CURRENT GRADING PRACTICES OF SECONDARY SCHOOL TEACHERS IN THE CORE CONTENT AREAS: WHAT DO GRADES MEAN?

by

Nathan E. McLean

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Education in
Educational Leadership

Charlotte

2018

Approved by:

_____
Dr. Claudia Flowers

_____
Dr. Jim Watson

_____
Dr. Debra Morris

_____
Dr. Paola Pilonieta

Abstract

NATHAN E. MCLEAN. Current grading practices of secondary school teachers in the core content areas: what do grades mean? (Under the direction of Dr. CLAUDIA FLOWERS and Dr. JIM WATSON).

Using survey research methodology, the study investigates assessment and grading practices of secondary school teachers. A total of 248 teachers in the Piedmont and coastal regions of North Carolina completed a web-based survey examining academic and nonacademic factors that impacted the final grades of their students and assessment types used in the classroom. The results suggested that the grades teachers assigned are a conglomerate of various factors and mainly use achievement but also incorporate non-achievement factors. The most used factors to determine student grades were: percent correct, use of zeros, objective mastery, academic performance, effort, homework completion, the adjustment of borderline grades, homework quality, the degree that a student pays attention or participates, and improvement since the beginning of the year. The most prevalent assessment type was summative assessments with objective questions created by the teacher. While teachers agreed that the grades assigned by teachers across the state of North Carolina are inflated and contribute to the continually increasing graduation rate, most teachers also believe their grades represent achievement, as opposed to any other factors. There were differences found between core content teachers in the factors used to assign final grades, with math teachers less likely to use objective mastery and homework completion. Implications for school leaders and suggestions for future research are discussed.

Acknowledgments

After intensively reading and writing for more than a year, this final note signifies the end of the dissertation. The process undertaken during this time has been a wonderful learning experience both professionally and personally, and an achievement I am very proud of. Although I have completed the doctoral program and my dissertation, I could not have done so without the help of several people, and I would like to recognize their contributions.

First, I would like to thank Dr. Claudia Flowers for working diligently with me from the beginning to narrow my topic and framing the process to keep me moving in the right direction. Your open communication and honest feedback were vital to completing the process while simultaneously instilling a desire to continue in academia. Your professionalism should be an example to others.

I would also like to recognize Dr. Jim Watson, Dr. Debra Morris, and Dr. Paola Pilonieta for serving on my Dissertation Committee. The feedback you have provided throughout the process was valuable and integral to the successful completion of the final document.

Lastly and most importantly, I would like to thank my wife Andrea, daughter Addison, and my son Bryson for their unwavering support of me throughout the entire process. The time they sacrificed allowing me to work on numerous vacations and weekends has finally paid off and they can have their husband and father back. Without

their help, this day of culmination would never have come and for this I am forever grateful.

Table of Contents

List of Tables

Chapter 1: Introduction

Assessment and grading in education have been perceived through history to measure and represent student achievement (Brookhart, 2015). The first major study of the public education system was revealed in 1983 with the publication of *A Nation at Risk*. The report revealed that students in the United States were lacking basic skills and behind other nations, even though students were successfully progressing through the school systems (National Commission for Excellence in Education, 1983). The report led to the standards-based reform movement emphasizing accountability, through measuring student achievement (Hamilton, Stecher, & Yuan, 2008) and it indicated that teacher grading practices were an area of concern (National Commission for Excellence in Education, 1983).

The study of grading practices became prevalent in the 1980s and into the 1990s, and researchers found that grades were composites of achievement and non-achievement related factors, such as behavior and effort (Brookhart, 2015). Guskey (2006) describes this grading system as hodgepodge grading because the items used in determining final grades frequently vary between teachers and schools. Many of these items do not provide any insight into achievement, and when they do, it rarely monitors student progress (Frey & Schmidt, 2010). As a result, the final grades created from these varying factors blur the intended meaning of grades making them difficult to interpret possibly leading to skill deficiency which could hinder their ability to graduate or to be successful after graduation (Bursuck et al., 1996).

Grading is an essential component of the educational process, impacts every student within its system, and is accepted, without criticism, as being a valid representation of student learning by all stakeholders (Schneider & Hutt, 2014). Parents of secondary school students use the information provided by teachers to discern their child's degree of performance or achievement in a course (Randall & Engelhard, 2010) and allow parents to intervene as necessary to assist the teacher or their child. Teachers must use the information gathered from assessments and grading to make informed decisions regarding the instruction provided in the classroom so that they can meet the needs of all students which are often difficult to distinguish and vary among students (Black & Wiliam, 1998; Curry, Mwavita, Holter, & Harris, 2016). The grades that teachers assign to students impact their futures and the most important aspect of grading for teachers to consider is that once a final grade is assigned, it can never be recovered (Brookhart, 1994).

Inaccurate grading practices can negatively impact students' futures by not providing them with the adequate knowledge or skills which inhibit their ability to obtain gainful employment after graduation or to succeed in post-secondary studies (Bursuck, et al., 1996). The importance of valid grading practices is emphasized because the students with high grade point averages typically are awarded more scholarships and afforded additional options in the colleges or universities they could attend, but, and as equally important, the opposite is also true (Allen, 2005). Although high school grades are a strong predictor of college success, there is a wide variation in student performance

among students with the same high school grades which forces universities to consider other factors (Cyrene & Chan, 2012).

Allen (2005) attributes the variation in performance due to biased and invalid grades that teachers assign because they include factors other than achievement. The review of the literature revealed that assessment and grading practices varied widely among studies and included various components of both achievement, through a variety of assessment types, and non-achievement factors including four common categories (a) achievement, (b) ability, (c) behavior, and (d) effort, among others. These achievement and non-achievement factors are combined without a standardized weighting system into a final grade making it difficult to determine what each grade means (Brookhart, 2015; Ornstein, 1994).

The ability of school leaders to determine the meaning of grades is vital in making informed decisions regarding reform efforts within their building or district. A positive result of assessment and grading at the secondary level is that it better represents tested achievement than any other level, indicating that it is useful in the approximation of achievement on standardized assessments, but improvement is necessary before grades can be trusted as a valid measure (Brookhart, 2015). Fully understanding assessment and grading practices of current teachers at the secondary level is important, especially in the core content areas, because the standards-based reform movement, that began in 1994 with the Improving America's School Act, or better known in 2001 as No Child Left Behind, emphasizes school accountability measured by standardized assessments that provide evidence of student performance in a specific subject area.

The large-scale study of assessment and grading practices of teachers has been declining since the 1990s (Brookhart, 2015) and there have been minimal studies examining the core content areas at the high school level. The lack of information in this area is important to investigate because it can have significant implications in the age of accountability regarding educational reform in assessment and grading. The current study examined assessment and grading practices post-1994, which marked the beginning of the standards-based educational reform.

Problem Statement

A critical issue in secondary education is the development of college-ready students. North Carolina's graduation rate has risen remarkably from 68.3% in 2005-2006, to 85.9% in 2014-2015, (North Carolina Department of Public Instruction, 2016) aligning with the change in the national graduation rate, 74.7 to 83.4%, over the same period (National Center for Education Statistics, 2015). These impressive results should produce an abundance of students better prepared to transition into post-secondary education, but in reality,

> nearly 60 percent of first-year college students discover that, despite being fully eligible to attend college, they are not academically ready for post-secondary studies…Increasingly, it appears that states or post-secondary institutions may be enrolling students under false pretenses. Even those students who have done everything they were told to do to prepare for college find, often after they arrive, that their new institutions have deemed them unprepared (National Center for

Public Policy and Higher Education and The Southern Regional Education Board, 2010, p. 1).

A key statistic in determining college readiness is a student's grade point average in college preparatory courses, because it is often the most important aspect in determining their acceptance into a post-secondary institution (Allen, 2005; Cliedinst, Korateng, & Nicola, 2015; Cyrenne & Chan, 2012: National Association of College Admission Counseling, 2015). The grades assigned by teachers are meant to align with achievement, and initially expected to fall into the typical bell-curve distribution with the "average" grade being considered a "C", but as early as the 1970s, more than half of the students earned an "A" or a "B" (Schneider & Hutt, 2014). The increasing trend of a classroom grade also has broader impacts by affecting overall student GPAs.

"Overall grade point averages increased from 2.68 in 1990 to 3.00 in 2009" (p. 13), but there was not a statistically significant increase from 2005 to 2009 (Nord et al., 2011). Over the same period, the authors found that the lowest grade point averages (GPAs) were in the core academic areas such as English, math, science and social studies. Among the core content areas, graduates in math and science earned lower GPAs when compared to English and social studies and this was the only statistically significant difference (Nord et al., 2011). After 2009, the graduation rate increased at a much larger rate nationally, almost 6% (National Center for Education Statistics, 2015) and in North Carolina nearly 11% (North Carolina Department of Public Instruction, 2016), suggesting significant changes in GPAs since the results of the 2009 NAEP High School transcript study.

It is obvious that grades teachers assign are changing, and current understandings of their meanings present a conundrum when used as a measure of student learning as evidenced by the increasing student GPAs, graduation rate, and the lack of preparedness for post-secondary studies. If grades, and the assessments used to calculate them, carry such weight, then the focus must shift to understanding the individual nuances of the grading practices existing within today's high schools in the core content areas.

Purpose Statement

The current study updated empirical evidence surrounding assessment and grading practices at the secondary level to address the gap between increasing graduation rates and the preparedness of incoming post-secondary students. The study provides a comprehensive description of assessment and grading practices within the four core content areas, English, math, science, and social studies, of current classroom teachers, and identifies meaningful relationships between the core content areas.

Research Questions

The following research questions guided the present study:

1. What factors do secondary teachers use in determining grades for students in the four core content areas, English, math, science, and social studies?

2. What types of assessments do secondary teachers use in grading students in the four core content areas, English, math, science, and social studies?

3. How do the types of assessments and factors used in grading students vary among English, math, science, and social studies teachers?

Significance Statement

Despite the increased emphasis on producing college and career-ready students, teacher-assigned grades inadequately predict student achievement. The findings of this study revealed the current assessment and grading practices of high school teachers within the four core content areas and delineated the meanings of grades. The understanding of grades and their meaning is vital to school leaders, parents, and individuals who must rely on accurate reporting records that represent student achievement to make decisions.

As school leaders, it is important to understand grading practices to ensure they accurately represent student learning because McMillian (2001) suggests that teachers use grading practices to ensure students succeed through establishing practices that allow them to earn high marks. Also, administrators may inadvertently influence grading practices when they base a teacher's success on the distribution of grades; low grades are generally thought of poorly, contrary to grading's purpose, leading to artificially inflated grades (Kunnath, 2017). Through a detailed examination of grades, administrators will be able to address assessment and grading practices at the school level through policy development and decision-making processes.

Parents believe that teacher-assigned grades at the secondary level represent achievement and trust these grades to be a representation of how well their children are doing in school (Brookhart, 1994; Pilcher, 1994; Randall & Engelhard, 2010). The literature suggests that grades consist of various factors which convolute their meaning

and therefore misinform the parents (Brookhart, 1994; Cross & Frary, 1999; Guskey, 2006), leading to confusion when their children arrive at college unprepared.

Lastly, understanding the assessment and grading practices of secondary school teachers provides insight into the discrepancy between grades, high school graduation rate increase, and the production of college-prepared graduates. It also will reveal the areas of need in training pre-service teachers on grading and assessment because the grading practices they experience as students may influence the practices they employ as teachers (Allen, 2005). The results will also inform college admissions counselors on the accuracy of grades to predict the performance of first-year students (Allen, 2005; Clinedinst, Korateng, & Nicola, 2015; Cyrenne & Chan, 2012).

Practical Framework

Grading practices were first developed with the intention of documenting learning, or student achievement, and the grades assigned to students are meant to be a representation of the degree of a student's achievement (Brookhart 2015). The parents of students also share the belief that the assigned grades represent achievement (Brookhart, 1994; Pilcher, 1994; Randall & Engelhard, 2010). While teachers and students believe the same, they also indicate other factors, such as effort, can also influence grades (Brookhart 1994). Regardless of the advice provided by research and grading experts, teachers continue to use multiple variables, and achievement and non-achievement factors to determine final grades (Randall & Engelhard, 2010).

