rural, suburban, and urban schools were included. Student achievement was measured by grade-level English language arts and mathematics assessments administered through the Missouri Assessment Program (MAP). Kelly found no significant difference in high school English II MAP scores, high school algebra 1 MAP scores, and eighth-grade English language arts MAP scores because of the implementation of standards-based grading. However, there was a significant difference in eighth-grade mathematics MAP scores of White students. The results of the study indicated a decline in the mean of eighth-grade mathematics MAP scores because of the implementation of standards-based grading. Kelly also examined the data by gender, ethnicity, special education, and free and reduced lunch status. The results showed no significant difference in females or males on the English II, algebra 1, or eighth-grade English MAP assessment. However, there was a significant difference for females but not males on the eighth-grade mathematics MAP assessment resulting in a decline in scores. Additionally, no significant difference was found among Black, Hispanic, or White

students on the English 2, algebra 1, and eighth-grade English MAP assessments. On the eighth-grade mathematics MAP assessment, there was no difference in Black or Hispanic students after the implementation of SBG. There was, however, a difference in White students who showed a statistically significant reduction in the mean score on the eighth-grade mathematics MAP assessment. There was no significant difference in the results based on special education status after SBG implementation. The results of the study provided evidence that there was a significant difference in English II MAP assessment scores between students classified as lower socioeconomic status and those students who were not after SBG implementation. However, there was no difference in scores for the algebra 1, eighth-grade English, or eighth-grade math MAP assessments when looking at students affected by low SES. Kelly found no significant difference in attendance or ACT scores, but there was a significant decrease in dropout rates after implementation.

Bosanec (2020) conducted a quantitative study to evaluate the impact of standards-based grading in English, mathematics, and science on predicting ACT scores compared to traditional grading. The midwestern, suburban high school switched to SBG in 2015. An impact program evaluation was used in this study. Participants in the study completed six consecutive semesters at the school without disruption and took the ACT exam by the end of their sixth semester. The results of the study indicated a significant difference in grades as measured by GPA in all three content areas. Students in the standards-based system had higher GPAs than students in the traditional grading system. ACT scores were only significantly different in mathematics and science but not English. For mathematics and science, it was shown that ACT scores were lower in the standards-

based system. Regardless of the system, there was a positive linear relationship meaning the lower the GPA, the lower the ACT score in all three contents.

d'Erizans (2020) conducted a correlational study comparing student achievement as measured by Northwest Evaluation Association's Measures for Academic Growth (NWEA-MAP) scores in language arts and mathematics and grades of middle school students in both a traditional system and a standards-based system. The goal was to determine if grades were a predictor of a student's academic achievement on the NWEA-MAP. This study was conducted at an American school in the Middle East that switched from traditional grading to standards-based in 2013. Data was used from 2009-2019. To compare mathematics scores, 3,481 data points were used, and to compare language arts scores, 3,343 data points were used. In mathematics and language arts, a statistically significant difference was found between scores and predictability of students in a standards-based system and those in a traditional system with higher scores resulting from the standards-based system.

Hargrove (2020) conducted a quantitative quasi-experimental study to determine if there was a difference in reading growth for 213 second- and third-grade students between those taught in a traditional system and those taught in a standards-based system. All students in the study had Measures of Academic Progress scores from the beginning and middle of the year. Second-grade students were taught using a standards-based system, while third grade students were taught in a traditional system. Hargrove found a statistically significant difference between the mean scores of those taught traditionally and those taught in a standards-based system and that those students in the standards-based system showed less growth than their traditionally taught peers.

Summary

Teachers using standards-based grading aim to add more meaning to a student's grade by making grades a better representation of what a student knows or does not know. Previous studies show mixed results and that further studies are needed. Grades play a large role in determining a student's post-secondary path. ACT is a commonly used by post-secondary institutions to see academic achievement. Therefore, this study examined the effects of standards-based grading compared to a traditional grading system and the effect on students' academic achievement on the ACT, including reading, mathematics, English, science, and composite scores. The next chapter includes the research design, data collection procedures, data analysis and hypothesis testing, and limitations used by the researcher.

Chapter 3

Methods

The first purpose of the current study was to determine the extent there is a difference in ACT composite, English, math, reading, and science scores of students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022). The second purpose of this study was to determine the extent the difference in ACT composite, English, math, reading, and science scores of first-time 11th-grade ACT test takers taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022) are affected by student gender and ethnicity. Included in this chapter is the methodology used to test the research questions. Chapter 3 provides a description of the research design, selection of participants, measurement, data collection, data analysis and hypothesis testing, and limitations.

Research Design

A causal-comparative design using archived data was utilized for this quantitative study. This design was used because the data analysis compared two or more experimental groups after a cause had already been implemented (Creswell & Creswell, 2018). The dependent variables were the composite scores and the subscores (English, mathematics, reading, and science) for first-time 11th-grade ACT test takers from 2018-2019 and 2021-2022. The independent variables were student gender and ethnicity.

Selection of Participants

The participants of this study were 11th-grade students who took the ACT for the first time in a midwestern suburban school district. A purposive sampling procedure was used to select 11th grade students enrolled at four high schools in District A. Lunenburg and Irby (2008) defined purposive sampling as "selecting a sample based on the researcher's experience or knowledge of the group to be sampled" (p. 175). A student's data was included in this if the following criteria were met:

1. The student attended District A during the 2018-2019 or 2021-2022 school years.
2. The student was enrolled in Grade 11.
3. The student was taking the ACT for the first time during the spring of each school year.
4. The student received a valid composite score on the ACT as well as subscores in English, mathematics, reading, and science.

Measurement

Student achievement was measured using composite scores and the English, mathematics, reading, and science subscores on the ACT college entrance exam. The ACT is scored on a scale of one to 36. A student earns a scaled score on each sub-section that is then averaged to obtain their composite score. The scale score is determined using a conversion from the student's raw score or number of questions they answered correctly (ACT, 2020).

The ACT is a timed assessment to be completed in 2 hours and 55 minutes

without breaktime and contains four sections. The first section is English. This section is 45 minutes long, contains 75 multiple-choice questions, and tests production of writing, knowledge of language, and conventions of standard English (ACT, 2020). Section two is mathematics. This section covers number and quantity, algebra, functions, geometry, statistics and probability, modeling, and integrating essential skills over 60 multiple-choice questions in 60 minutes. The third section is reading. It is 40 multiple-choice questions in 35 minutes. It is comprised of four passages that test key ideas and details, craft and structure, and integration of knowledge and ideas. The final section is science. This section is comprised of 40 multiple-choice questions to be completed in 35 minutes. Topics tested include interpretation of data, scientific investigation, and evaluation of models, inferences, and experimental results (ACT, 2020). The breakdown of sections of the ACT with test lengths and the number of questions can be found in Table 6.

Table 6

Breakdown of ACT College Entrance Exam

Section	Breakdown of Assessed Content (percent of test over standard)
English	Production of writing (29-32%) Knowledge of language (13-19%) Conventions of standard English (51-56%)
Mathematics	Preparing for higher math (57-60%) • Number & Quantity (7-10%) • Algebra (12-15%) • Functions (12-15%) • Geometry (12-15%) • Statistics & Probability (8–12%) Integrating essential skills (40-43%) Modeling (≥ 27%)
Reading	Key ideas and details (55-60%) Craft and structure (25-30%) Integration of knowledge and ideas (13-18%)
Science	Interpretation of data (45-55%) Scientific investigation (20-30%) Evaluation of Models, inferences, and experimental results (25-35%)

Note. Adapted from *ACT Technical Manual*, by ACT, 2020, pp. 3.3, 3.9, 3.11, 3.14. https://success.act.org/s/article/The-ACT-Technical-Manual

Lunenburg and Irby (2008) stated that validity is "the degree to which an instrument measures what it purports to measure" (p. 181). To ensure test validity, ACT developed College and Career Readiness Standards. These standards were written by highly qualified subject-matter experts in each area. These teams of experts reviewed normative data, college admissions criteria, and information obtained through ACT's

course placement service. Standards were written “based on their analysis of the knowledge and skills students need to respond successfully to test items that were answered correctly by 80% or more of the examinees who scored within each score range” (ACT, 2020, p. 8.3).

Content validity refers to “the degree to which an instrument measures an intended content area” (Lunenburg & Irby, 2008, p. 181). To ensure content validity, ACT asked nationally recognized scholars in English, math, reading, and science, as well as high school and university education departments, to review the College and Career Readiness Standards (ACT, 2020). Teams of six, comprised of three college curriculum and instruction instructors as well as a classroom teacher from eighth, tenth, and twelfth grades, were formed (ACT, 2020). These teams were asked to determine if the College and Career Readiness Standards “(a) accurately reflected the skills and knowledge needed to correctly respond to test items (in specific score ranges) on the ACT and (b) represented a continuum of increasingly sophisticated skills and understandings across the score ranges” (ACT, 2020, p. 8.7). Teams were given a complete set of standards as well as a random sampling of 17 test items from each score range per test section. To obtain the 17 test items, 85 observations were needed in the five score ranges. The target criterion for agreeance amongst the teams was 70% and the standard error could be no greater than 0.05 for the accuracy in the percentage of matches.