Allen (2005) indicates grading is a function of many variables rather than expressing the degree of knowledge in a content area. The way grades are assigned seems

to vary, based upon local agreements made between teachers and students and are motivated by external factors dictated by the accountability system in place (Willingham, Pollack, & Lewis, 2002). The standards-based reform (SBR) movement returned the focus of assessment and grading to performance which has led to increased achievement (Hamilton, et al., 2008, December), and a continual increase in graduation rate (National Center for Education Statistics, 2015; North Carolina Department of Public Instruction, 2016), and climbing GPAs (Nord et al., 2011). The resulting changes arising from the SBR suggest that grading practices are better aligned to achievement, yet students are still not prepared for post-secondary studies (Clinedinst, et al.,, 2015; Cyrenne & Chan, 2012).

The assumption is that grades communicate the same information and mean the same thing across classrooms and schools and because of this, grades are often accepted at face value and not viewed critically (Schneider & Hutt, 2014). This assumption is a direct result that teacher-assigned grades and the standardized assessments produced by the standards-based reform measure the same thing, when in fact they vary widely, having a moderate to weak relationship (Brookhart, 2015). In some classes, grades are an accurate reflection of achievement, but in others, this is not the case (Brookhart, 2015). Recent research suggests (see Bol, Stephenson, O'Connel, & Nunnery, 1998; Cross and Frary, 1999; McMillian, 2001; Pilcher, 1994; Willingham et al., 2002) that significant relationships exist among subject-area teachers and assessment and grading practices, which can be used to examine the disconnect between the continued rise in the number of graduates and the lack of preparedness in post-secondary studies.

Methodology

The research questions were examined using a survey research design method to target high school teachers in school districts from the Piedmont and coastal region of North Carolina. An instrument developed by McMillian (2001) was modified for this study. Dillman, Smyth and Christian's (2014) Tailored Designed Method (TDM), which is based on social exchange theory and is used to develop the trust of participants by emphasizing the benefits of participation aimed at reducing the overall survey error, was used to implement the survey procedures.

Limitations

The present study focused on eight of the one-hundred fifteen school districts throughout North Carolina, representing approximately seven percent of the state's districts and total high schools. The demographics of the participants also vary significantly from the reported demographics of the state's teachers, with 93.1% of the participants being white as opposed to 49.5% represented statewide (Division of School Business, Financial and Business Services & North Carolina Department of Public Instruction, 2016, January). Taking these factors into consideration, the results discussed may not accurately represent North Carolina's teachers use of assessment and grading factors and therefore should be cognizant of these concerns in the discussion to follow.

In addition, the survey relies on self-reported assessment and grading practices occurring during the spring semester of 2018 and may result in teachers reporting socially acceptable practices rather than those implemented, even after the assurance of

confidentiality. Also, all items in the survey were required to be responded to which could have resulted in a lower return rate should a participant not be willing to answer a single question. Lastly, because the survey was distributed prior to the end of the semester, it may not accurately represent the assessment and grading practices used to determine the final grades after submission.

Delimitations

The researcher elected to limit the review of the literature surrounding assessment and grading practices after 1994 due to the political influence of the standards-based education reform, marked by the passage of Improving America's Schools Act (IASA). The IASA began the accountability movement and the emphasis on producing globally competitive students prepared for post-secondary studies and careers. Additionally, the target population of this study is focused on eight districts in the Piedmont and coastal regions of North Carolina, due to the convenience of access to the researcher. Only the four core content areas are being examined in this study, not other subject matter areas. Also, only the factors included in the survey, both achievement and non-achievement, were examined and may not be all-inclusive of factors that impact assessment and grading practices.

Summary

The current non-experimental study examined the assessment and grading practices of English, math, science, and social studies high school teachers in the Piedmont and coastal regions of North Carolina. The study provides current information

to be used in defining grades and gives insight into the discrepancy between the steadily increasing graduation rate and the production of graduates that are unprepared for post-secondary studies. In the next chapter, a thorough review of the literature will begin with the historical development of grading in the United States, discussion of state and federal legislative and policy changes, and will conclude with the summary of assessment and grading practices at the secondary level.

Chapter three will provide a detailed description of the methods, procedures, a description of the instrument and participants of the proposed study. Chapter four will be a review of the research design, instrumentation, data analysis, and a summary of the results. Chapter five will provide a summary and discussion of the results, implications for practice, and suggestions for future research. At the conclusion of chapter 5, the references used in the development of the dissertation and the appendices referenced throughout will be included.

Chapter 2: Review of Literature

The subjectivity of a single grade assigned by teachers to document student learning has been criticized as an adequate measure for more than 100 years (Brookhart, 2015), even predating standardized grades (Curreton, 1971). The grading system used presently has been altered throughout history due to political influences, changes in technology, and changes in laws impacting education. Although the system is standardized, teachers often have differing belief systems regarding assessment and grading influencing how the system is implemented within their classrooms (Kunnath, 2017; McMillian, 2001; Randall & Engelhard, 2010), which convolutes the meanings conveyed by the grades assigned.

This study updated empirical evidence surrounding assessment and grading practices at the secondary level to address the gap between increasing graduation rates and the preparedness of incoming post-secondary students. The study provides a comprehensive description of assessment and grading practices within the four core content areas, English, math, science, and social studies, of current classroom teachers and identified meaningful relationships between the core content areas, ability levels, and grade levels of students.

The following literature review examined current grading system at the secondary level with a primary focus on the core content areas, English, math, science, and social studies. The review begins with a historical examination of the development of the assessment and grading system used today, major educational reforms impacting

assessment and grading practices, and current research, post-1994. The results provide a baseline to examine the variation of grading between the core content areas and to develop the meaning of grades at the secondary level.

The Development of an Assessment and Grading System

The first formal test on record, implemented in China in 210 BC, the Letters examination was a pass or fail examination with both objective, performance-based, and subjective, oral responses, portions. The success a student had on this exam dictated his/her future role in society and the subjective nature of the knowledge portion raised concerns with government officials. The officials responded by standardizing the responses to the oral exam to eliminate any bias and deviations from the responses they deemed appropriate (Madaus & O'Dwyer, 1999). Since the inception of assessment and grading, the validity of these practices has been a persistent problem, and politicians and reformers alike have sought improvement.

During the Middle Ages leading to the 20th Century, Madaus and O'Dwyer (1999) discovered the assessment system began transitioning from a qualitative view to a quantitative one emphasizing the trend from subjective to objective testing. The development of the factory system in Europe during the 14th Century further pushed the trend of objectivity because the formerly skilled master craftsman no longer created individual pieces, but rather mass replication requiring the uniformization of a task to allow for simple verification of a skill with minimal subjectivity (Madaus & O'Dwyer, 1999). The factory model influenced the education system in Europe and produced the

catechetical method of instruction and assessment, very similar to methods the United States use today.

> The catechetical method emphasized three things,
>
> first, the domain of knowledge is delimited – that is, small enough, so that sampling from it to assess achievement is relatively easy; second, based on a transmissionist view of learning and instruction; and third, instruction and assessment are seamlessly linked in an ongoing formative way (Madaus & O'Dwyer, 1999, p. 691).

Through this method and the advent of the quantitative score in the late 18th Century, concerns surrounding subjectivity of assessments began to diminish (Madaus & O'Dwyer, 1999). The ability to collect data from assessments allowed instruction to be modified to meet the needs of the students, average scores could be used to document progress, and students could be ranked based upon their performance.

As the nature of assessment and grading began to evolve, the European model transitioned to achievement and competition rather than the pedagogical function in the catechetical method and was exemplified best by the Cambridge Mathematical Tripos. The Tripos was a multi-day high-stakes exam in which students would compete head-to-head to test their knowledge until a single victor emerged. The students' academic ranking would determine their placement in the tournament bracket, and as they progressed through the examination, they would face more difficult questions and opponents, to compete for the highest marks of distinction that separates them from their peers (Schneider & Hutt, 2014). Those who performed well generally were very successful financially after leaving Cambridge with the top student earning a portion of

Cambridge's endowment, but those who did poorly were marked for the remainder of their lives (Schneider & Hutt, 2014). The European model of assessment and grading, both in the secondary and post-secondary realm, emphasized the ranking of students based upon achievement tying education directly to the degree of success in life.

Grading and Assessment in the United States

The European idea of grading and assessment transitioned to Massachusetts, approximately in 1709, in the form of oral evaluations (Madaus & O'Dwyer, 1999) but were largely used for pedagogical purposes (Schneider & Hutt, 2014, p. 202) linking assessment to teaching and learning, differing from their European counterparts of the time. Early educational reformers eventually adopted the French multilevel, age-graded and time-constrained learning, educational system which we know today near the turn of the 19th century (Hadley, 1902; Rossides, 1984).

In 1845, Horace Mann revived the emphasis on linking assessments to learning by introducing the written exam to increase the efficiency of assessments, hold the schools accountable for student learning, and to create grade-level gateway exams to prove students are prepared to move on to the next step in learning (Madaus & O'Dwyer, 1999). The move was politically driven and used testing to "promote the social movement of industrial capitalism's developing commitment to standardization, uniformity, precision, clarity, quantification, and rational tactics" (Madaus & O'Dwyer, 1999). The accountability trend shifted to the classroom where student learning and assessment began to link with the first summative test given, in the United States, in a content area, a spelling test in 1887, rather than a general end of grade exam (Madaus &

O'Dwyer, 1999). As testing and educational assessments began to be used more frequently, the necessity to develop a way to document student learning became evident (Brookhart, 2015; Schneider & Hutt, 2014).

The use of gradebooks was first documented around the mid-1820s in colleges and some secondary schools, and they often went beyond including achievement grades by incorporating items such as whether they attended chapel, class, and student behaviors (Schneider & Hutt, 2014). The original gradebooks used narrative descriptors of student performance and did not contain quantitative grades until after the middle of the 19th century (Brookhart, 2015; Durm, 1993). Grades at the secondary level began to be standardized in the 1870s and were used as a ranking system that indicated achievement, served to motivate and prevented excessive competition among students through periodic calculations (Schneider & Hutt, 2014), which resemble the reporting periods in current secondary schools.

The concerns for appropriate assessments and grading became a larger focus as educators sought to improve the assessment and grading practices within their classrooms in the early 20th Century (Madaus & O'Dwyer, 1999). Teaching seminars encouraged adding points for general ability if the grade was not a *fair assessment* of the student's general work because giving a student, who is very ambitious, an average score would break his/her heart and encourage laziness (Curreton, 1971). The adjustments of grades, as suggested by these seminars, began to skew the grades from the intended purpose of objectively reporting achievement and using assessment information in a formative way.

The end of the 1920s brought the industrial era, compulsory attendance laws, and the ideology of the assembly line into the realm of the public education system (Stiggins, 1991) much like the impact the factory system had on education in Europe. Students were treated as products on an assembly line, learning was bound to a timeline, one school year, while adding components, specific units or topics, at various intervals as they progressed along the continuum (Stiggins, 1991). Grades signified the quality of progress and the schools served as the quality control system indicating which student would go into the factory or onto post-secondary studies (Stiggins, 1991). The previously detailed accounts of written student performance teachers gave to students documenting their performance in early grading became increasingly difficult as school populations increased due to the enforcement of compulsory attendance laws. Transient students added another level of difficulty with secondary schools being widely separated geographically, making it increasingly difficult to communicate, so school officials began to require a method for student tracking (Schneider & Hutt, 2014). However, "in order for [grades to be used for communication], they would have to mean something similar across classrooms, and ideally across schools" (Schneider & Hutt, 2014, p. 209).

The A-F, 4.0 scale, and the 100-point scale, first developed in the American University system, were fused and used at the secondary level mid-twentieth century to create the standardized grading system we know today while also solving the needs of the schools at the time (Schneider & Hutt, 2014). The standardized grading system was created to measure achievement, aligning with the distribution of intelligence on standardized exams, such as IQ tests, placing "C" in the middle to represent the average. Although the grades were intended to reflect the purpose of grading, measuring

achievement and linking assessments to instruction, non-academic factors began to impact grading and caused a shift in the *typical* distribution of grades in the 1970s, so that at least half of the students, according to Furguson and Massey (1975) earned an A or a B, (as cited in Schneider & Hutt, 2014).

Schneider and Hutt (2014) discovered that by the 1970s grading was not producing learning, but instead encouraged cheating, led to the narrowing of the curriculum, and instruction was loosely aligned to the items on summative assessments, which was an expressed concern of many educators of the time. The concerns are compounded when realizing the importance of accurate grading practices because the grades students received in high school became, and still are, vitally important for college admission (Schneider & Hutt, 2014). As a result, grades became the currency of the educational process, and the students learned to *play school* by exerting the least amount of effort to earn the highest grade possible, which was often dependent upon the course and teacher selection, because of varying assessment and grading practices (Schneider & Hutt, 2014).

The originally developed grading system that mirrored the catechetical method emphasizing delineated content that linked assessment in an ongoing and formative way to instruction had slowly deteriorated to include factors unrelated to achievement. The results are grades that have blurred meanings and difficult to use for the documentation of learning. Nevertheless, teacher-assigned grades are still perceived throughout the history of grading to be a representation of student achievement (Brookhart, 2015).

Political Influence of Assessment and Grading

The publication of *A Nation at Risk* in 1983 set the stage for educational reform in the United States and the concerns regarding assessment and grading practices became evident. The United States, according to the report, was failing to increase global competitiveness and producing students who were unable to think critically. It reported that "nearly 40 percent [of secondary students] cannot draw inferences from written material; only one-fifth can write a persuasive essay; and only one-third can solve a mathematics problem requiring several steps" (National Commission for Excellence in Education, 1983, p. 9). The report identified one area of deficiency as the grading practices of teachers.

The National Commission for Excellence in Education (1983) defined grades as the "degree to which students demonstrate their mastery of the subject matter" (p. 17). They discovered that while grades have continued to increase, the level of achievement, as measured by standardized tests, has declined. The report suggests that the teachers' assessment of student work is inaccurate, the curriculum is not aligned with the standardized assessments used to determine achievement, and the overall calculation of a grade is not in alignment with its stated purpose. Thus began the standards-based reform movement.

The standards-based reform sought to improve the educational system in the United States by increasing academic expectations for students, aligning instruction and content, by assessing and documenting student achievement through standardized assessment, and holding the schools accountable for student learning (Hamilton et al.,

2008). The reform emphasized the idea that assessments, in addition to accountability, should inform instruction and it produced efforts to improve the systems and methods used to track student progress (Hamilton et al., 2008). The standards-based reform essentially sought to return the assessment and grading system to the catechetical methods originally brought to the United States from Europe.

Improving America's School Act

The reauthorization of the Elementary and Secondary Schools Act in 1994, also known as the Improving America's Schools Act (IASA), sought to "promote the alignment of all education components—curriculum and instruction, professional development, school leadership, accountability, and school improvement…to ensure that all children can attain challenging standards" (Riley, 1995, September, para. 4). The IASA changed education at the secondary level by requiring the states to adopt uniform standards that address higher level thinking skills in math and English language arts and to link this to student performance on state assessments (Education Week, 1994; Riley, 1995, September). The IASA also required the states to devise a system to monitor the attainment of *adequate yearly progress (AYP)* using disaggregated data on student achievement from state-developed assessments and to identify the schools who did, or did not, meet AYP (Education Week, 1994). The passage of the IASA created a high-stakes environment where teachers must focus their instruction and adequately assess student achievement to monitor student progress toward meeting the end-of-year benchmarks.

No Child Left Behind

The passage of No Child Left Behind (NCLB) in 2001 expanded IASA through requiring the adoption of state standards in all core content areas, and the development of a standardized assessment in math, English, and science at the secondary level (Hamilton et al., 2008). NCLB also included strict sanctions for schools whose students did not meet the expected AYP but left the state to set the guidelines regarding proficiency, most likely because more than half of the states did not meet the original requirements of IASA (Hamilton et al., 2008).

North Carolina's Educational Reform

Shortly after the publication of *A Nation at* Risk, North Carolina implemented a standards-based reform of its own, like IASA, in the 1980s using its own funding (Hamilton et al., 2008) and by 2012 had expanded its accountability model, under *The Ready Initiative,* to include students, teachers, and schools alike (North Carolina Department of Public Instruction, 2015, January). North Carolina's School Performance Grade, approved in 2013 as part of the Ready Initiative (G.S. §115C-83.15), uses a combination of achievement, 80%, and student growth, 20%, to calculate a final school grade expressed in the traditional letter-percentage combination, A-F on a 10-point increment scale. Student achievement calculations included standardized scores on the End-of-Course exams in English II, Math 1, and Biology, the ACT and ACT WorkKeys tests, and 4 and 5-year graduation cohort rates.

The school report card grade aligns somewhat with the traditional meaning of student grades, documentation of achievement, but also considers students may learn a considerable amount yet not achieve proficiency on the standardized assessments, i.e.,

student growth. The use of letters and percentages suggests that the grades assigned to schools should correspond to student assigned grades, thus communicating the same information. Given the increased demand for measuring and communicating student achievement, it would be expected that the assessment and grading practices of North Carolina Teachers would reflect achievement and student learning.

Student accountability and grading. The student accountability model, North Carolina State Board of Education policy GCS-A-016, requires that, at the secondary level, the End-of-Course, North Carolina Final Exams, and teacher-made final exams count as 20% of a student's final grade for a course, while the other 80% remains at the discretion of the classroom teacher's assessment and grading practices. This policy attempts to address concerns of the subjectivity in grading and re-emphasizes that grades reflect the degree of student achievement and forces teachers, to some degree, to adhere to this when calculating final grades.

Other concerns with current grading systems are reflected in North Carolina State Policy GCS-L-004, which documents the change in the grading system at the secondary level to a 10-point scale in 2015 instead of the 7-point scale adopted in 1995, coming shortly after the passage of IASA, and removed district variability in grade point average (GPA) assignments for letter grades by limiting the correspondence to whole numbers only, such as an A is a 4.0, and a B is a 3.0. These changes were necessary to create uniformity across the state by standardizing the grading scales, allowing easier grade transfers between schools, to address concerns with student competitiveness from surrounding states, and to align grading methods with the system devised by the

University of North Carolina and NC Community College System (North Carolina State Board of Education, 2014, October).

Although the legislative and policy changes seek to address concerns surrounding grading variability, the effect, barring no changes in current practices, would raise student GPAs, reduce the ability for it to appropriately rank students using GPAs, and further separate grades from the historically established intentions, to document student mastery. It is obvious that throughout history the validity of assessment and grading has been a concern and frequently addressed by grade reformers and politicians alike, but "is it the best we can accomplish? Probably not" (Durm, 1993, p. 297).

Results of the Standards-Based Reform

The standards-based reform led to experimentation with innovative forms of assessment that would be sensitive to high-quality, cognitively challenging instruction, but eventually gave way to standardized, time-efficient forms (Hamilton et al., 2008). Student achievement in most states has risen since the implementation of NCLB, but Hamilton et al. (2008) suggest that the progress is artificially inflated because teachers are preparing students for taking a standardized assessment rather than improving instructional practices in the classroom. The authors note these standardized assessments only measure narrow portions of the curriculum, and that the less frequently tested standards may not be taught without it being apparent, or what is considered *teaching to the test.* The *teaching to the test* is a result of the sanctions associated with schools who are low performing and the modification of current practices to meet the demands of the higher standards demanded by NCLB (Hamilton et al., 2008).

In summary, although the purpose of IASA and NCLB was to increase rigor and student learning in current practice, it produced a testing-based reform in education instead and did not address any aspects of grading validity which has been a continued concern throughout the history of assessment and grading. The standards-based reform altered assessment practices in the classroom by de-emphasizing the use of formative assessment due to a "distrust of teachers and a belief that external testing will on its own, improve learning" (Black & Wiliam, 1998, p. 5). As a result, the standardized paper-and-pencil, multiple-choice exams, backed by scientific results, became the political tool to hold schools accountable because assessments and grading practices used do not adequately document student learning (Stiggins, 1991).

Graduates and Remedial Courses

The graduation rate in North Carolina has increased nearly 18% (North Carolina Department of Public Instruction, 2016) and 10% nationally (National Center for Education Statistics, 2015) between 2006 and 2015 but yet more than half of the students have to take remedial courses (National Center for Public Policy and Higher Education and The Southern Regional Education Board, 2010) even though 86% of students believe they are prepared for college (Center for Community College Student Engagement, 2016). The traditional metric, GPA, is considered an accurate representation of a student's preparedness for post-secondary work but those who earn better grades are not necessarily more prepared; forty percent of students who report having an "A" average still have to take remedial courses (Center for Community College Student Engagement, 2016).

Remedial courses are defined by the University of North Carolina General Administration (2015, February) as:

> Courses in reading, writing, or mathematics for college-level students lacking those skills necessary to perform college-level work at the level required by the institution. Students participating in remedial education while in college may not earn credit toward their degrees by completion of these courses (p. 9).

The criteria for those required to take remedial courses varies among post-secondary schools within the University of North Carolina (UNC) system and the North Carolina Community Colleges System. The UNC system, consisting of seventeen universities, provided remedial courses in eight of its locations with seven in math and five in English during the 2012-2013 school year. In those schools offering remedial math, see Table 1 (University of North Carolina General Administration, 2015, February), five use standardized assessment data from students' applications, SAT and/or ACT scores, and placement test scores, either the ACCUPLACER or internally developed tests, or rely solely on placement tests scores. For Remedial English, see Table 2, only two schools use application data, either SAT scores or SAT scores and GPA in combination, and placement tests, such as the ACCUPLACER or internally-developed tests, whereas the remaining three solely use placement tests (University of North Carolina General Administration, 2015, February).

Table 1

Data to Determine Remedial Math Placement

University	Application Data	Test
ASU	SAT	Internal Test
ECU	SAT/ACT	ACCUPLACER
ECSU	SAT	ACCUPLACER
NC&T	SAT/ACT	Internal Test
NCSU	SAT	Internal Test
UNCC		Internal Test
WSSU		ACCUPLACER

Table 2

Data to Determine Remedial English/Reading Placement

University	Application Data	Test
ASU	SAT	Internal Test
ECSU		ACCUPLACER
NCCU		ACCUPLACER
UNCP	SAT, High School GPA	
WSSU		ACCUPLACER

The North Carolina Community College System (NCCCS) has changed its policies around the criteria used in determining who is required to take developmental education courses in 2013 and was updated in 2015. The State Board of Community Colleges approved the Multiple Measures for Placement policy, rather than using placement tests such as the ACCUPLACER or COMPASS, and instead exempting students within five years of graduation from a North Carolina High School, whom have taken four eligible math classes, requiring Algebra 2, and earning an overall unweighted GPA of 2.6 or greater (State Board of Community Colleges, 2015, February 20). Those who do not meet the GPA requirements can be exempted if they meet the college-ready standards on the SAT or ACT, but if they do not, they are subject to the required placement tests (State Board of Community Colleges, 2015, February 20). The decision

to use GPA, in addition to the math courses taken, is because it was determined that "GPA is a cumulative measure earned while taking multiple subjects over multiple years, is a strong indicator of students' motivation, study habits, and persistence" (SuccessNC, 2013, p. 35), which is contradictory to the traditional meaning of grades.

Remedial Course Enrollment and Impacts

Nationally, approximately forty percent of students at public four-year colleges and 68% of community college students require at least some developmental education and take at least one remedial course (Chen & Simone, 2016, September). Jaggars and Stacey (2014, January) found "twenty-eight percent of community college students who take a developmental education course go on to earn a degree within eight years" (p.1) as "compared to 43% of students who don't take remedial course" (Attewell, Lavin, Domina, & Levey, 2006 as cited in Jaggars & Stacey, 2014, January, p. 6). While Chen and Simone (2016, September), reported that approximately 45% of public four-year college students, who took remedial courses, earned a certificate or degree as opposed to 71% percent of students who were not required to take these courses. When considered collectively, approximately 37% of students that have taken at least one remedial course earn a certificate or degree within six years of their initial enrollment.

In examining the persistence rate of students required to take more remedial courses prior to enrolling in credit-bearing courses, they generally complete the sequence at a lower rate than those required to take less while those in two-year programs, required to complete three or more courses, do so more frequently than students in a four-year program, see Table 3 and Table 4 (Chen & Simone, 2016, September). While exploring English and math remedial course completion rates, see Table 5 and Table 6 (Chen &

Simone, 2016, September), the two have similar numbers although students who take courses in math do so at a lesser rate. Nationally, the overall impact of remedial courses on certificate/degree complete is most pronounced in those students who are required to take more remedial math courses. Approximately half of the students requiring more than two remedial courses do not complete the sequence and never enroll in a college level course, while those who are required to take remedial English are more likely to complete the sequence as opposed to those who required Math (Chen & Simone, 2016, September).