Reliability is “the consistency or repeatability of an instrument” (Creswell & Creswell, 2018, p. 154). Inter-item covariances were used by ACT to estimate the reliability coefficient or internal consistency reliability (ACT, 2020). Cronbach’s alpha was used to determine estimates for the reliability of correct raw scores while r_t was used

to estimate the reliability of scale scores. To find r_t the following formula was used where SEM_t standard error of measurement for the estimated scale score and s_t^2 is the sample variance of the observed scale score.

$$r_t = 1 - \frac{SEM_t^2}{s_t^2}$$

ACT used operational data to calculate the alphas that provided evidence of reliability for composite scores and subscores. Data from seven test forms was gathered. Table 7, adapted from ACT technical manual, contains the scale score median, minimum, and maximum reliability estimates across the test forms. Values closer to one show a higher consistency. The level of the scale score reliability coefficients listed below is strong evidence of internal consistency reliability.

Table 7

Summary Statistics for Scale Score Reliability

	Reliability		
Test	Median	Minimum	Maximum
English	0.93	0.92	0.94
Mathematics	0.92	0.91	0.93
Reading	0.87	0.85	0.89
Science	0.85	0.84	0.88
Composite	0.97	0.97	0.97

Note. Adapted from *ACT Technical Manual*, by ACT, 2020, p. 10.2. https://success.act.org/s/article/The-ACT-Technical-Manual. Sample sizes were not included because ACT did not provide them with the reliability information.

Data Collection Procedures

Oral consent was given by the director of data and accountability for this study to be conducted in July 2022 with the condition of having the study approved by Baker University's Institutional Review Board (IRB). On July 25, 2022, the written consent form was signed by the director of data and accountability in District A for this study to be conducted with the condition of having the study approved by Baker University's IRB (see Appendix B). On August 27, 2022, a request for permission to conduct the study was submitted to Baker University's IRB committee, which was approved on August 30, 2022 (see Appendix C). Upon approval by the IRB, the director of data and accountability collected archival test score data and student demographic data. Several Excel worksheets were sent to the researcher. The data was coded to ensure the anonymity of the students. The worksheets were merged into one file and imported into IBM SPSS Statistics Grad Pack Base 28 for Mac for data analysis.

Data Analysis and Hypothesis Testing

ACT assessment data was analyzed to address the research questions in this study. A two-factor analysis of variance (ANOVA) was conducted to test each hypothesis. Research questions, hypotheses, and data analysis follow below.

RQ1

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT composite scores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022)?

H1. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT composite scores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022).

A two-factor Analysis of Variance (ANOVA) was conducted to test H1 and H2. The two categorical variables used to group the dependent variable, ACT composite scores, were grading scale type (traditional, standards-based) and gender (male, female). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical variable among three or more groups, including a main effect for grading scale type, a main effect for gender, and a two-way interaction effect (Grading Scale Type x Gender). The main effect for grading scale type was used to test H1. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

RQ2

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT composite scores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

H2. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT composite scores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender.

The interaction effect (Grading Scale Type x Gender) from the first ANOVA was used to test H2. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

H3. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT composite scores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student ethnicity.

A second two-factor ANOVA was conducted to test H3. The two categorical variables used to group the dependent variable, ACT composite scores, were grading scale type (traditional, standards-based) and ethnicity (Black, Hispanic, White, Other). The interaction effect (Grading Scale Type x Ethnicity) was used to test H3. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

RQ3

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT English subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022)?

H4. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT English subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022).

A third two-factor ANOVA was conducted to test H4 and H5. The two categorical variables used to group the dependent variable, ACT English subscores, were grading scale type (traditional, standards-based) and gender (male, female). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical among three or more groups, including a main effect for grading scale type, a main effect for gender, and a two-way interaction effect (Grading Scale Type x Gender). The main effect for grading scale type was used to test H4. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

RQ4

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT English subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

H5. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT English subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender.

The interaction effect (Grading Scale Type x Gender) from the third two-factor ANOVA was used to test H5. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

H6. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT English subscores for students taught using a traditional grading scale

(school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student ethnicity.

A fourth two-factor ANOVA was conducted to test H6. The two categorical variables used to group the dependent variable, ACT English subscores, were grading scale type (traditional, standards-based) and ethnicity (Black, Hispanic, White, Other). The interaction effect (Grading Scale Type x Ethnicity) was used to test H6. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

RQ5

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT math subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022)?

H7. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT math subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022).

A fifth two-factor ANOVA was conducted to test H7 and H8. The two categorical variables used to group the dependent variable, ACT math subscores, were grading scale type (traditional, standards-based) and gender (male, female). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical among three or more groups, including a main effect for grading scale type, a main effect for gender, and a two-way interaction effect (Grading Scale Type x Gender). The main

effect for grading scale type was used to test H7. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

RQ6

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT math subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

H8. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT math subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender.

The interaction effect (Grading Scale Type x Gender) from the fifth two-factor ANOVA was used to test H8. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

H9. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT math subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student ethnicity.

A sixth two-factor ANOVA was conducted to test H9. The two categorical variables used to group the dependent variable, ACT math subscores, were grading scale type (traditional, standards-based) and ethnicity (Black, Hispanic, White, Other). The interaction effect (Grading Scale Type x Ethnicity) was used to test H9. The level of

significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

RQ7

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT reading subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022)?

H10. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT reading subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022).

A seventh two-factor ANOVA was conducted to test H10 and H11. The two categorical variables used to group the dependent variable, ACT reading subscores, were grading scale type (traditional, standards-based) and gender (male, female). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical among three or more groups, including a main effect for grading scale type, a main effect for gender, and a two-way interaction effect (Grading Scale Type x Gender). The main effect for grading scale type was used to test H10. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

RQ8

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT reading subscores for students taught using a traditional grading scale (school year 2018-

2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

H11. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT reading subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender.

The interaction effect (Grading Scale Type x Gender) from the seventh two-factor ANOVA was used to test H11. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

H12. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT reading subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student ethnicity.

An eighth two-factor ANOVA was conducted to test H12. The two categorical variables used to group the dependent variable, ACT reading subscores, were grading scale type (traditional, standards-based) and ethnicity (Black, Hispanic, White, Other). The interaction effect (Grading Scale Type x Ethnicity) was used to test H12. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

RQ9

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT science subscores for students taught using a traditional grading scale (school year

2018-2019) and students taught using standards-based grading scale (school year 2021-2022)?

H13. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT science subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022).

A ninth two-factor ANOVA was conducted to test H13 and H14. The two categorical variables used to group the dependent variable, ACT science subscores, were grading scale type (traditional, standards-based) and gender (male, female). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical among three or more groups, including a main effect for grading scale type, a main effect for gender, and a two-way interaction effect (Grading Scale Type x Gender). The main effect for grading scale type was used to test H13. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

RQ10

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT science subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

H14. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT science subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender.

The interaction effect (Grading Scale Type x Gender) from the ninth two-factor ANOVA was used to test H14. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

H15. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT science subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student ethnicity.

A tenth two-factor ANOVA was conducted to test H15. The two categorical variables used to group the dependent variable, ACT science subscores, were grading scale type (traditional, standards-based) and ethnicity (Black, Hispanic, White, Other). The interaction effect (Grading Scale Type x Ethnicity) was used to test H12. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

Limitations

Lunenburg and Irby (2008) stated, "Limitations are factors that may have an effect on the interpretation of the findings or on the generalizability of the results" (p. 133). The researcher does not have control over the limitations, and misapprehensions can be avoided by explicitly stating the limitations (Lunenburg & Irby, 2008). Limitations of this study included:

1. Different high schools in District A have different SBG expectations.
2. Students may have participated in varying amounts of ACT preparation.
3. Student experiences prior to the school years involved in this study are unknown.

4. The impact of the COVID-19 pandemic on student's education experience is unknown.

Summary

The overall purpose of this chapter was to describe the methodology used to determine how first-time 11th-grade ACT test-takers score on the ACT after the implementation of SBG. The participants chosen for this study were 11th-grade students who took the ACT for the first time in a midwestern suburban school district. A purposive sampling procedure was used to select 11th grade students enrolled at four high schools in District X. The validity and reliability of the ACT were presented. The data collection procedures, data analysis and hypothesis testing, and the limitations were detailed. Results of the study are presented in the following chapter.

Chapter 4

Results

The overall purpose of this study was to determine how first-time 11th-grade ACT test-takers score on the ACT. The first purpose was to determine the extent there is a difference in ACT composite, English, math, reading, and science scores of students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022). The second purpose of this study was to determine the extent the difference in ACT composite, English, math, reading, and science scores of first-time 11th-grade ACT test takers taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022) are affected by student gender and ethnicity. To address the purposes of this study, 10 research questions were posed and 15 hypotheses were tested. The results of the hypothesis testing are included in this chapter.

Descriptive Statistics

For this study, 1,095 first time 11th-grade ACT takers' scores were used from the 2018-2019 school year. This number represents the participants taught in a traditional grading system. For the 2021-2022 school year, 885 first time 11th-grade ACT takers' scores were used. This number represents the participants taught in a standards-based grading system. Students who did not have scores for all sections of the ACT were eliminated from both years of data.

The first demographic variable was gender. Gender included females and males. In Table 8, the frequencies for student gender for students in traditional and standards-based grading scales are found.

Table 8

Crosstabulation of Student Gender by Grading Scale

Gender	Grading Scale	
	Traditional (*N*)	Standards-Based (*N*)
Females	557	451
Males	538	434

The second demographic variable was ethnicity. In Table 9, the frequencies for original and recoded student ethnicity categories for students in traditional and standards-based grading scales are found. The original data include seven groups: American Indian/Alaska Native, Asian, Black, Hispanic, Multi-racial, Pacific Islander, and White. Students who were Asian, American Indian/Alaska Native, Multi-racial, and Pacific Islander were recoded as Other.

Table 9

Crosstabulation of Original and Recoded Ethnicity Categories by Grading Scale

Race	Grading Scale	
	Traditional (*N*)	Standards-Based (*N*)
Original		
Asian	56	38
Black	179	171
Hispanic	178	135
American Indian/Alaska Native	5	1
Multi-racial	110	92
Pacific Islander	10	14
White	557	434
Recoded		
Black	179	171
Hispanic	178	135
White	557	434
Other	181	145

Hypothesis Testing

To address the purposes of this study, 10 research questions were posed and 15 hypotheses were tested. The results of these tests are contained in this section. In this section, each research question is addressed individually. Each research question is

followed by the corresponding hypotheses for that question. Following each hypothesis are the corresponding analysis and results for each.

RQ1

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT composite scores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022)?

H1. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT composite scores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022).

A two-factor ANOVA was conducted to test H1 and H2. The two categorical variables used to group the dependent variable, ACT composite scores, were grading scale type (traditional, standards-based) and gender (male, female). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical among three or more groups, including a main effect for grading scale type, a main effect for gender, and a two-way interaction effect (grading scale type x gender). The main effect for grading scale type was used to test H1. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated a statistically significant difference between the means, $F(1, 1976) = 49.516$, $p = .000$, $\eta^2 = .024$. See Table 10 for the means and standard deviations for this analysis. The traditional grading scale mean ($M = 18.80$) was

higher than the standards-based grading scale mean (*M* = 17.27). H1 was supported. The effect size indicated a small effect.

Table 10

Descriptive Statistics for the Results of the Test for H1

Grading Scale Type	*M*	*SD*	*N*
Traditional	18.80	4.75	1,095
Standards-based	17.27	4.84	885

RQ2

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT composite scores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

H2. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT composite scores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022) affected by student gender.

The interaction effect (Grading Scale Type x Gender) from the first ANOVA was used to test H2. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated there was not a statistically significant difference between at least two of the means, $F(1, 1976) = 0.081$, $p = .776$. See Table 11

for the means and standard deviations for this analysis. A post hoc was not warranted. H2 was not supported.

Table 11

Descriptive Statistics for the Results of the Test for H2

Grading Scale Type	Gender	*M*	*SD*	*N*
Traditional	Female	18.93	4.53	557
	Male	18.66	4.97	538
Standards-based	Female	17.47	4.67	451
	Male	17.08	5.01	434

H3. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT composite scores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student ethnicity.

A second two-factor ANOVA was conducted to test H3. The two categorical variables used to group the dependent variable, ACT composite scores, were grading scale type (traditional, standards-based) and ethnicity (Black, Hispanic, White, Other). The interaction effect (Grading Scale Type x Ethnicity) was used to test H3. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated there was not a statistically significant difference between at least two of the means, $F(3, 1972) = 0.591$, $p = .621$. See Table 12

for the means and standard deviations for this analysis. A post hoc was not warranted. H3 was not supported.

Table 12

Descriptive Statistics for the Results of the Test for H3

Grading Scale Type	Ethnicity	*M*	*SD*	*N*
Traditional	Black	16.30	3.57	179
	Hispanic	17.53	3.76	178
	White	19.78	5.13	557
	Other	19.48	4.84	181
Standards-based	Black	14.99	3.64	171
	Hispanic	16.38	3.59	135
	White	18.40	5.05	434
	Other	17.44	5.37	145

RQ3

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT English subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022)?

H4. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT English subscores for students taught using a traditional grading scale

(school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022).

A third two-factor ANOVA was conducted to test H4 and H5. The two categorical variables used to group the dependent variable, ACT English subscores, were grading scale type (traditional, standards-based) and gender (male, female). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical among three or more groups, including a main effect for grading scale type, a main effect for gender, and a two-way interaction effect (Grading Scale Type x Gender). The main effect for grading scale type was used to test H4. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated a statistically significant difference between the means, $F(1, 1976) = 42.909$, $p = .000$, $\eta^2 = .021$. See Table 13 for the means and standard deviations for this analysis. The traditional grading scale mean ($M = 17.86$) was higher than the standards-based grading scale mean ($M = 16.14$). H4 was supported. The effect size indicated a small effect.

Table 13

Descriptive Statistics for the Results of the Test for H4

Grading Scale Type	*M*	*SD*	*N*
Traditional	17.86	5.85	1095
Standards-based	16.14	5.83	885

RQ4

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT English subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

H5. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT English subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender.

The interaction effect (Grading Scale Type x Gender) from the third ANOVA was used to test H5. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared,* is reported.

The results of the analysis indicated there was not a statistically significant difference between at least two of the means, $F(1, 1976) = 0.117$, $p = .732$. See Table 14 for the means and standard deviations for this analysis. A post hoc was not warranted. H5 was not supported.

Table 14

Descriptive Statistics for the Results of the Test for H5

Grading Scale Type	Gender	*M*	*SD*	*N*
Traditional	Female	18.33	5.81	557
	Male	17.37	5.86	538
Standards-based	Female	16.70	5.91	451
	Male	15.56	5.70	434

H6. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT English subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student ethnicity.

A fourth two-factor ANOVA was conducted to test H6. The two categorical variables used to group the dependent variable, ACT English subscores, were grading scale type (traditional, standards-based) and ethnicity (Black, Hispanic, White, Other). The interaction effect (Grading Scale Type x Ethnicity) was used to test H6. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared,* is reported.

The results of the analysis indicated there was not a statistically significant difference between at least two of the means, $F(3, 1972) = 0.585$, $p = .625$. See Table 15 for the means and standard deviations for this analysis. A post hoc was not warranted. H6 was not supported.

Table 15

Descriptive Statistics for the Results of the Test for H6

Grading Scale Type	Ethnicity	*M*	*SD*	*N*
Traditional	Black	15.05	4.59	179
	Hispanic	16.23	4.56	178
	White	19.05	5.91	557
	Other	18.59	6.60	181
Standards-based	Black	13.61	4.73	171
	Hispanic	15.00	4.54	135
	White	17.45	6.00	434
	Other	16.26	6.39	145

RQ5

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT math subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022)?

H7. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT mathematics subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022).

A fifth two-factor ANOVA was conducted to test H7 and H8. The two categorical variables used to group the dependent variable, ACT mathematics subscores, were

grading scale type (traditional, standards-based) and gender (male, female). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical among three or more groups, including a main effect for grading scale type, a main effect for gender, and a two-way interaction effect (Grading Scale Type x Gender). The main effect for grading scale type was used to test H7. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated a statistically significant difference between the means, $F(1, 1976) = 42.909$, $p = .000$, $\eta^2 = .032$. See Table 16 for the means and standard deviations for this analysis. The traditional grading scale mean ($M = 18.56$) was higher than the standards-based grading scale mean ($M = 17.00$). H7 was supported. The effect size indicated a small effect.

Table 16

Descriptive Statistics for the Results of the Test for H7

Grading Scale Type	*M*	*SD*	*N*
Traditional	18.56	4.46	1095
Standards-based	17.00	4.01	885

RQ6

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT math subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

H8. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT math subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender.

The interaction effect (Grading Scale Type x Gender) from the fifth ANOVA was used to test H7. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated there was not a statistically significant difference between at least two of the means, $F(1, 1976) = 0.239$, $p = .625$. See Table 17 for the means and standard deviations for this analysis. A post hoc was not warranted. H8 was not supported.

Table 17

Descriptive Statistics for the Results of the Test for H8

Grading Scale Type	Gender	*M*	*SD*	*N*
Traditional	Female	18.23	4.15	557
	Male	18.91	4.74	538
Standards-based	Female	16.76	3.70	451
	Male	17.25	4.30	434

H9. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT mathematics subscores for students taught using a traditional grading

scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student ethnicity.