Table 3

Public 4-Year College Overall Remediation Completion Rate

# of Remedial Course	Completion – None	Completion – Some	Completion - All
1	20.2%	N/A	79.8%
2	12.9%	29.1%	57.4%
3	12.6%	53.9%	33.5%
3+	6.5%	76.8%	16.17%

Table 4

Public 2-Year College Overall Remediation Completion Rate

# of Remedial Course	Completion – None	Completion – Some	Completion - All
1	26.7%	N/A	73.3%
2	17.4%	27.7%	54.8%
3	8.0%	15.7%	40.4%
3+	8.9%	66.1%	25%

Table 5

Public 2-Year College Math/Reading Remedial Completion Rate

Remedial Subjects	Completion – None	Completion- Some	Completion - All
Math	20%	29%	50%
Reading/English	21%	46%	63%

Table 6

Public 4-Year College Math/Reading Remedial Completion Rate

Remedial Subjects	Completion – None	Completion- Some	Completion - All
Math	20%	22%	58%
Reading/English	21%	15%	64%

In the North Carolina University System (UNC System) the total enrollment in remedial instruction of freshmen who were recent graduates included 1,749 students in the 2013-14 school year (University of North Carolina General Administration, 2015, February) representing 5.8% of the total recent high school graduating freshman class (University of North Carolina System, 2013). According to the University of NC Remedial and Development Activities Report, the total enrollment in remedial English only included 363 students, 1,244 in math, and 142 in both English and math (University of North Carolina General Administration, 2015, February). While these numbers may be expected to be evenly distributed across the schools in the UNC System offering remedial courses to recent graduates, it is evident that the universities have varying needs with some only offering either English or math remedial courses while others offer both, see Table 7. Also, the data reveals that some students are required to take multiple sections of one subject or at least one section from each subject as evidenced by the enrollment numbers in Table 7 (Data from The University of North Carolina General Administration, 2015, February).

Table 7

UNC System Remedial Courses by School and Recent Graduate Enrollment

School	Recent Grad Fr	Total Remedial	Remedial Math	Remedial English
ASU	2,782	180	151	34
NCCU	856	151	0	0
ECU	4,297	800	818	0
ECSU	266	254	206	244
NC&T	1,690	422	0	437
UNCC	3,883	902	0	937
UNCP	985	160	163	0
WSSU	684	324	140	330

Note. Total remedial is unduplicated, Math and English are Duplicated. Recent Grad Fr = recently graduated college freshman

Aligning with national statistics, North Carolina post-secondary students, enrolled from 2009-2015, who took any remedial courses were found to earn a degree within six years, 49.4%, which is more than 20% lower than those who were not required to take any, while those taking both English and math had the lowest graduation rate, 43.2%. (University of North Carolina General Administration, 2015, February). Remedial English students had the lowest retention rate, 79%, compared to 87.6% of academically ready peers, while students who only were required to take remedial math were most successful in retention and graduation rates (University of North Carolina General Administration, 2015, February). Although the data here is representative of the entire UNC System, it is obvious that the majority of students enrolling are prepared for college-level work suggesting that the students who are generally underprepared for post-secondary work would enroll in North Carolina community colleges.

"In 2011, 69% of recent high school graduates placed into at least one developmental course when they enrolled in a North Carolina community college" (SuccessNC, 2013, p. 32). According to Clotfelter, Muschin, and Vigdor (2015), North

Carolina graduates, subsequently enrolling in a North Carolina community college between 2001 and 2009, were required to take developmental math 48% of the time while English accounted for a somewhat smaller amount, 37%. The authors also add that of the remedial math students enrolled, nearly 40% never enrolled in a college math course while more than 60% of remedial English students never enrolled in a college English course. The students who persisted on to the college level courses, although successful in the remedial courses, were unable to pass these courses 68% of the time in math and 45% for English (Clotfelter, Muschin, and Vigdor, 2015). A particularly concerning trend in math, according to Bailey (2010), is that students who fall three or more levels below college readiness standards only 8% complete a college-level course and 23% never enrolled in the first remedial course (as cited in SuccessNC, 2013, p. 32).

College Ready Graduates

Prior to 2013, the majority of students seeking enrollment in North Carolina community colleges and the UNC System relied heavily on placement tests and standardized test scores, but now greater responsibility is being placed upon the high schools in preparing students for post-secondary studies. In 2015, North Carolina passed Career-and College-Ready Graduates (CCRG), *S.L 2015-241* (*see Sec. 10.13),* but it was amended, Senate Bill 99, in 2018, delaying the implementation until the 2020-2021 school year. The CCRG legislation introduces developmental math and reading courses into the senior year of high school to allow students to be exempted from taking remedial courses post-secondarily should they meet the requirements in Table 8 below (State Board of Community Colleges and the State Board of Education, 2018, March). Students earning above a 2.79 GPA are exempt from the ACT College-Ready benchmark

requirements and are not required to take remedial courses. In addition, the State Board of Community Colleges must approve the curriculum to be used in the high school remedial courses, provide training to the high school faculty regarding the implementation of the courses, and will provide general oversight regarding the instructional delivery of these courses (State Board of Community Colleges and the State Board of Education, 2018, March).

Table 8

Criteria to Exempt Remedial Courses per CCRG

Subject	Unweighted GPA	Test Score
English	2.2 - 2.7999	ACT English 18+ or ACT Reading 22+
Math	2.2 - 2.7999	ACT 22+

Note. Applies only to students in the identified GPA range and to be exempt must earn the associated test score. Those earning 2.8 or higher are exempted on GPA alone.

As part of North Carolina's READY Initiative, the ACT became a required component of North Carolina's school accountability model in the 2011-2012 school year, and all high school Juniors are required to take it. The performance on the ACT is concerning given the gap between the continually improving graduation rate and the lack of improvement on the standardized assessment. According to ACT (2017), from 2013 to 2017, 46% of North Carolina graduates met ACT's college readiness standards in English, 33% in Reading, and thirty-two percent in math. It was also reported in that in 2017, 47% of North Carolina's graduates met zero of ACT's College Readiness Benchmarks (ACT, 2017), see Table 9. The data indicate that more than half of high school graduates in the

fall of 2018 could be required to take remedial courses pending GPA and could influence grade inflation to avoid the remedial course offerings at the high school or encourage more students to play school to avoid taking these courses.

Table 9

North Carolina ACT Performance Data 2013-2017

Subject	2013	2014	2015	2016	2017	5yr Avg.
English	43	47	47	47	46	46
Reading	31	30	34	34	36	33
Math	33	33	32	32	30	32
Science	25	23	26	26	27	25
All Four	17	16	18	18	18	17

Note. Percentage of all NC students meeting or exceeding college-ready standards

Summary

The requirement of remedial coursework in post-secondary studies does not ensure these students will matriculate through to graduation at the same rate as those who are not required to take the courses. Many students depend on financial aid to pay for their education which can cause remedial students to reach the 30-hour credit maximum for developmental education, or 150-percent ceiling on total credit hours attempted in a program of study forcing these students to drop out, while others required to take non-credit bearing courses may never enroll (SuccessNC, 2013). The cause of the need for remediation is that the students have graduated "without all of the core skills will need to ably succeed at the next level (ACT, 2017). Now, due to *Sec. 10.13* of *S.L 2015-241* the responsibility falls squarely on the high schools of North Carolina to ensure the students are prepared for college-level work, and the degree of success of future students hinges on the ability of teachers to "raise the bar on learning standards and focusing on the essential skills necessary for success" (ACT, 2017).

Grading practices are more important than ever before; it is necessary that they appropriately measure student performance because the more students who do not meet the college-ready standards it further limits the availability of staff, elective course offerings for all students, and further limits resources a high school has. Nevertheless, "Far too many students continue to graduate from high school without all of the core skills they will need to ably succeed at the next level" (ACT, 2017, p. 8).

Research on Assessment and Grading

Guskey (2006) describes the current grading system of classroom teachers as *hodgepodge grading* because each grade is calculated by averaging student performance related to various items, frequently varying between teachers, such as attendance, behavior, formative assessments, group projects, extra credit, effort, and tests. Many of these items do not provide any insight into the learning process, and when they do, it rarely monitors student progress and is almost always included in the final grade, even when they are still struggling with the concepts (Frey & Schmitt, 2010). As a result, the final grades assigned to students have blurred meanings and are difficult to interpret.

Black and Wiliam (1998) indicate three issues with assessment and grading practices that contribute to the current system's lack of ability to adequately communicate student learning to parents, students, and other teachers (Guskey, 2001, Guskey, 2006, Muñoz & Guskey, 2015; Scriffiny, 2008). First, students are not learning effectively due to using low-level thinking and measuring the recall of facts rather than the degree of understanding, even when the teachers state this as their purpose. Furthermore, teacher-developed assessments or the use of assessments that have not had their items critically

reviewed to ensure they measure what is intended further adds to the inconsistencies found in grades (Black & Wiliam, 1998).

Secondly, grades tend to have a negative impact. The system emphasizes the sorting of students, grades serve to compare creating a competitive environment, leading to students believing they cannot learn, and separating those who *know* from those who *do not know* (Anderson, 1998). The single grade alone, a percentage or letter, does not provide feedback as to what can be done to improve learning. This system separates process from product, effectively hiding patterns of learning preventing the identification of areas of student learning difficulties (Anderson, 1998; Ornstein, 1994; Scriffiny, 2008), and hinders the ability of grades to be used in pedagogical decisions which is contradictory to the intended purpose of grading in the United States (Madaus & O'Dwyer, 1999; Schneider & Hutt, 2014).

Thirdly, grades also serve as a social and managerial function, at the expense of learning (Black & Wiliam, 1998). The assignment of a grade takes precedence over its analysis to inform instruction, they are used as a payment to control classroom behaviors and thought to hold students accountable for their academic actions, such as the completion of homework (Pilcher, 1994; Schneider & Hutt, 2014).

Guskey (2006) adds that the current grading system is flawed because there is not a clear definition of the purpose of grading and as a result, two students earning the same grade can have vastly different levels of content achievement. Muñoz and Guskey (2015) define the purpose of grading "is to describe how well students have achieved the learning objectives or goals established for a class or course of study" (p. 65) and emphasize that grades are most effective only when achievement data are used.

Ironically, the intended purpose of grading has not changed throughout history but what teachers use to determine final grades has changed.

Formative Assessments

The term *formative assessment* does not refer to any instrument specifically but rather a process developed by the teacher, which can involve the use of student self-evaluation, observation, discussion, and written work (Black & Wiliam, 1998; Curry, et al., 2016). Formative assessments are used to guide the student learning processes (Wiliam, 2011), provide substantial feedback to students regarding their thought process occurring during instruction, and used to adjust instruction to meet the needs of each student (Black & Wiliam, 1998; Black & Wiliam, 2003; Curry et al., 2016; Hongli, 2016) rather than to be used to document achievement (Black & Wiliam, 2003; Wormeli, 2006). The purpose and the process of formative assessment are not revolutionary; they are "at the heart of effective teaching" (Black & Wiliam, 1998, p. 2) and reflective of one of the early purposes of grading (Schneider & Hutt, 2014).

Research supports the use of formative assessment to raise the achievement of students on standardized tests (Black & Wiliam, 1998; Black & Wiliam, 2003; Curry et al., 2016; Hongli, 2016; Wiliam, 2011), such as those used in accountability models. Black and Wiliam's (1998) review of 20 studies, ranging from 5-year-olds to university graduates around the world across several subjects and countries, indicate that improved formative assessment implementation, across all grade levels, has an average effect size on achievement between 0.4 and 0.7. To explain this further, "an effect of 0.4 would mean that the average pupil involved in an innovation [strengthening formative assessment use] would record the same achievement as a pupil in the top 35% of these

not so involved. An effect size gain of 0.7…. would have raised the United States' national math score from the middle of 41 countries to the top 5" (p. 141). Wiliam (2011) also identified effect sizes ranging from 0.32 to 0.96 indicating that formative assessment substantially increases student learning in the classroom. Despite this research, Frey and Schmitt (2010) found only 12% of the assessments given by teachers were truly formative in nature, and the clear majority replicate the tests used in the accountability system produced by IASA and NCLB, therefore limiting the data available to inform instruction.