A sixth two-factor ANOVA was conducted to test H9. The two categorical variables used to group the dependent variable, ACT mathematics subscores, were grading scale type (traditional, standards-based) and ethnicity (Black, Hispanic, White, Other). The interaction effect (Grading Scale Type x Ethnicity) was used to test H9. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated there was not a statistically significant difference between at least two of the means, $F(3, 1972) = 1.185$, $p = .314$. See Table 18 for the means and standard deviations for this analysis. A post hoc was not warranted. H9 was not supported.

Table 18

Descriptive Statistics for the Results of the Test for H9

Grading Scale Type	Ethnicity	*M*	*SD*	*N*
Traditional	Black	16.27	3.14	179
	Hispanic	17.36	3.50	178
	White	19.40	4.63	557
	Other	19.44	4.84	181
Standards-based	Black	15.16	3.01	171
	Hispanic	16.20	2.73	135
	White	17.89	4.23	434
	Other	17.26	4.52	145

RQ7

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT reading subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022)?

H10. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT reading subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022).

A seventh two-factor ANOVA was conducted to test H10 and H11. The two categorical variables used to group the dependent variable, ACT reading subscores, were grading scale type (traditional, standards-based) and gender (male, female). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical among three or more groups, including a main effect for grading scale type, a main effect for gender, and a two-way interaction effect (Grading Scale Type x Gender). The main effect for grading scale type was used to test H10. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated a statistically significant difference between the means, $F(1, 1976) = 20.065$, $p = .000$, $\eta^2 = .010$. See Table 19 for the means and standard deviations for this analysis. The traditional grading scale mean ($M = 19.03$) was higher than the standards-based grading scale mean ($M = 17.77$). H10 was supported. The effect size indicated a small effect.

Table 19

Descriptive Statistics for the Results of the Test for H10

Grading Scale Type	*M*	*SD*	*N*
Traditional	19.03	5.87	1095
Standards-based	17.77	6.66	885

RQ8

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT reading subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

H11. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT reading subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender.

The interaction effect (Grading Scale Type x Gender) from the seventh ANOVA was used to test H10. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated there was not a statistically significant difference between at least two of the means, $F(1, 1976) = 0.187$, $p = .666$. See Table 20 for the means and standard deviations for this analysis. A post hoc was not warranted. H11 was not supported.

Table 20

Descriptive Statistics for the Results of the Test for H11

Grading Scale Type	Gender	*M*	*SD*	*N*
Traditional	Female	19.42	5.71	557
	Male	18.63	6.01	538
Standards-based	Female	18.28	6.57	451
	Male	17.25	6.72	434

H12. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT reading subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student ethnicity.

An eighth two-factor ANOVA was conducted to test H12. The two categorical variables used to group the dependent variable, ACT reading subscores, were grading scale type (traditional, standards-based) and ethnicity (Black, Hispanic, White, Other). The interaction effect (Grading Scale Type x Ethnicity) was used to test H12. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated there was not a statistically significant difference between at least two of the means, $F(3, 1972) = 0.196$, $p = .899$. See Table 21 for the means and standard deviations for this analysis. A post hoc was not warranted. H12 was not supported.

Table 21

Descriptive Statistics for the Results of the Test for H12

Grading Scale Type	Ethnicity	*M*	*SD*	*N*
Traditional	Black	16.30	4.65	179
	Hispanic	17.92	4.98	178
	White	20.08	6.09	557
	Other	19.59	6.06	181
Standards-based	Black	15.15	5.01	171
	Hispanic	16.87	5.08	135
	White	19.03	7.16	434
	Other	17.96	7.15	145

RQ9

For first-time 11th-grade ACT test takers, to what extent is there a difference in ACT science subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022)?

H13. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT science subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022).

A ninth two-factor ANOVA was conducted to test H13 and H14. The two categorical variables used to group the dependent variable, ACT science subscores, were

grading scale type (traditional, standards-based) and gender (male, female). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical among three or more groups, including a main effect for grading scale type, a main effect for gender, and a two-way interaction effect (Grading Scale Type x Gender). The main effect for grading scale type was used to test H13. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated a statistically significant difference between the means, $F(1, 1976) = 46.549$, $p = .000$, $\eta^2 = .023$. See Table 22 for the means and standard deviations for this analysis. The traditional grading scale mean ($M = 19.22$) was higher than the standards-based grading scale mean ($M = 17.70$). H13 was supported. The effect size indicated a small effect.

Table 22

Descriptive Statistics for the Results of the Test for H13

Grading Scale Type	*M*	*SD*	*N*
Traditional	19.22	4.89	1,095
Standards-based	17.70	4.98	885

RQ10

For first-time 11th-grade ACT test takers, to what extent is the difference in ACT science subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender and ethnicity?

H14. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT science subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student gender.

The interaction effect (Grading Scale Type x Gender) from the ninth ANOVA was used to test H13. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated there was not a statistically significant difference between at least two of the means, $F(1, 1976) = 0.009$, $p = .922$. See Table 23 for the means and standard deviations for this analysis. A post hoc was not warranted. H14 was not supported.

Table 23

Descriptive Statistics for the Results of the Test for H14

Grading Scale Type	Gender	*M*	*SD*	*N*
Traditional	Female	19.18	4.47	557
	Male	19.26	5.30	538
Standards-based	Female	17.68	4.54	451
	Male	17.72	5.41	434

H15. For first-time 11th-grade ACT test takers, there is a statistically significant difference in ACT science subscores for students taught using a traditional grading scale

(school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022) affected by student ethnicity.

A tenth two-factor ANOVA was conducted to test H15. The two categorical variables used to group the dependent variable, ACT reading subscores, were grading scale type (traditional, standards-based) and ethnicity (Black, Hispanic, White, Other). The interaction effect (Grading Scale Type x Ethnicity) was used to test H15. The level of significance was set at .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated there was not a statistically significant difference between at least two of the means, $F(3, 1972) = 0.483$, $p = .694$. See Table 24 for the means and standard deviations for this analysis. A post hoc was not warranted. H15 was not supported.

Table 24

Descriptive Statistics for the Results of the Test for H15

Grading Scale Type	Ethnicity	*M*	*SD*	*N*
Traditional	Black	16.98	3.40	179
	Hispanic	18.01	4.29	178
	White	20.15	4.94	557
	Other	19.76	5.16	181
Standards-based	Black	15.44	4.17	171
	Hispanic	16.90	4.10	135
	White	18.80	5.12	434
	Other	17.80	5.21	145

Additional Analyses

After completing the hypothesis testing, the researcher analyzed the data further. ACT composite and subscores were recoded into two categories: Met and not met. These categories were based on the college readiness benchmarks set by ACT. College readiness benchmarks for each subsection were averaged to determine a college readiness benchmark for an ACT composite score (ACT, 2022). ACT college readiness benchmarks can be found in Table 25. Students were sorted into two categories based on meeting or not meeting the scores in Table 25.

Table 25

ACT College Readiness Benchmarks

Test	ACT College Readiness Benchmark
Composite	21
English	18
Mathematics	22
Reading	22
Science	23

Note. Adapted from "ACT college readiness benchmarks," by ACT, 2022. https://www.act.org/content/act/en/college-and-career-readiness/benchmarks.html

Additional analyses were conducted to determine for first-time 11th-grade ACT test takers, to what extent there is a difference in students meeting the ACT college readiness benchmark for the composite score for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading

scale (school year 2021-2022). A two-factor ANOVA was conducted. The two categorical variables used to group the dependent variable, ACT Composite scores, were the recoded ACT scores based on meeting or not meeting ACT college readiness benchmark and grading scale type (traditional, standards-based). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical variable among three or more groups, including a main effect for grading scale type, a main effect for meeting the college readiness benchmark, and a two-way interaction effect (Grading Scale Type x Benchmark). For this analysis of composite scores, the interaction effect was used. The level of significance used was .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated a statistically significant difference between at least two of the means, $F(1, 1976) = 9.5178$, $p = .002$, $\eta^2 = .005$. See Table 26 for the means and standard deviations for this analysis. The traditional grading scale mean for those who met the ACT college readiness benchmark ($M = 24.59$) and for those who did not ($M = 16.14$) was higher than the standards-based grading scale mean for those who met the ACT college readiness benchmark ($M = 24.43$) and for those who did not ($M = 15.12$). The effect size indicated is small.

Table 26

Descriptive Statistics for the Results of the Additional Analysis of ACT Composite Scores

Grading Scale Type	College Readiness Benchmark	*M*	*SD*	*N*
Traditional	Met	24.59	3.31	1,095
	Not Met	16.14	2.35	
Standards-based	Met	24.43	3.25	885
	Not Met	15.12	2.68	

Additional analyses were conducted to determine for first-time 11th-grade ACT test takers, to what extent there is a difference in students meeting the ACT college readiness benchmark for English subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022). A two-factor ANOVA was conducted. The categorical variable used to group the dependent variable, ACT English subscores recoded by meeting ACT college readiness benchmark, was grading scale type (traditional, standards-based). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical among three or more groups, including a main effect for grading scale type, a main effect for meeting the college readiness benchmark, and a two-way interaction effect (Grading Scale Type x Benchmark). For this analysis of composite scores, the interaction effect was used. The level of significance used was .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis did not indicate a statistically significant difference between at least two of the means, $F(1, 1976) = 3.608$, $p = .058$. See Table 27 for the means and standard deviations for this analysis. A post hoc was not warranted.