North Carolina's Department of Public Instruction (NCDPI) realized the benefits of improving the use of and informing the teachers within the state about formative assessments. NCDPI developed the Formative Assesment Learning Community's Online Network (NC Falcon) introduced in June 2010 which provided professional development to North Carolina Teacher's "to develop a basic understanding of formative assessment and illustrate the role it should play in a comprehensive, balanced assessment system" (North Carolina Department of Public Instruction, 2010, May, p. 1). The professional development of NC Falcon emphasized the use of formative assessments and provided resources and support to administrators, as well as teachers, to increase the frequency and quality of these assessments within each school throughout the state. The intention was to target student learning by encouraging teachers to use formative assessments as interim benchmarks prior to administering the summative assessments. Although this system has been in place since 2010, there has not been any empirical studies completed to document the use of formative assessments in North Carolina, but it supports the idea that the teachers of the state may use formative assessments at a higher rate.

The use of formative assessment complements high stakes assessments forcing teachers to reflect on the effectiveness of their instruction (Curry et al., 2016), necessary to break the low-performance testing cycle (Black & Wiliam, 1998), and reduce the achievement gap (Hongli, 2016). Ultimately, the use of formative assessment will produce a student who is better able to think critically, retain more information, and, as a result, the standardized assessments, which emphasize higher-order thinking skills, will show more gains in achievements than the use of summative assessments alone (Black & Wiliam, 2003). The use of formative assessments in this manner, increases the alignment of assessment and grading with its intended purpose by removing information that does not accurately reflect the degree of mastery at the end of a unit.

Summative Assessments

Summative assessments are defined as an evaluation of learning and instruction at the end of teaching to document achievement (Black & Wiliam, 1998; Black & Wiliam, 2003; Hongli, 2016, Stiggins, 2017) and can consist of multiple formats, such as authentic assessments, projects, objective and subjective assessments, major exams, and performance quizzes. Willingham et al. (2002) suggest that rather than using classroom assessment and grading practices to represent student mastery of learning objectives, standardized assessments, such as those produced by the standards-based reform, are being used to keep the grading scales honest and to provide an accurate representation of student learning. “One likely source of the apparent contradiction is the tendency to assume that a grade average and a test score are, in some sense, mutual surrogates; that is, measuring much the same thing, even in the face of obvious differences” (Willingham et al., 2002, p. 2).

Non-Achievement Factors

Often included in the assignment of a final grade, without a standardized weight, are non-achievement factors which are referred to as "academic enablers" (McMillian, 2001) and they make up more than 15% of a final grade (Ornstein, 1994). McMillian (2001) found that significant weight is given to effort, but also participation and improvement were components of the grading system used. Teachers also measure the degree of responsibility students have by including zeros to document failure for not turning in homework. Extra credit, which is used more prevalently in lower level classes (McMillian, 2001), contributes to the further masking of student deficiencies when using the final grade as an indication of achievement. Willingham et al. (2002) also discovered that teachers included grades for attendance and behavior to document the overall performance in their class. Ultimately, the inclusion of these non-achievement factors provides a false sense of achievement, contradictory to the historical and research-based purpose of grading (Allen, 2005; Brookhart, 1994; Brookhart, 2015; Guskey, 2006; Randall & Engelhard, 2010; Wormeli, 2006).

Current Practices in Research

Current research on assessment and grading practices at the secondary level reveals three broad categories used in the grading of secondary students: summative, formative and non-achievement factors. Academic achievement, consisting of both summative and formative assessments, represents the largest component of all secondary grades (Bursuck et al., 1996; Cross & Frary, 1999; Duncan & Noonan, 2007; Frey & Schmitt, 2010; Kunnath, 2017; McMillian, 2001; Ornstein, 1994; Randall & Engelhard, 2009; Willingham et al., 2002). The non-achievement category varies widely in their

components, but four common factors found within general grading practices became evident: (a) achievement, (b) ability, (c) behavior, and (d) effort and assessment types used by teachers varied among a multitude of other elements (see Table 10). These factors resemble the findings of Brookhart (1994) suggesting that grading practices have remained unchanged over the past quarter-century.

Overall, nearly 90% of all assessments given are used in the calculation of final grades (Frey & Schmitt, 2010), which indicates the little use of grades in a formative manner. In comparing the use of traditional to alternate assessment methods, neither were used more frequently, and both were used approximately half of the time (Bol et al., 1998) while Frey and Schmitt (2010) and McMillian (2001) found performance assessments were used about one-third of the time. In general, teachers were found to develop their own assessments about half of the time or less (Frey & Schmitt, 2010; McMillian, 2001), and these exams consist mainly of multiple choice and short answer/fill-in-the-blank type (McMillian, 2001). Dunan and Noonan (2007) suggest that these assessments are developed to mirror the end-of-course tests used in the state accountability program and may account for the reason teachers use tests created by someone other than themselves.

The review of the literature reveals that teachers' use of formative assessments is generally contradictive to its meaning and purpose within the classroom by being included in the calculation of final grades. Often, quizzes are given during instruction and the grading of homework, both of which are assumed to monitor student learning and to inform instruction, are included in the calculations of final grades. Formative

assessments, as described above, and their use was found to be similar across all subjects and does not vary by teacher gender. Frey and Schmitt (2010) indicate that assessments were given during instruction 25% of the time, but were only used formatively, absent of impacting the final grade, half of the time contrary to its definition. Wormeli (2006) states and most teachers would agree, that the purpose of homework is for the students to *practice* what they have learned within the classroom and therefore be used in a formative manner, that is to inform instruction. Instead, it was identified as a significant component of grades for 25% of teachers and a moderate component in 46% (Cross & Frary, 1999). More specifically, it was graded for correctness by slightly more than half of the teachers while 16% used classmates to grade, 43% only checked for completion, and 5% did not assign a grade at all (Cross & Frary, 1999). McMillian (2001) found that 51% of the teachers responding use homework in the calculation of their final grades, but only 28% of them graded it. The lack of assessing homework makes the formative use of homework unlikely, especially when only half or less review it, and implies that the use of homework is simply a measure of effort, which is a non-achievement factor.

Non-achievement factors, part of a teacher's *professional judgment,* include student behaviors, effort, attitude, organization, progress, participation, assigning zeros for not turning in homework, participation, ability, and growth consist of approximately one-fifth of a final grade (Bursuck et al., 1996; Ornstein, 1994). Duncan and Noonan (2007) identified effort as the most used factor, McMillian (2001) found that 78% of teachers used effort to some extent, and Cross and Frary (1999) found that effort positively impacted lower-ability students. Cross and Frary (1999) also found that 53% of teachers said that participation had a moderate or strong influence on the grades they

assigned, while 39% included behavior. Regarding ability, more than 50% of teachers consider it when assigning grades while frequently, 72% of teachers raised grades of low-ability students, and 20% used growth to a "substantial extent" in grading (Cross & Frary, 1999).

Of all the grading factors identified in the literature, the variation of teacher grading practices within the area of non-achievement factors continues to convolute the meanings of grades (see Table 10). Its variance is so substantial and inconsistent that Willingham et al., (2002) was unable to correct for its lack of reliability even when considering all other factors. Although the majority of these studies generally focus on assessment and grading practices among all teachers, a few suggest that the variation in the factors of grading in the core content areas could have some similarities assisting in the understanding of grades in each area.

Table 10

Assessment and Grading Practices of High School Teachers

Study	Major Findings
Brookhart (1994)	Variety of data used but consistent; Mostly achievement but also includes non-achievements; Ability and effort affects grades; Teacher variation; Grades communicate expectations; Discrepancies in grades; As grade level increases so does written test, ability effects grades, use of multiple assessment types
Pilcher (1994)	Math grades are a function of achievement, effort, and attitude; English grades are a function of ability, effort, and attitude (behavior); Parents believe grades represent achievement; Math teachers use of effort increased as ability level decreased
Ornstein, (1994)	15% of grades are based on professional judgment (suggestings -ability, behavior, effort); High schools mainly use test results (Achievement)
Bursuck, et al. (1996)	No difference in years of experience, education level, and training; Tests (Achievement) and homework were major components of grades – 67.5%; Papers/reports = 12%; About 20% were non-achievement factors such as effort/attitude, organization, progress, ability, attendance, and participation.
Bol, et al. (1998)	Observational and assessment methods and performance tasks used most frequently; No difference in traditional or alternate assessment use; Alternate assessment methods used more with older teachers; Math used traditional methods less frequently than other subjects
Cross & Frary (1999)	More than 50% of teacher consider ability; 72% raised grades of low-ability students; Math teachers were less likely to consider ability 54%; Discrepancy between the grading and use of homework; behavior, effort, and growth were used between 25% and 40% of teachers; 53% used participation
Mertler (2000)	Females used alternative assessment techniques more frequently; N.S. between years of experience and traditional assessments but did with alternative assessments; N.S. performance assessments; Assessment practices vary by level;
McMillian (2001)	Major exams used most; Discrepancy between subjects regarding non-achievement factors and assessments; Assessments varied by grade level and class ability
Willingham et al., (2002)	Grading practices contract with teacher and varies by course name; Significant grade variation; Grades used for motivation
Duncan & Noonan (2007)	Most used academic enabler is effort; Discrepancies between subjects and assessment type used; Summative assessment most used
Randall & Engelhard (2009).	Teachers assign grades on achievement mainly, but also use ability, behavior, and effort; Better behavior leads to better grades

Table 10 Continued

Study	Major Findings
Frey & Schmitt (2010)	12% of assessments are formative; Large majority of teachers replicate the traditional multiple-choice tests; no major differences between years of experience, and grade level. Females used performance-based assessment 50% more than males at all grade levels. Teacher created tests themselves 55% of the time.
Randall & Engelhard (2010)	Four major factors used in grading – achievement, ability, effort, behavior; Student characteristics relied on most in borderline cases; Behavior impacted the grades the most, even with high effort and ability
Brookhart (2015)	Relationship between grades and tests are 0.4-0.7; Much variation in grades; Student, teacher, and school or classroom context characteristics affected grades; personality, effort, behavior, classroom citizenship or learning skills were used in class grades
Kunnath (2017)	Four factors -achievement, learning targets mastered, ability, effort; Inclusion of zeros for incomplete assignments or assessments nearly a 5th; Use of effort varied significantly; achievement largest factor; Effort is used to prevent failure of the low-ability student; High achieving – high ability students' grades less subjective

Content areas. Overall, Pilcher (1994) found that English grades are a function of ability, effort, and attitude. Teachers in English were more likely to create their own assessments that measure higher-order thinking skills and were more likely to have constructed responses (McMillian, 2001). Frey and Schmitt (2010) found that English teachers were more likely to give performance-based assessments than any other subject. In the calculation of the final grade, academic achievement was emphasized more in English than any other content area (McMillian, 2001).

The overall findings surrounding math teachers varied and were inconsistent. Of all the content areas, math teachers used non-achievement factors, like extra credit and assigned zeros for homework that was not completed, the least (Duncan & Noonan, 2007; McMillian, 2001), but math teachers were found to use effort more in final grades than other subjects. Pilcher (1994) found that grades in this subject area are a function of

achievement, effort, and attitude, with the use of effort having an inverse relationship to ability when determining final grades.

Regarding assessments, Duncan and Noonan (2007) found that math assessments were more likely to be major exams that could be scored objectively, and they were also the least likely to use constructed responses. On the contrary, Bol et al. (1998) found that they used traditional methods less frequently than other subjects while the authors suggested the teachers are more interested in problem-solving aspect rather than selecting an answer choice. Duncan and Noonan (2007) also found they had the highest degree of aversion from using benchmark tests in final grades, and Cross and Frary (1999) added that math teachers were less likely to consider ability than other content areas, only 54% did.

Like math, science also preferred objective assessments, but did more so (McMillian, 2001), emphasizing the use of major exams, and equally avoiding using benchmarks and the use of academic enablers in the calculation of final grades (Duncan & Noonan, 2007). Science teachers also developed assessments that emphasized recall more often than math and English (Duncan & Noonan, 2007). No additional significant findings were reported.

Social studies teachers used recall assessments similarly to science teachers (Duncan & Noonan, 2007; McMillian, 2001) and academic achievement was emphasized less than English and math subjects (Duncan & Noonan, 2007). Exam structure also differed between 10-12th grade and 9th-grade students in that as a student's grade level increases the more likely constructed responses were used (McMillian, 2001).