Table 27

Descriptive Statistics for the Results of the Additional Analysis of ACT English Subscores

Grading Scale Type	College Readiness Benchmark	*M*	*SD*	*N*
Traditional	Met	23.13	4.34	1,095
	Not Met	13.56	2.37	
Standards-based	Met	22.84	4.30	885
	Not Met	12.68	2.62	

Additional analyses were conducted to determine for first-time 11th-grade ACT test takers, to what extent there is a difference in students meeting the ACT college readiness benchmark for mathematics subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022). A two-factor ANOVA was conducted. The categorical variable used to group the dependent variable, ACT mathematics subscores recoded by meeting ACT college readiness benchmark, was grading scale type (traditional, standards-based). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical among three or more groups, including a main effect for grading scale type, a main effect for meeting the college readiness benchmark, and a two-way interaction effect (Grading Scale Type x Benchmark). For this analysis of

composite scores, the interaction effect was used. The level of significance used was .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated a statistically significant difference between at least two of the means, $F(1, 1976) = 4.180$, $p = .041$, $\eta^2 = .002$. See Table 28 for the means and standard deviations for this analysis. The traditional grading scale mean for those who met the ACT college readiness benchmark ($M = 25.30$) and for those who did not ($M = 16.40$) was higher than or equal to the standards-based grading scale mean for those who met the ACT college readiness benchmark ($M = 25.30$) and for those who did not ($M = 15.80$). The effect size is small.

Table 28

Descriptive Statistics for the Results of the Additional Analysis of ACT Math Subscores

Grading Scale Type	College Readiness Benchmark	*M*	*SD*	*N*
Traditional	Met	25.30	2.65	1,095
	Not Met	16.40	2.18	
Standards-based	Met	25.30	2.81	885
	Not Met	15.80	2.42	

Additional analyses were conducted to determine for first-time 11th-grade ACT test takers, to what extent there is a difference in students meeting the ACT college readiness benchmark for reading subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022). A two-factor ANOVA was conducted. The categorical variable

used to group the dependent variable, ACT reading subscores recoded by meeting ACT college readiness benchmark, was grading scale type (traditional, standards-based). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical among three or more groups, including a main effect for grading scale type, a main effect for meeting the college readiness benchmark, and a two-way interaction effect (Grading Scale Type x Benchmark). For this analysis of composite scores, the interaction effect was used. The level of significance used was .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis indicated a statistically significant difference between at least two of the means, $F(1, 1976) = 27.854$, $p = .000$, $\eta^2 = .014$. See Table 29 for the means and standard deviations for this analysis. The traditional grading scale mean for those who met the ACT college readiness benchmark ($M = 26.56$) was lower than the standards-based grading scale mean for those who met the ACT college readiness benchmark ($M = 27.20$). The traditional grading scale mean for those who did not meet the ACT college readiness benchmark ($M = 16.05$) was higher than the standards-based grading scale mean for those who did not meet the ACT college readiness benchmark ($M = 14.69$). The effect size is small.

Table 29

Descriptive Statistics for the Results of the Additional Analysis of ACT Reading Subscores

Grading Scale Type	College Readiness Benchmark	*M*	*SD*	*N*
Traditional	Met	26.56	4.11	1,095
	Not Met	16.05	3.17	
Standards-based	Met	27.20	4.43	885
	Not Met	14.69	3.73	

Additional analyses were conducted to determine for first-time 11th-grade ACT test takers, to what extent there is a difference in students meeting the ACT college readiness benchmark for science subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using standards-based grading scale (school year 2021-2022). A two-factor ANOVA was conducted. The categorical variable used to group the dependent variable, ACT science subscores recoded by meeting ACT college readiness benchmark, was grading scale type (traditional, standards-based). The results of the two-factor ANOVA can be used to test for differences in the means of a numerical among three or more groups, including a main effect for grading scale type, a main effect for meeting the college readiness benchmark, and a two-way interaction effect (Grading Scale Type x Benchmark). The level of significance used was .05. When appropriate, an effect size, as indexed by *eta squared*, is reported.

The results of the analysis did not indicate a statistically significant difference between at least two of the means, $F(1, 1976) = 2.711$, $p = .100$. See Table 30 for the means and standard deviations for this analysis. A post hoc was not warranted

Table 30

Descriptive Statistics for the Results of the Additional Analysis of ACT Science Subscores

Grading Scale Type	College Readiness Benchmark	*M*	*SD*	*N*
Traditional	Met	25.88	3.11	1,095
	Not Met	17.07	3.10	
Standards-based	Met	25.30	2.88	885
	Not Met	15.91	3.44	

Summary

In this chapter, the descriptive statistics and the results of the statistical analysis for the 10 research questions and the associated 15 hypotheses were presented. The results of the additional analysis of the data were also presented. A summary of the study, the findings related to the literature, and the conclusion can be found in Chapter 5.

Chapter 5

Interpretation and Recommendations

Examined in this study were the differences in ACT composite scores and the subscores in English, mathematics, reading, and science of students taught in a standards-based grading environment compared to that of a traditional grading environment. Additionally, this study also examined the effects of ethnicity and gender on those differences. There are three main sections in Chapter 5: study summary, findings related to the literature, and conclusions.

Study Summary

This section contains an overview of the problem researched in this study, the lack of information on whether standards-based grading practices used in the classroom have impacted 11th grade students' performance on their first attempt at taking the ACT. This section also includes the purpose of this study and the research questions. A summary of the methodology used in this study and the major findings are found in this section.

Overview of the Problem

According to Durm (1993), the 100-point grading system dates to the early 19th century, and at the time of this study, most American schools continue using it. Finkelstein (1913) wrote how the traditional 100-point grading system was uncalibrated. Educational reforms such as the reauthorization of the Elementary and Secondary Education Act in 1994 and the signing of the No Child Left Behind Act in 2001, followed by the adoption of the Common Core State Standards in 2009, opened the door for educators to look at differently at grading systems. Through these reforms, accountability

systems began to look different in education and caused some schools to change their accountability system and student grades. The effects of this change on student success at all grade levels were unknown, making it an important topic to study. Because of the lack of data surrounding SBG, there is a need to explore the effect grading practices, specifically traditional versus standards-based, have on a student's academic achievement. This researcher sought to determine if District A's move to standards-based grading positively impacted ACT scores. Additionally, this researcher sought to identify if there was a difference by content: English, mathematics, reading or science. The researcher also wanted to determine if there was a difference by gender and ethnicity. Specifically, the researcher sought to determine if one gender or ethnicity was thriving more than others with the implementation of standards-based grading.

Purpose Statement and Research Questions

The overall purpose of this study was to determine how first-time 11th-grade ACT test-takers score on the ACT. The first purpose was to determine the extent there is a difference in ACT composite, English, math, reading, and science scores of students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022). The second purpose of this study was to determine the extent the difference in ACT composite, English, math, reading, and science scores of first-time 11th-grade ACT test takers taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022) are affected by student gender and ethnicity. To address the purposes of this study, 10 research questions were posed and 15 hypotheses were tested.

Review of the Methodology

A causal-comparative design using archived ACT data to measure student achievement from 2018-2019 and 2021-2022 was utilized for this quantitative study to determine if there was a significant difference in student achievement after the implementation of standards-based grading. The participants of this study were 11th-grade students who took the ACT for the first time in a midwestern suburban school district. The dependent variables were the ACT composite score and the subscores (English, mathematics, reading, and science). The independent variables were student gender and ethnicity. To test the 15 hypothesis, 10 two-factor ANOVAs were conducted.

Major Findings

The researcher for this study examined data for first time 11th-grade ACT takers scores used from the 2018-2019 school year to represent students taught in a traditional grading system and from 2021-2022 to represent students taught in a standards-based system. The researcher sought to determine what effect the implementation of standards-based grading had on student achievement, as represented by ACT scores, for this study. The researcher also examined whether the differences were affected by gender and ethnicity. After each of the hypothesis testing results are reported, the results of an additional analysis involving grading scale type and a third categorical variable based on the ACT college readiness benchmarks (met, unmet) are reported.

After analyzing the composite scores for those taught in the traditional grading system (2018-2019) and those taught in the standards-based system (2021-2022), the research found that overall, there was a significant difference and that composite scores decreased in the standards-based system. Gender and ethnicity did not affect the

differences found between the traditional and the standards-based grading scales. The results of the additional analysis involving grading scale type and ACT college readiness benchmarks were the traditional grading scale mean for those who met the ACT college readiness benchmark and those who did not was higher than the standards-based grading scale mean for those who met the ACT college readiness benchmark and those who did not.

The analysis of English subscores for those taught in the traditional grading system and those taught in the standards-based system showed that, overall, there was a significant difference and that mean scores were lower in the standards-based system. Gender and ethnicity did not affect the difference in ACT English subscores between the traditional and the standards-based grading scales. The results of the additional analysis involving grading scale type and ACT college readiness benchmarks were a significant difference between the traditional grading scale mean and the standards-based grading scale mean for those who met the ACT college readiness benchmark in English and those who did meet the ACT college readiness benchmark in English.

The analysis of mathematics subscores for those taught in the traditional grading system and those taught in the standards-based system showed that overall, there was a significant difference and that mean scores lowered in the standards-based system. Gender and ethnicity did not affect the differences in ACT mathematics subscores between the traditional and the standards-based grading scales. The results of the additional analysis involving grading scale type and ACT college readiness benchmarks showed a significant difference in ACT mathematics subscores for students meeting and not meeting the college readiness benchmark set by ACT. There was not a significant

difference between the traditional grading scale mean and the standards-based grading scale mean for those who met the ACT college readiness benchmark in mathematics and those who did meet the ACT college readiness benchmark in mathematics.

The analysis of reading subscores for those taught in the traditional grading system and those taught in the standards-based system showed that overall, there was a significant difference and that mean scores lowered in the standards-based system. Gender and ethnicity did not affect the differences found in ACT reading subscores between the traditional and the standards-based grading scales. An additional analysis was conducted to determine if there was a significant difference in students meeting and not meeting the college readiness benchmark set by ACT. The reading subscore mean for students meeting the benchmark increased in the standards-based grading system, while those not meeting the benchmark decreased in the standards-based system.

The analysis of science subscores for those taught in the traditional grading system and those taught in the standards-based system showed that overall there was a significant difference and that mean scores decreased in the standards-based system. Gender and ethnicity did not affect the differences found in ACT science subscores between the traditional and the standards-based grading scales. The results of the additional analysis provided evidence that there was not a significant difference in students meeting and not meeting the college readiness benchmark in science set by ACT. The mean for students meeting and not meeting the benchmark decreased in the standards-based system.

Findings Related to the Literature

The findings of the current study related to the literature are found in this section. A statistically significant difference was found in ACT composite scores and subscores (English, mathematics, reading, and science) for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022). Gender and ethnicity did not affect the differences in ACT composite and subscores found between the traditional and the standards-based grading scales. Findings from the additional analysis after scores were recoded based on the ACT college readiness benchmark were mixed. There was no significant difference in the means of English and science subscores as recoded by meeting the college readiness benchmark. There were significant differences for ACT composite, mathematics, and reading. The means were equal to or higher in the traditional system for ACT composite, mathematics and for those coded as not meeting the ACT college readiness benchmark in reading. However, means were higher in the standards-based system for those coded as meeting the ACT college readiness benchmark in reading.

Townsley (2017), Kelly (2018), and Bosanec (2020) all also studied the effects of grading type on ACT scores at the secondary level. Townsley (2017) examined the effects on ACT composite scores, as well as mathematics and English subscores. The findings in the current study support Townsley's results. In both studies, the difference in composite ACT score and sub sections in mathematics and English were statistically significant. Both studies also showed that the mean decreased with the implementation of standards-based grading. Bosanec (2020) studied the effects of grading type on English,

mathematics, and science. The results of the current study support Bosanec (2020), who found a statistical difference in mathematics and science scores and a decrease in the mean scores after the implementation of standards-based grading. However, the results of the current study are in contrast to Bosanec (2020), who did not find a statistically significant difference in English scores. The results of the current study did not support Kelly's (2018) findings that indicated no statistically significant difference in ACT composite scores.

Kemp (2007), Yoakum (2014), Graves (2016), and Poll (2019) studied the effects of grading type on student achievement at the secondary level, but unlike the current study, the ACT was not the measure of student achievement. Kemp (2007) measured student achievement using the Plato eduTest for 7th graders. Unlike the current study, Kemp showed higher means in the standards-based system, but the difference was not statistically significant. Both of these findings are not supported by the current study. Yoakum (2014) measured student achievement using high school end-of-course exams in communication arts and mathematics. Similar to the current study, Yoakum (2014) found a statistically significant difference in communication arts scores and a decrease in the mean score after the implementation of standards-based grading. However, the results of the current study did not support Yoakum's (2014) finding of a statistically significant difference in mathematics scores with an increase in the mean after the implementation of standards-based grading. Graves (2016) used state assessments to measure student achievement by grading type. Graves's (2016) study showed mixed results, with 50% of the schools showing an increase in student achievement in English after the implementation of standards-based grading, while the other 50% showed no difference.

The results of Graves' analyses are not supported by the current study results, which found a decrease in English scores after the implementation of standards-based grading. Poll (2019) focused his sample on SAGE students and used the end-of-level SAGE scaled scores to measure student achievement. This study examined mathematics, science, and language arts. Poll (2019) found a statistically significant difference in all three areas, with the scores increasing at implementing standards-based grading. The results of the current study contrast with Poll's (2019) results.

Buttrey (2014), Norton (2014), and Hargrove (2020) all studied the effects of grading type on student achievement but at the elementary level. Buttrey (2014) and Norton (2014) used the K-PREP state assessment to measure student achievement. Both studies showed a significant difference in mathematics scores, which is supported by the current study; however, unlike the current study Buttrey (2014) and Norton (2014) found an increase in the mean scores while the current study showed a decrease. Buttrey (2014) and Norton (2014) also found a difference in reading/language arts, but unlike the current study, it was not significant. Buttrey (2014) and Norton (2014) did show an increase in reading/language arts, just not significant, while the current study showed a decrease. Hargrove (2020) used Measures of Academic Progress to measure student achievement, specifically reading growth, after the implementation of standards-based grading. Similar to the current student Hargrove (2020) found a statistically significant difference by grading type, with the standards-based system showing less growth than the traditional system.

Conclusions

Conclusions for this study about the effect of standards-based grading compared to traditional grading on student achievement as measured by the ACT college entrance exam can be found in this section. Also, in this section are the conclusions for this study about the effects of gender and ethnicity on the differences by grading scale can be found in this section. This section also contains implications for action and recommendations for future research. Finally, there are closing remarks from the researcher.

Implications for Action

The results of this study can help provide feedback to the leaders of District A on the effectiveness of the standards-based grading system as it is currently implemented. The results showed a significant difference in composite scores as well as English, mathematics, reading, and science subscores for students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022). Based on these results, it is recommended that district leaders take the actions outlined in this section.

First, it is recommended that this study be continued by the district for three or more years to help decrease the effect, if any, of the COVID-19 pandemic had on the data. This data analysis would help district leaders better understand the impact of the standards-based grading system. It is important for district leaders to use data when determining what is best for student academic achievement.

Second, based on the results of this study, it is recommended that district leaders could invest time in professional development on standards-based grading for secondary teachers. It is recommended that the professional development begin with how standards-

based grading affects day-to-day teaching so that it becomes less about the grading scale and more about classroom practice. Essentially, district leaders should focus on standards-based learning, not grading.

Third, District A would not provide the researcher with data on student socio-economic status (SES). It is recommended that the district, even though that data could not be provided to the researcher, analyze the data by SES. This would help determine whether the difference in ACT scores for students taught using a traditional grading scale and students taught using a standards-based grading scale was affected by student SES.

Finally, based on the results of this study, it is recommended that district leaders look at data by school. Because each school used a different grading scale, it would be valuable information for district leaders to know if one scale is producing better results. Once the scale is established, it is recommended that all secondary schools in the district use the same scale.

Recommendations for Future Research

A standards-based approach is common at the elementary level but less common at the secondary level. The uncommon use of minimal research of a standards-based approach at the secondary level drove this researcher to conduct this research. Recommendations for continued and future research can be found in this section.

The first recommendation for future research would be to continue this study but use the data from the next three years. This recommendation is made for a couple of reasons. First, the COVID-19 pandemic hit in 2020, causing schools to close and forcing some schools to go to a hybrid schedule where students were only at school two days a week. That school year was in the middle of the two years of data used in this study, so it

is recommended that this study be replicated without a global pandemic possibly effecting the data. Second, the data used in this study was from the beginning of the implementation of standards-based grading. Using data from when it is a more established practice could yield different results. Also, the schools in this study implemented a change in their grading system without necessarily changing classroom practices. The school district in this study has since partnered with experts in standards-based grading who are focused on changing classroom practice before changing the grading scale. Because of this, using data from future years may again yield different results.

The second recommendation for future research would be to change the instrument used to measure student achievement. For this study, the ACT college entrance exam was used. Using student grade point averages might be a better indicator since many colleges are switching to using that for admittance instead of the ACT.

A third recommendation for future research would be to include student SES as an independent variable. For this study, the SES data was not released by District A. Analyzing this data would be beneficial in determining if the difference in ACT scores for students taught using a traditional grading scale and students taught using a standards-based grading scale are affected by student SES. The addition of this variable could provide more insight especially for more urban-like districts that have a higher percentage of low SES students.

Finally, the three schools in this study used different grading scales (see Tables A1-A3 in Appendix A). Future researchers could compare the data among the three

schools. This comparison could show if there were statistically significant differences based on the scale used.