Teacher and student characteristics. In general, male teachers use alternative assessment techniques half as much as their female counterparts (Mertler, 2000), and of alternate assessments, females are 50% more likely to use performance-based assessment at all grade level and create their own tests (Frey & Schmitt, 2010). The use of formative assessments was not found to have a statistically significant difference between male and female teachers (Frey & Schmitt, 2010).

Frey and Schmitt (2010) found no significant difference between years of experience regarding grading practices, and there was no correlation between experience and the development of their own assessment versus them being created by another individual. Older teachers tend to avoid short answer test formats and are more likely to use performance-based exams rather than multi-choice, matching, true-false, and essay (Frey and Schmitt, 2010). The use of alternate assessments was inconclusive among varying years of experience as Bol et al. (1998) found alternate assessment methods were used more with older teachers, 20 + years, while Mertler (2000) found that younger teachers used alternative assessments more often.

McMillian (2001) found that externally developed tests were used more frequently in the lower grades, constructed responses the most in 12th grade and emphasized higher-order thinking skills were used more in the 12th grade than in the 9th. They continue by identifying that homework, and the use of zeros was more prevalent in ninth and 10th grade than 12th grade. Frey and Schmitt (2010) found no significant differences between grade-level and grading practices.

The use of assessment as a major component of grades varied significantly, but students with higher ability had assessments comprise a larger component of their final grade (Kunnath, 2017). Higher ability students were also tested using more constructed responses than lower ability students (McMillian, 2001). Effort as a grading factor also varied between students of ability level. McMillian (2001) found that the lack of perceived effort by teachers of higher ability students did not negatively impact their grade, but higher effort raised grades in 25% of the teachers surveyed. Cross and Frary (1999) found that in 72% of the teachers who taught low-ability students, they raised their grades. Ultimately, as the students' ability level increase, their grades are less subjective and more likely to reflect their achievement in each subject area (Kunnath, 2017; Cross & Frary, 1999; McMillian, 2001).

Conclusions

In the years following IASA, the graduation rate has continued to climb, nearly 10% nationally (National Center for Education Statistics, 2015) and 18% in North Carolina (North Carolina Department of Public Instruction, 2016). Simultaneously, the reports revealed, that national GPAs had an upward trend between 1990 and 2009, from a 2.68 to a 3.0, and the rapid graduation rate increase after 2009, 6% nationally and 11% in North Carolina, suggest that GPAs should reflect a similar increase. These collective statistics show an overall positive impact on education resulting from the standards-based educational reform, but when compared to success in post-secondary studies, "60% of first-year students find themselves academically unprepared" (National Center for Public Policy and Higher Education and the Southern Regional Education Board, 2010). The

passage of IASA in 1994 changed the curriculum across the United States and resulted in assessment methods reflecting the standardized accountability tests (Hamilton et al., 2008), but the overall impact on grading practices at the secondary level is not yet evident.

Brookhart's (2015) meta-analysis of the past 100 years of grading indicates there were a lot of descriptive studies on assessment and grading practices occurring in the 1980s and 1990s, but the studies have decreased in frequency over time following the standards-based reform. The literature review, post-IASA, revealed that achievement and non-achievement factors were mixed to create conglomerate grades which makes it difficult to describe their components. Only five studies examining specifically the four content areas (see Bol et al., 1998; Cross and Frary, 1998; Duncan & Noonan, 2007; McMillian, 2001; Pilcher, 1994) were discovered and they have some areas of commonality suggesting the ability to infer the meaning of grades among subject areas, while student and teacher characteristics have been found to vary among other studies. The following chapter will describe the methodology, participants, and procedures of the current study.

Chapter 3: Methodology

This non-experimental descriptive study using survey research methods examined the current assessment and grading practices of high school English, math, science, and social studies teachers. A survey was administered to current high school, grade 9-12, teachers of English, math, science, and social studies to address the following research questions:

1. What factors do secondary teachers use in determining grades for students in the four core content areas, English, math, science, and social studies?
2. What types of assessments do secondary teachers use in grading students in the four core content areas, English, math, science, and social studies?
3. How do the types of assessment and factors used in grading students vary among English, math, science, and social studies teachers?

This chapter will review the targeted sample, a description of the instrumentation, research design, the procedures, and data analysis methods that were used in the study.

Description of Participants

Eight school districts in the western and coastal regions of North Carolina participated in the current study, containing 34 high schools from rural, suburban, and urban areas. The collective high schools have approximately 212 English, 236 math, 199 science, 217 social studies teachers, and six teaching all subjects totaling 870 possible subject-teacher combinations, but there are only 867 possible participants due to three participants having more than one subject listed as a teaching responsibility. The eight

districts were selected due to their immediate access to the researcher and were selected to obtain a representative sample of teachers from the state of North Carolina. The districts' demographics vary in degrees of diversity, poverty, and size. The participating districts' demographics are described in Table 11 below.

Table 11

Participating District Demographics

PK-12	9-12	Grad	FRL	White	AA	His	Asian	Multi	Haw/Pac	NA
11,500	3,706	87.9	48.6	85.0	5.0	7.0	0.0	2.0	0.0	0.0
3,090	943	93.5	63.0	50.0	13.0	25.0	6.0	6.0	0.0	0.0
4,480	1,250	85.0	47.2	45.2	21.0	22.9	4.0	6.7	0.0	0.0
17,370	5,380	91.4	48.4	75.0	6.0	12.0	7.0	4.0	0.0	0.0
5,084	1,519	86.1	54.2	80.9	4.1	9.3	1.9	3.9	0.1	0.1
20,858	9,222	88.1	43.4	83.1	12.2	7.3	2.4	1.8	0.1	0.5
26,214	8,116	82.4	51.0	81.3	14.6	5.4	1.4	2.0	0.0	0.6
5,870	1,706	90.0	30.0	73.0	14.0	8.0	1.0	3.0	0.2	0.1

Note. Grad = Graduation Rate; FRL = Free and Reduced Lunch; AA = African American; His = Hispanic; Haw/Pac = Hawaiian/Pacific Islander; NA= Native American

Research Design

This study employed a non-experimental descriptive design using web-based survey research methods to examine current assessment and grading practices of high school English, math, science, and social studies teachers. The study used a quantitative data analysis because it can be used to "develop a precise representation of a sample's behavior or personal characteristics" (p. 214) and can be used as a starting point for educational reform (Gall, Gall, & Borg, 2010) around assessment and grading practices of students.

Dillman et al's. (2014) Tailored Designed Method (TDM), which is based on social exchange theory that explains how people interact in daily life by considering the

costs and benefits, ultimately leads to the building of trust or a lack thereof. The TDM "is the development of survey procedures that create respondent trust and perceptions of increased rewards and reduced costs for being a respondent, that take into account features of the survey situation, and that have as their goal the overall reduction of survey error" (Dillman, 2000, p. 4). In this study, respondents remained anonymous, and the informed consent assured all participants that their participation would be kept secure and confidential. Furthermore, the questions were non-threatening and did not require participants to reveal embarrassing information. There was no correct or incorrect response to the items on the survey, and respondents were asked to give their opinion or perception.

The TDM emphasizes four carefully timed contacts and the use of respondent-friendly surveys to maximize response rate (Dillman et al., 2014), but the number of contacts was reduced to three to avoid overburdening the teachers and encourage districts to participate. For this study, the survey was designed to ensure that participants would not take more than 10 minutes to complete the survey; two volunteers, not included as part of the sample, completed the survey in 6-8 minutes. The survey was distributed one month prior to schools opening the testing window at the end of the school year when teachers are finalizing grades for the academic year.

Description of Instrument

The instrument used in this study, with permission, was developed by McMillian (2001) to explore variations in assessment and grading practices teachers use across a variety of classes. The original instrument was developed through a review of research on

assessment and grading practices and the adoption of items from previous survey instruments. The original survey contained 47 closed-ended items identified, including both achievement and non-achievement factors, e.g., behavior, ability, and effort, the types of assessments used, and the cognitive level of assessment (e.g., recall knowledge, understanding, application, and reasoning).

The original 47 were checked for content-validity by asking 15 teachers to review the items for clarity and to ensure that all aspects of items considered in assessment and grading were included. The revised items were subjected to a second pilot test with 23 teachers from a school district outside of the sample to gather additional feedback on clarity, correlations among items, item response distributions, and reliability. Test-retest reliability, completed with the second pilot group four weeks after the initial test, showed on average an exact match on 46% of the teachers; 89% of the matches were within 1 point of the 6-point scale. Coefficients of agreement were also calculated, and items that showed an exact match of less than 60% were deleted or combined with other items. Item statistics were used to reduce the number of items to 34 in three categories (19 items assessed grading factors, 11 assessed different types of assessment, and four items assessed the cognitive levels) in the revised survey used in McMillian (2001).

McMillian's (2001) instrument was adapted by the researcher through an item restriction process to align with the research questions and purpose of the proposed study. The items related to the cognitive levels of Bloom's taxonomy a student was assessed were removed, and additional items were developed from the literature review. The questionnaire was checked for clarity through a talk-aloud protocol with three high

school teachers outside of the targeted sample, resulting in revisions of item wording, the inclusion of real-classroom examples to clarify, and the elimination of duplicate items. Twenty-eight Likert-type items remained from the original instrument to be used in the present study, three were adopted from Frey and Schmitt (2010) to address the degree of formative and summative assessment variance and to include assessments created by others which were not included on the original survey, and three additional questions were adapted from the literature review to examine grade inflation.

The resulting questionnaire from the process above, see Appendix A, has 42 items consisting of 6 demographic questions, absent of identifying information, and 36 questions targeting assessment and grading practices have a common stem for all the Likert-type items; *To what extent will the final grades of students in your class, as described above, be based on.* The questionnaire responses are a six-point Likert rating scale which ranges from *Completely* to *Not at all* and was constructed to allow teachers to indicate usage without the constraints of a forced choice response. Research question and survey item alignment are indicated in Table 12 below.

Table 12

Alignment of Research Questions and Survey Items

Research Question	Survey Question Item Number
1. What factors do secondary teachers use in determining grades in the four core content areas, English, math, science, and social studies?	1-19, 32 -36
2. What types of assessments do secondary teachers use in the four core content areas, English, math, science, and social studies?	20-31
3. How do the types of assessment and factors used in grading vary among English, math, science, and social studies teachers? [dependent variables]	1-36

The selection of a specific semester as a target for the proposed study, spring of 2018, is necessary because this makes "recall simple and related to recent events [which] helps to produce high-quality survey data" (p. 37) and increase survey response rate (Dillman, 2000). Likert-type questions were used because they provide a linear representation of the sample's degree of use and practices (Rattray & Jones, 2007) surrounding assessment and grading. These items also eliminate the participants' necessity to calculate a percentage or to estimate the degree of use from the procedures implemented in their classrooms (Rattray & Jones, 2007).

The questionnaire was divided into five sections with like questions grouped to aid in the understanding of content and the framing of responses thereby reducing response error (Dillman & Smyth, 2007; Dillman et al, 2014). The demographic sections are divided to keep the most interesting and necessary items at the top of the survey, and

the least interesting, participant characteristics, at the bottom. The item responses are arranged in a single vertical row to encourage top-to-bottom reading and avoid skipping response choices as have been found in multi-column or horizontal response layouts (Dillman, & Smyth, 2007).

The survey was deployed using Google Forms, a web-based application, due to its ease of use and the ability to easily export the data and merge it with the PASW Student Version 18 statistical software package (SPSS Inc, 2009). An email was distributed with the Google Forms link to the survey embedded into an e-mail which allowed an anonymous response and provided a summary of cumulative responses to participants upon submission. The use of Google Forms allowed the questionnaire to be combined into a single scrolling design rather than a multi-screen approach as suggested by Dillman and Smith (2007) because it " (a) requires less computer time and computer resources to contact with the web server because it only requires one single submission of the final responses and (b) provides richer context for respondents to respond because all questions are on one page" (as cited in Weimiao & Yan, 2009, p. 134).

Procedures

This study implemented the TDM, slightly modified to meet the needs of the researcher, which consisted of three carefully planned contacts instead of four, outlined below, based on social exchange theory, and aims to reduce overall survey error (Dillman et al., 2014). The survey was delivered electronically through the participants' school provided email addresses because participants are more likely to respond honestly when

they can preview questions as opposed to over the phone or in-person interviews where they may not have the same opportunity (Dillman et al., 2014).