Concluding Remarks

Educational reforms such as the reauthorization of the Elementary and Secondary Education Act in 1994 and the signing of the No Child Left Behind Act in 2001 followed by the adoption of the Common Core State Standards in 2009 led to changes in how schools in the United States were held accountable for student success. Due to the ever-changing accountability system mandated by reforms such as those mentioned above, secondary schools are searching for better ways to track student learning. A standards-based system can often be found at the elementary level but is relatively new to the secondary level. Because standards-based is a relatively new approach for secondary schools, school leaders must continue making informed decisions that best support student achievement.

References

ACT. (2020). *ACT Technical Manual.* https://success.act.org/s/article/The-ACT-Technical-Manual

ACT. (2022). *ACT college readiness benchmarks.* https://www.act.org/content/act/en/college-and-career-readiness/benchmarks.html

American Psychological Association. (n.d.). *Socioeconomic status*. https://www.apa.org/topics/socioeconomic-status

Bosanec, D. (2020). *An impact program evaluation of traditional and standards-based grading practices in relation to ACT performance in a suburban high school* (Publication No. 27742225) [Doctoral dissertation, Cardinal Stritch University]. ProQuest Dissertations and Theses Global.

Bouchrika, I. (2021, August 9). *Standards-based grading: Definition, benefits & comparison with traditional grading*. Research.com. https://research.com/education/standards-based-grading

Brookhart, S. M., Guskey, T. R., Bowers, A. J., McMillan, J. H., Smith, J. K., & Smith, F. (2016, December). A century of grading research: Meaning and value in the most common educational measure. *Review of Educational Research, 86*(4), 803-848. https://uknowledge.uky.edu/edp_facpub/2

Buttrey, K. (2014). *The Effects of standards-based grading on student achievement, grade alignment, and teacher perceptions* (Publication No. 3582782) [Doctoral dissertation, Union University]. ProQuest Dissertations and Theses Global.

Creswell, J. W., & Creswell, J. D., (2018). *Research design: qualitative, quantitative, and mixed methods approaches*. (5th ed.). SAGE.

Decker, M., Kleppinger, K., & Pewitt, A. (2018). *Standards-based grading and its impact on student achievement* (Publication No. 10937636) [Doctoral dissertation, Lipscomb University]. ProQuest Dissertations and Theses Global.

Department of Elementary and Secondary Education. (2022). *District demographic data*. https://apps.dese.mo.gov/MCDS/Reports/SSRS_Print.aspx?Reportid=6c5b805c-5af7-4c33-be41-dc2b83ded4aa

d'Erizans, R. (2020). *A correlational study of standards-based grading, traditional grades, and achievement* (Publication No. 28092216) [Doctoral dissertation, Northeastern University]. ProQuest Dissertations and Theses Global.

Durm, M. (1993). An A is not an a: A history of grading. *The Educational Forum, 57*(3), 294-297. doi:10.1080/00131729309335429

EdPost Staff. (2015). *The ABC's of ESEA, Essa and No Child Left Behind*. Ed Post. (2015, January 14). https://educationpost.org/the-abcs-of-esea-essa-and-no-child-left-behind/

Edwards, C. (n.d.). *Department of Education timeline*. Cato Institute. https://www.downsizinggovernment.org/education/timeline-growth

Finkelstein, L., (1913). The marking system in theory and practice. *Educational Psychology Monographs,* (No. 10). Warwick & York. https://www.google.com/books/edition/The_Marking_System_in_Theory_and_Practic/2pxJAAAAIAAJ?hl=en&gbpv=1&printsec=frontcover

Gewertz, C. (2020, December 10). The Common Core explained. *Education Week.* https://www.edweek.org/teaching-learning/the-common-core-explained/2015/09

Graves, C., (2016). *Competency based grading in students' achievement.* (Publication No. 11012802) [Doctoral dissertation, University of Missouri-Columbia]. ProQuest Dissertations and Theses Global.

Guskey, T., (2006). "*It wasn't fair!*" *Educators' recollections of their experiences as students with grading* [Paper presentation]. American Educational Research Association, San Francisco, California.

Haptonstall, K. G., (2010). *An analysis of the correlation between standards-based, non-standards-based grading systems and achievement as measured by the Colorado Student Assessment Program (CSAP)* (Publication No. 3397087) [Doctoral dissertation, Cappella University]. ProQuest Dissertations and These Global.

Hardegree, A. C., (2012). *Standards-based assessment and high stakes testing: Accuracy of standards-based grading* (Publication No. 3540113) [Doctoral dissertation, Liberty University]. ProQuest Dissertations and Theses Global.

Hargrove, M. H. (2020). *Quantitative study exploring traditional and standards-based grading practices' effects on student yearly reading achievement growth as measured by MAP* (Publication No. 128150997) [Doctoral dissertation, North Central University]. ProQuest Dissertations and Theses Global.

Hattie, J. (n.d.). *Hattie effect size list - 256 influences related to achievement.* Visible Learning. https://visible-learning.org/hattie-ranking-influences-effect-sizes-learning-achievement/

Heflebower, T., Warrick, P., & Hoegh, J. (2014). *A school leader's guide to standards-based grading*. Marzano Resources.

Hobden, H. (2019, March 29). *Traditional grading scale: Pros, cons & alternatives*. Kidio. https://kid.io/traditional-grading-scale-pros-cons-alternatives/

Iamarino, D. (2104). The benefits of standards-based grading: A critical evaluation of modern grading practices. *Current Issues in Education, 17*(2), 1-10. http://cie.asu.edu/ojs/index.php/cieatasu/article/view/1234

Kelly, D. (2018). *The relationship between standards-based grading and student achievement in secondary math and English language arts, and other annual performance measures including attendance, dropout, and ACT scores* (Publication No. 10931019) [Doctoral dissertation, Evangel University]. ProQuest Dissertations and Theses Global.

Kemp, M. D., (2007). *A comparison of traditional instruction and standards-based instruction on seventh-grade mathematics achievement.* [Doctoral Dissertation, Mississippi State University]. https://scholarsjunction.msstate.edu/td/76

Khan, A. (2019, July 7). *Pros and cons of standards-based grading – TOPPR bytes*. Toppr bytes. https://www.toppr.com/bytes/pros-and-cons-of-standards-based-grading/

Klein, A. (2015, April 10). No Child Left Behind: An overview. *Education Week*. https://www.edweek.org/policy-politics/no-child-left-behind-an-overview/2015/04

Knight, M., & Cooper, R. (2019). Taking on a new grading system: The interconnected effects of standards-based grading on teaching, learning, assessment, and student behavior. *NASSP Bulletin, 103*(1), 65-92. https://doi.org/10.1177%2F0192636519826709

Lee, A., Carberry, A. R., Diefes-Dux, H. A., Atwood, S.A., & Siniawski, M. T. (2018). Faculty perception before, during, and after implementation of standards-based grading. *Australasian Journal of Engineering Education*, *23*(2), 53-61. https://doi.org/10.1080/22054952.2018.1544685

Lee, C. (2020, October 14). *What is the history of grading?* Turnitin. Retrieved from https://www.turnitin.com/blog/what-is-the-history-of-grading

Lunenburg, F. C., & Irby, B. J. (2008). *Writing a successful thesis or dissertation: Tips and strategies for students in the social and behavioral sciences.* Corwin Press.

Mahr, L. (2020). *Education reform and the transition to standards based grading in the United States* [Master's thesis, Northwestern College]. NWCommons. https://nwcommons.nwciowa.edu/cgi/viewcontent.cgi?article=1262&context=education_masters

Mann, H. (1872). *Annual reports on education*. Lee and Shepard Publishers. https://play.google.com/books/reader?id=1Dk4AAAAYAAJ&pg=GBS.PP8&hl=en

Marzano, R. J. (2010). *Formative assessment & standards-based grading*. Solution Tree Press.

McCarthy, J. & Sharp, T. (2016). *A study on student science achievement as measured by a standardized test of the effect of traditional and standards-based grading practices* [Specialist thesis, Northwest Missouri State University]. https://www.nwmissouri.edu/library/fieldstudies/2016/mccarthyjsharpt.pdf

Meador, D. (2019, July 15). *The traditional grading scale: Endangered species or ageless wonder?* ThoughtCo. https://www.thoughtco.com/the-pros-and-cons-of-utilizing-a-traditional-grading-scale-3194752

Mink, A. (2015). *A comparative analysis of a standard based grading system versus a traditional based grading system with respect to student academic achievement and motivation* (Publication No. 3689505) [Doctoral dissertation, Concordia University]. ProQuest Dissertations and Theses Global.

Muñoz, B. M. A., & Guskey, T. R. (2015). Standards-based grading and reporting will improve education. *Phi Delta Kappan, 96*(7), 64–68. doi:10.1177/0031721715579043

National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* https://edreform.com/wp-content/uploads/2013/02/A_Nation_At_Risk_1983.pdf

Norton, J. B. (2014). *The Impact of traditional and standards-based grading on teacher-issues grades, student achievement, and parental understanding of grades* (Publication No. 3582795) [Doctoral dissertation, Union University]. ProQuest Dissertations and Theses Global.

Poll, T. R. (2019). *Standards-based grading: A correlational study between grades and end-of-level test scores* (Publication No. 13428322) [Doctoral dissertation, Utah State University]. ProQuest Dissertations and Theses Global.

Rainey, L. (2016). *Effects of standards-based grading on student achievement: A correlational investigation* (Publication No. 102399760) [Doctoral dissertation, Dallas Baptist University]. ProQuest Dissertations and Theses Global.

Reese, B. A. (2015). *The impact of standards-based grading on student academic performance on standardized assessments* (Publication No. 3732038) [Doctoral dissertation, Wingate University School of Graduate Education]. ProQuest Dissertations and Theses Global.

Schinske, J., & Tanner, K. (2014). Teaching more by grading less (or differently). *CBE—Life Sciences Education, 13*(2), 159–166. https://doi.org/10.1187/cbe.cbe-14-03-0054

Schneider, J., & Hutt, E. (2013). Making the grade: A history of the A–F marking scheme. *Journal of Curriculum Studies, 46*(2), 201–224. https://doi.org/10.1080/00220272.2013.790480

Scriffiny, P. L. (2008). Seven reasons for standards-based grading. *Educational Leadership, 66(*2), 70–74. https://www.ascd.org/el/articles/seven-reasons-for-standards-based-grading

Sieling, C. (2013). *Standards-based grading in mathematics: Effects on student achievement and attitude* (Publication No. 1523840) [Masters of science in education action research project, Southwest Minnesota State University]. ProQuest Dissertations and Theses Global.

Spencer, K., (2012, September/October). Standards based grading: New report cards aim to make mastery clear. *Harvard Education Letter, 28*(5), 4-10.

Swan, G. M., Guskey, T. R., & Jung, L. A. (2014). Parents' and teachers' perceptions of standards-based and traditional report cards. *Educational Assessment, Evaluation and Accountability, 26*(3), 289–299. https://doi.org/10.1007/s11092-014-9191-4

Thompson, C. J. (2009). Preparation, practice, and performance: An empirical examination of the impact of standards-based instruction on secondary students' math and science achievement. *Research in Education, 81*(1), 53-62. doi:10.7227/RIE.81.5.

Tierney, R. D., Simon, M., & Charland, J. (2011). Being fair: Teachers' interpretations of principles for standards-based grading. *The Educational Forum, 75*(3), 210-227. http://dx.doi.org/10.1080/00131725.2011.577669

Townsley, M. (2017). *Impact of standards-based grading practices on post-secondary readiness as measured by standardized college entrance exams.* (Publication No. 10253938) [Doctoral dissertation, University of West Georgia]. ProQuest Dissertations and Theses Global.

Townsley, M. (2019). Considering standards-based grading. *Journal of School Administration Research and Development*, *4*(1), 35–38. https://doi.org/10.32674/jsard.v4i1.1941

Tyree-Hamby, A. L. (2015). *An examination of the correlation between teacher-assigned standards-based grades and teacher-assigned traditional grade and student achievement* (Publication No. 3732246) [Doctoral dissertation, Lindenwood University]. ProQuest Dissertations and Theses Global.

Washington University. (2019, November 22). *Race and ethnicity self-study guide*. https://students.wustl.edu/race-ethnicity-self-study-guide/

Weinhold, S. P. (2015). *A quantitative study of standards-based grading and the Missouri Assessment Program* (Publication No. 10933997) [Doctoral dissertation, Missouri Baptist University]. ProQuest Dissertations and Theses Global.

Welsh, M. E., D'Agostino, J. V., & Kaniskan, B. (2013). Grading as a reform effort: Do standards-based grades converge with test scores? *Educational Measurement: Issues & Practice*, *32*(2), 26-36. doi:10.1111/emip.12009

Wulick, D. A. (2020, February 29). *How do you calculate your GPA? Step-by-step instructions*. https://blog.prepscholar.com/how-do-you-calculate-gpa

Yoakum, K. (2014). *The affect of standards based grading on student achievement* [Research paper, Northwest Missouri State University]. https://www.nwmissouri.edu/library/researchpapers/2014/Yoakum,%20Katrina.pdf

Appendices

Appendix A. School Grading Scales

Table A1

School A Grading Scale

Score	Description
10.0	Advanced – a performance beyond what a standard requires
9.5	Proficient– the level of learning required for all students
8.5	Nearing proficient - No major errors or omissions regarding score 7.5 content and partial success with score 9.5 content
7.5	Basic – basic knowledge or skill necessary for mastering the target content
6.5	Below basic - Partial success with score 7.5 content and major errors or omissions regarding score 9.5 content
5	No Evidence

Table A2

School B Grading Scale

4.0 Scale Score	Traditional Letter Grade
3.00-4.00	A
2.50-2.99	B
2.00-2.49	C
1.00-1.99	D
Below 1.00	F

Table A3

School C Grading Scale

Standards-Referenced Score	Traditional Scoring %	Description
4	100	Student has mastered and demonstrated deep understanding beyond the learning target.
3	90	Student has shown complete mastery of the learning target.
2	70	Student has shown evidence of the basic knowledge/ skills towards mastering the learning target
1	60	Student evidence demonstrates progress towards the basics of the learning target.
0	0	Student has shown no evidence of mastery, even with help.

Appendix B: Consent to Conduct Research

Request to Conduct Research with the [redacted] School District

Please read and follow the directions in this document.

A copy of this form must be returned to [redacted] Director of Data and Accountability with the necessary signatures BEFORE approval can be granted to conduct research. Please include your last name as you save your document e.g. Shanks-Research Request.

Name of Applicant: Molly Smith
Employee of [redacted] Schools? Yes☒ No ☐
If YES, location of building and your position with the [redacted] District
[redacted] School Principal

Is the research in fulfillment of a graduate program requirement and/or in partnership with an external organization? Yes ☒ No ☐
If YES, what is the name of external organization and lead contact person?
External Organization: Baker University
Lead Contact Person and Position. Dr. Susan Rogers, Dissertation Advisor
Briefly describe the purpose of the research: The purposes of this study are to determine how first-time 11th grade ACT test-takers score on the ACT, the extent there is a difference in ACT composite, English, math, reading, and science scores of students taught using a traditional grading scale (school year 2018-2019) and students taught using a standards-based grading scale (school year 2021-2022, and the extent the difference in students taught using a standards-based grading scale affected by student gender, ethnicity, and socio-economic status.

Submission Requirements – please mark check boxes as appropriate

1. A copy of the complete application submitted for formal approval by a human subjects review board. This application should include, at a minimum:
   a. A brief summary of the purpose and scope of the research including:
   - ☐ The extent to which the research addresses and/or aligns with the goals of the school district
   - ☐ Potential benefit of the research to positively impact district, building, or classroom practice

   b. A brief summary of the research methods including:
   - ☐ Participants
   - ☐ Selection process
   - ☐ Remuneration procedures (if applicable)
   - ☐ Assurance of confidentiality of participant identification
   - ☐ Consent and assent procedures and documents
   - ☐ Activities related to the research, including proposed survey, interview, and/or questions/instruments
   - ☐ Extent of intrusiveness/disruption regarding classroom instruction
   - ☐ Time/effort requirements of participants
2. ☐ Evidence that the proposed research has been formally approved through a review board for protection of human subjects.
3. ☐ Assurance from the researcher that building principals, teachers, students and/or their parents may opt out of participation without consequence even with approval by the district team.
4. **Signature of Principal(s)** of building(s) impacted by research study before approval.

X

Signature of Director of Data and Accountability: [redacted]

Team Review Date: July 25, 2022 Approved: X Not Approved: _____

Appendix C: Institutional Review Board Approval

BAKER UNIVERSITY
Own Confidence

Baker University Institutional Review Board

August 30th, 2022

Dear Molly Smith and Susan Rogers,

The Baker University IRB has reviewed your project application and approved this project under Exempt Status Review. As described, the project complies with all the requirements and policies established by the University for protection of human subjects in research. Unless renewed, approval lapses one year after approval date.

Please be aware of the following:

1. Any significant change in the research protocol as described should be reviewed by this Committee prior to altering the project.
2. Notify the IRB about any new investigators not named in original application.
3. When signed consent documents are required, the primary investigator must retain the signed consent documents of the research activity.
4. If this is a funded project, keep a copy of this approval letter with your proposal/grant file.
5. If the results of the research are used to prepare papers for publication or oral presentation at professional conferences, manuscripts or abstracts are requested for IRB as part of the project record.
6. If this project is not completed within a year, you must renew IRB approval.

If you have any questions, please contact me at npoell@bakeru.edu or 785.594.4582.

Sincerely,

Nathan D. Poell

Nathan Poell, MLS
Chair, Baker University IRB

Baker University IRB Committee
Tim Buzzell, PhD
Nick Harris, MS
Scott Kimball, PhD
Susan Rogers, PhD