Dillman et al's. (2014) TDM emphasizes the use of the three elements of the social exchange theory critical to achieving the maximum return rate in survey research methods which include rewards, what is expected to be gained by the participant for participating, costs, what participant must sacrifice to obtain the rewards, and the building of trust that the rewards of participating will eventually outweigh the costs of participating in the study. The researcher offered a drawing for one of four possible VISA $25 gift cards for the participants to use for their classroom as an incentive to share their grading practices as well as to emphasize the importance of understanding how grades are assigned at the secondary level through each of the contacts made with the participants. In order to be included in the drawing, the respondents were required to submit their email address to the researcher after they have responded to the survey. The degree of participant costs were reduced by ensuring total anonymity in their responses to the survey by not collecting any identifiable information; unless they elected to participate in the drawing, then their email address was removed prior to importing the data into the data analysis software. In addition, the information provided by the potential participants would be very low risk and feel comfortable revealing this information. As a whole, these processes reduced the overall risk of participation and building the trust that what they provide to the researcher is valuable in improving the field of education.

The emphasis in the reduction of overall survey error, consisting of four components, coverage, sampling, measurement, and non-response error, is integral to Dillman et al's. (2014) TDM. Due to the researcher's choice to use a convenience sampling method targeting eight school districts in the Piedmont and coastal region of North Carolina, the ability to reduce coverage and sampling error are limited when compared to a true random sampling technique targeting the entire state. Attempts to mitigate the limitations of the sampling technique include targeting districts that are diverse and located in rural, suburban, and urban areas. Measurement error was reduced through the use of direct survey methods with Likert-type items developed through the use of teacher content-validity checks for clarity, test-retest reliability from two pilot tests, and the use of item statistics to develop the instrument.

The ability to reduce non-response error with email survey methods is a concern because Weimiao and Yan (2009) found that web-based surveys have a response rate of approximately 11% less than other methods and Kittleson (1997) had similar findings with the response rates falling between 25-30% for web-based surveys. Although the response rate for email surveys is somewhat lower than other survey modes, through the use of multiple reminders, as suggested by the TDM, Kittleson (1997) found it could be doubled, Sánchez-Fernández, Muñox-Leiva, and Monotor-Rios (2011) found a 25% increase, and Dillman (2000) found an improved response rate of 20-40%. The researcher in the current study used only three contacts with potential participants, as opposed to the four suggested in the TDM, raising concerns of a poor response rate but Sánchez-Fernández et al. (2011) indicated no significant changes in response rate after the second

reminder. Therefore, the impact on response rate due to employing one less contact should be minimal.

Each of these three contacts addressed the three elements of social theory incrementally, appearing different than the previous with the final contact containing language expressing the importance of the considerable efforts and resources used to collect their information. The initial contact, as well as all subsequent contacts, was made via the participants' school district assigned e-mail addresses collected from their respective district's school webpage. The letter was embedded in an email, and a link to the survey was included, opening a new window to prevent technological issues with the various email hosting services. The introductory letter, see Appendix B, provided a positive initial contact, informed the participants of the purpose of the study, emphasized the benefits of participating, included why the participants were needed to complete the survey, its usefulness, a statement of confidentiality, and a how they could contact the researcher should they have questions or comments regarding the study. A note indicating the endorsement of the district was included, as suggested by Weimiao and Yan (2009), to convey the idea that the survey is important. Each participant was notified that no identifying information was going to be collected and, as a result, may be contacted in the future even if they have already submitted the survey. The researcher attempted to mitigate repeated future contacts by removing participants from the mail merge that submitted their school district assigned email address, or an email address which was easily identifiable to a specific participant, to participate in the VISA gift card drawing.

The first reminder to complete the survey, and second subsequent contact with participants, see Appendix C, occurred five days after the sending of the survey based upon the suggestions made by Dillman et al. (2014), because typically half of all returned surveys occur within the first three days. The purpose of this contact was to remind targeted participants about the survey if they have not completed it, to thank them for their participation if it has been returned, how to contact the researcher should they have any questions, and to provide an additional link to the survey.

The second reminder mirrored the delivery of the original survey but with stronger language emphasizing the importance of responding. The third and final contact, delivered 12 days after the first reminder, included a new cover letter, via email indicating that it has been nearly three weeks since they first received the questionnaire. The letter, see Appendix D, provided instructions to complete the survey with a link to the questionnaire embedded at the end of the e-mail, reminded them about the drawing, thanked those who have responded, and indicated the study will soon come to an end.

In summary, the TDM implemented in the current study occurred over approximately 46 days beginning on May 1, 2018, and the survey remained open through June 15, 2018, for all participating districts. These dates were selected to avoid summer vacation while simultaneously occurring nearest the end of the school year to allow participants to have the most recent knowledge of their assessment and grading practices they implemented during the second semester of the school year.

Data Analysis

The responses to the questionnaire were extracted from Google Sheets, the data collection application linked to the survey, and imported into the PASW Student Version 18 (SPSS Inc, 2009) for analysis, see Table 13 for a summary of research question alignment and the statistical analysis method. For Research Question one, frequency distributions, means, and standard deviations were reported for items 1 to 19 and 32 to 36 on the survey. These items were rank-ordered by means from highest to lowest. The same statistical procedure was used to examine Research Question two using survey questions 20 to 31. For Research Question three, an ANOVA was completed for each of the 36 items on the instrument with each item's mean used as the dependent variables and the content area (i.e., English, math, science, and social studies) served as the independent variable. Using an alpha level of .05, statistically significant differences in means among the content areas were identified, and post hoc analyses were completed to identify the specific content area differences. Finally, Cohen's *d* was calculated for items with significant post hoc analyses to determine practical effect sizes.

Table 13

Research Questions and Statistical Analysis Method

Research Question	Statistics	Data Analysis
1. What factors do secondary teachers use in determining grades in the four core content areas, English, math, science, and social studies?	Descriptive Statistics	Means, Medians, Standard Deviations, Frequencies, and Percentages
2. What types of assessments do secondary teachers use in the four core content areas, English, math, science, and social studies?	Descriptive Statistics	Means, Medians, Standard Deviations, Frequencies, and Percentages
3. How do the types of assessment and factors used in grading vary among English, math, science, and social studies teachers?	Descriptive and Inferential Statistics	Means, Medians, Standard Deviations, ANOVA

Summary

The researcher used a non-experimental survey research design to examine current assessment and grading practices of high school English, math, science, and social studies teachers. The present study updated previous studies surrounding assessment and grading practices at the secondary level to address the gap between increasing graduation rates and higher GPAs but a lower preparedness of incoming post-secondary students in institutions of higher education. The study provides a description of secondary teachers' perceptions of their assessment and grading practices of their students within the four core content areas, English, math, science, and social studies will identify meaningful relationships between the core content areas. The targeted sample of this study included only high school teachers who are currently teaching at least one, English, math, science, or social studies class during the spring semester of 2018.

The data were analyzed through descriptive and inferential processes using SPSS software, and the final two chapters will reveal the findings of the study and discuss their significance as it applies to the education of high school students.

Chapter 4: Data Analysis

Introduction

The current study examined grading practices at the secondary level and addresses the gap between increasing graduation rates and the preparedness of post-secondary students. The study provides a description of assessment and grading practices within the four core content areas, English, math, science, and social studies, of current high school teachers and identifies relationships between the core content areas. This chapter is organized by first describing the participants followed by a report of the findings of the three research questions and concluding with a focus on the significant findings.

The instrument used in this study required the participants to rate the degree in which they used a factor in their assessment and grading practices or agreed with a specific statement using a six-point Likert-type scale (*1-Not at all, 2- Very Little, 3-Some, 4-Quite a bit, 5-Extensively, or 6-Completely).* The results are first reported as descriptive statistics for the items on the instrument based upon its three distinct portions: grading factors which consisted of achievement and non-achievement factors, factors impacting the final grade, and the assessments used by participants. Following the descriptive reporting, results of one-way ANOVAs are reported to determine statistically significant mean differences among each of the core content areas. The responses to the survey for achievement factors, both in aggregate and by subject area, reporting means, standard deviation, and the response frequencies by percentages, can be found in Appendix E. The response frequency for all grading factors influencing the final grade in

English is found in Appendix F, math in Appendix G, science in Appendix H, and social studies in Appendix I. The response frequency by percentage for items related to the types of assessments used are found in Appendix J, with subject area responses specific to English in Appendix K, math in Appendix L, science in Appendix M, and social studies in Appendix N.

Participants

The survey was emailed to 867 participants with 21 emails being returned because they no longer existed, and two teachers indicated they did not meet the requirements of the study, which yielded 844 valid surveys distributed. Of the 844 sent, 248 surveys were returned giving an overall return rate of 29% for this study. The 248 teachers participating in this study consisted of 158 females (63.7%), 90 males (36.3%), and the racial demographics included 93.1% White, 3.2% Black or African American, 1.6% Hispanic or Latino, 1.6% Multi-Racial, and 0.4% Asian or Pacific Islander.

The number of participants declined as the years of experience increased with teachers having five or fewer years representing 21%, six to ten years 19.8%, eleven to fifteen years 21.8%, sixteen to twenty years 16.5%, twenty-one to twenty-five years 9.7%, and those with more than twenty-five years representing 11.3% of the sample. Most of the participating teachers taught Honors and Regular/College prep courses, 37.9%, and 38.3% respectively, while 17.7% taught Advanced Placement and 6% Remedial level courses. The participants were asked to select a specific course they were currently teaching to report assessment and grading practices; the responses indicate a

nearly equal representation across all subject areas with English courses representing 27%, math 25.8%, science 26.2%, and social studies 21% (see Table 14).

Table 14

Number and Percentage of Participants by Core Content Area

Subject	Frequency	Percent
English	67	27.0
Math	64	25.8
Science	65	26.2
Social Studies	52	21.0
Total	248	100.0

Grading Factors Used by Secondary Teachers

The sample sizes, means, and standard deviations for the four achievement factor items for all 248 responses are reported in Table 15. Recall that the items are measured on a one to six-point Likert scale. Three of the items had means above 4.0 (*Quite a bit*). The teachers reported they compare student performance to a percentage correct (M = 4.47), grade according to the objectives students master (M = 4.41), and use academic performance as opposed to other non-academic factors (M = 4.21). The use of graded homework had a mean, on average, below 4.0 (*Some*; M = 3.21) and was the lowest ranked achievement factor.

Table 15

Sample Size, Means, Standard Deviations for Achievement Factors in Aggregate Across All Participants

Item	*N*	*Mean*	*SD*
Percent correct (Item 1)	248	4.47	1.16
Objective mastery (Item 2)	248	4.41	1.05
Academic performance (Item 3)	248	4.21	1.00
Homework quality (Item 7)	248	3.21	1.38

Note. Item numbers refer to the instrument, see Appendix A.

The sample sizes, means, and standard deviations for the four achievement factors by core content areas are reported in Table 16. Generally, the trend across the content areas is similar to the collective results with three of the means being above 4.0 (*Quite a bit*). The teachers reported, with an average mean above 4.0 (*Quite a bit)* using student performance compared to a percentage correct (English, *M* = 4.63; Math, *M* = 4.34; Science, *M* = 4.55; Social Studies, *M* = 4.33*)*, grading according to the objectives students mastered (English, *M* = 4.37; Math, *M* = 4.17; Science, *M* = 4.72; Social Studies, *M = 4.37)*, and academic performance as opposed to other non-achievement factors (English, *M* = 4.06; Math, *M* = 4.30; Science, *M* = 4.32; Social Studies, *M = 4.15)*, while also reporting the use of homework quality, on average, below 4.0 (*Some;* English, *M* = 3.60; Science, *M* = 3.26; Social Studies, *M = 3.58).* The most noticeable difference in means, among all achievement factors, is found within the grading of homework for correctness and math teachers report using it less than 3.0 (*Very Little*; *M* = 2.47), ranging nearly a full rating point from English teachers who had the largest mean (*Some*; *M* = 3.60).

Table 16

Sample Size, Means, Standard Deviations for Achievement Factors by Core Content Area

	English (*N* = 67)		Math (*N* = 64)		Science (*N* = 65)		Social Studies (*N* = 52)	
Item	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Percent correct	4.63	1.13	4.34	1.06	4.55	1.23	4.33	1.25
Objective mastery	4.37	1.08	4.17	1.03	4.72	.91	4.37	1.12
Academic performance	4.06	1.03	4.30	.81	4.32	1.08	4.15	1.07
Homework quality	3.60	1.31	2.47	1.17	3.26	1.40	3.58	1.35

The sample sizes, means, and standard deviations for the non-achievement factors are reported in Table 17. The item with the highest mean was the use of zeros for incomplete or missing assignments (*Quite a bit; M = 4.42)* in the determination of their final grades. Five of the 15 non-achievement factors on the survey had a mean greater than 3.0 (*Some*). Teachers reported they used student effort (*M* = 3.44), completion of student homework, not graded, (*M* = 3.29) adjusted borderline grades based upon effort, improvement, behavior, or other "non-test" indicators (*M* = 3.22), the degree students pay attention or participates in class (*M* = 3.15), and student improvement since the beginning of the year (*M* = 3.10). The teachers also reported, on average below 3.0 (*Very Little*), using ability (*M* =2.60), student work habits (*M* =2.40), and extra credit for academic reasons (*M* = 2.05). In addition, teachers reported they did not use, on average below 2.0 (*Not at all*), performance compared to other students in their class (*M* = 1.89), disruptive behavior (*M* =1.70), grading based upon a policy restricting the number of letter grades that can be assigned (*M* =1.54), past performance of other students (*M* =1.52), extra

credit for non-achievement reasons (*M* =1.22), and the grades other teachers assigned (*M* =1.19) in determination of final grades.

Table 17

Sample Size, Means, Standard Deviations for Non-Achievement Factors in Aggregate Across All Participants

Item	*N*	*Mean*	*SD*
Zeros (Item 4)	248	4.42	1.49
Effort (Item 6)	248	3.44	1.26
Homework completion (Item 9)	248	3.29	1.33
Borderline grades (Item 10)	248	3.22	1.38
Attention or participation (Item 8)	248	3.15	1.33
Improvement (Item 11)	248	3.10	1.33
Ability (Item 5)	248	2.60	1.24
Work habit (Item 12)	248	2.40	1.22
Extra credit academic (Item 13)	248	2.05	1.13
Performance of current students (Item 14)	248	1.89	1.10
Behavior (Item 15)	248	1.70	1.03
Letter policy (Item 17)	248	1.54	1.18
Performance of past students (Item 18)	248	1.52	.98
Extra credit non-academic (Item 16)	248	1.22	.54
Grades other teachers assign (Item 19)	248	1.19	.59

Note. Item numbers refer to the instrument, see Appendix A.

The means of non-achievement factors used within the core content areas are reported in Table 18. Teachers in the core content areas reported the use of non-achievement factors similar to the aggregate group, with all reporting the use of zeros the highest, on average, above 4.0 (*Quite a bit)*. The only reported difference among English teachers, when compared to the aggregate results, is that they use extra credit less, and reported a mean average of less than 2.0 (*Not at all*; *M* = 1.88) which was also the least among all content areas. Math teachers had a sole difference as well, and reported using

attention and participation less than 3.0, on average (*Very Little; M* = 2.92). Science teachers were found to use two grading factors differently from the aggregate. The first, the use of student improvement, was reported below 3.0 (*Very Little*; *M* = 2.97), which is the least among all content areas, and the second, extra credit for academic reasons was reported less than 2.0 (*Not at all*; *M* = 1.97). Social studies teachers only varied in one area as well by comparing student performance to current students, reporting on average greater than 2.0 (*Very Little*; M = 2.08), which is more than most of the content areas and the group means.

Table 18

Sample Size, Means, Standard Deviations for Non-Achievement Factors by Core Content Area

	English (*N*=67)		Math (*N* = 64)		Science (*N* =65)		Social Studies (*N* = 52)	
Item	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Zeros	4.33	1.55	4.19	1.37	4.48	1.63	4.75	1.34
Effort	3.66	1.27	3.33	1.11	3.26	1.22	3.52	1.42
Homework completion	3.28	1.39	3.47	1.34	3.20	1.21	3.21	1.39
Borderline grades	3.45	1.36	3.11	1.51	3.00	1.27	3.35	1.33
Attention or participation	3.39	1.27	2.92	1.24	3.02	1.36	3.29	1.43
Improvement	3.30	1.22	3.03	1.25	2.97	1.35	3.08	1.54
Ability	2.76	1.29	2.52	1.04	2.48	1.25	2.67	1.38
Work habit	2.49	1.28	2.03	1.01	2.49	1.20	2.63	1.31
Extra credit academic	1.88	1.01	2.08	1.10	1.97	1.03	2.33	1.37
Performance of current students	1.81	1.08	1.81	1.08	1.91	1.10	2.08	1.17
Behavior	1.60	.89	1.69	.94	1.63	1.02	1.92	1.28
Letter policy	1.43	1.05	1.28	.88	1.60	1.30	1.94	1.41
Performance of past students	1.49	.98	1.34	.76	1.52	.94	1.79	1.23
Extra credit non-academic	1.22	.55	1.16	.41	1.20	.47	1.31	.73
Grades other teachers assign	1.24	.70	1.14	.39	1.12	.38	1.27	.82

Regarding the final grade, see Table 19, teachers agree with four out of the five factors on the survey reporting, on average, more than 3.0 (*Some*). The highest reported means, above 4.0 (*Quite a bit*), indicated teachers believe their grades represented achievement only ($M = 4.54$) while also believing ($M = 4.47$) that inflated grades contributed to the overall increase in the graduation rate in North Carolina. The participants also agreed, reporting on average more than 3.0 (*Some*), that a minimum grading policy limits the assignment of low grades for mid-semester grades ($M = 3.34$), and teachers recognize and care that administrators and other teachers judge their professional effectiveness upon the grades they assign ($M = 3.28$). Teachers also indicated they are not encouraged (*Very Little;* $M = 2.75$) to increase grades due to the offering of credit recovery to prevent student failure.

Table 19

Sample Size, Means, Standard Deviations for Final Grade Factors in Aggregate Across All Participants

Item	*N*	*Mean*	*SD*
Grades represent achievement only (Item 32)	248	4.54	.81
Grade inflation causing grad rate increase (Item 34)	248	4.47	1.25
Minimum grading policy limits assignment of low grades (Item 36)	248	3.34	1.75
Administrators and teachers judge grades assigned (Item 33)	248	3.28	1.24
Credit recovery encourages grade bumping (Item 35)	248	2.75	1.57

Note. Item numbers refer to the instrument, see Appendix A.

The responses among the core content area teachers, see Table 20, were relatively similar and only varied among social studies teachers. They reported, on average more than 3.0 (*Some*), that credit recovery encourages them to increase borderline failing grades ($M = 3.06$) as opposed to less than 3.0 *(Very little*).

Table 20

Sample Size, Means, Standard Deviations for Final Grade Factors by Core Content Areas

	English (*N* = 67)		Math (*N* = 64)		Science (*N* =65)		Social Studies (*N* = 52)	
Item	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Grades represent achievement only	4.54	.84	4.53	.78	4.52	.75	4.58	.89
Grade inflation causing grad rate increase	4.49	1.31	4.61	1.20	4.34	1.25	4.42	1.23
Minimum grading policy limits assignment of low grades	3.37	1.76	3.48	1.84	3.06	1.57	3.46	1.86
Administrators and teachers judge grades assigned	3.27	1.24	3.13	1.19	3.37	1.21	3.37	1.36
Credit recovery encourages grade bumping	2.75	1.65	2.78	1.55	2.46	1.42	3.06	1.65

Assessment Types Used by Secondary Teachers

The means for assessment types used by secondary teachers are reported in Table 21. The highest reported means, on average, were higher than 4.0 (*Quite a bit*) and indicated that the participants use summative assessments they create (*M* = 4.48) that contained objective items *(M* = 4.17) the majority of the time. The collective group also used, reporting more than 3.0 on average (*Some*), major exams (*M* =3.75), essays (*M* = 3.29), individual projects (*M* = 3.26), and formative assessments (*M* = 3.12). The remaining assessment types were reported, on average, less than 3.0 (*Very Little*) and they included the use of team projects (*M* = 2.91), tests made by others (teachers or found online) (*M* = 2.88), performance (*M* = 2.85) and authentic assessments (*M* = 2.63). The

lowest reported means, on average, was the use of assessments created by textbook publishers or designed by the district (*Very Little; M* = 2.23). Collectively, the results indicate most assessments used in secondary classrooms are created by the classroom teacher and are summative with objective tests items.

Table 21

Sample Size, Means, Standard Deviations for Assessment Types in Aggregate Across All Participants

Item	*N*	*Mean*	*SD*
Summative (Item 21)	248	4.48	.87
Designed by self (Item 22)	248	4.48	.96
Objective (Item 23)	248	4.17	1.08
Major exams (Item 27)	248	3.75	1.14
Essay (Item 24)	248	3.29	1.40
Project individual (Item 26)	248	3.26	1.18
Formative (Item 20)	248	3.12	1.24
Project team (Item 29)	248	2.91	1.05
Test made by others (Item 31)	248	2.88	1.23
Performance (Item 25)	248	2.85	1.23
Authentic (Item 28)	248	2.63	1.12
Publisher or district designed (Item 30)	248	2.23	1.11

Note. Item numbers refer to the instrument, see Appendix A.

The assessments used in each subject area and who creates them vary between one another but also share various factors with the aggregate, see Table 22. English teachers reported the use of assessments designed by themselves the most (*M* = 4.58), while also reporting a very high use of summative assessments (*M* = 4.46) that are objective (*M* = 4.01) while also reporting using essays (*M = 4.26)* on average above 4.0 (*Quite a bit*). They also used major exams, individual projects, formative assessments, and performance assessments, on average less than 3.0 (*Some*) and reported the use of

tests made by other teachers or found online less than the aggregate. English teachers also used authentic assessments and publisher, or district-designed assessments the least.

Math teachers also reported the use of summative assessments that they designed the most, on average more than 4.0 (*Quite a bit*) but reported they used objective assessments less (*Some*; *M* = 3.72), which is the least of all content areas. In addition, they used individual projects and formative assessments with a mean average below 3.0 (*Very Little*), while reporting the use of performance assessments, on average, below 2.0 (*Not at all*; *M* = 1.89), which is the least of all the core content areas and the lowest reported item by science teachers. Math teachers reported, on average less than 3.0 (*Very little*), using formative assessments (*M* = 2.80) and individual projects (*M* = 2.56), which is also less than the other content areas. Math teachers were also more likely to use assessments created by other teachers or found online (*Some*; *M* = 3.14).

Science teachers reported using objective, summative assessments they designed and major exams more than 4.0 (*Quite a bit)*, on average, which is the most of all the core content areas. Also, science teachers used essay, team projects, tests made by other teachers or found online, and performance assessments less than the entire group, reporting their use less than 3.0 (*Very Little*) on average. Social studies teachers reported means like the group, only reporting the use of assessments created by other teachers or found online less than 3.0 (*Very Little*; *M* = 2.67) which is also the lowest among all content areas.

English and social studies teachers were found to be similar in their responses with both reporting similar use of assessments types, except English use essays more

(*Quite a bit*; *M* = 4.13) than social studies on average (*Some; M* = 3.71), but both content area teachers used them more than math and science teachers reporting, on average, less than 2.0 (*Very Little*). English and social studies teachers used team projects more, with means, on average, being above 3.0 (*Some*), than math and science, with means less than 3.0 (*Very Little*). The results suggest that English and social studies teachers assess similar to one another, while math and science differ with math teachers use of assessments varying the most.

Table 22

Sample Size, Means, Standard Deviations for Assessment Types by Core Content Area

	English (*N* = 67)		Math (*N* = 64)		Science (*N* = 65)		Social Studies (*N* = 52)	
Item	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*	*Mean*	*SD*
Summative	4.46	.80	4.44	.85	4.54	.99	4.46	.83
Designed by self	4.58	.91	4.61	.83	4.20	1.12	4.56	.94
Objective	4.01	1.07	3.72	1.23	4.54	.92	4.46	.83
Major exams	3.40	1.07	3.69	1.11	4.02	1.17	3.92	1.15
Essay	4.13	1.04	2.53	1.54	2.83	1.21	3.71	1.07
Project individual	3.94	1.07	2.56	1.02	3.18	1.17	3.35	1.01
Formative	3.30	1.35	2.80	1.20	3.12	1.24	3.27	1.07
Project team	3.12	1.14	2.48	.89	2.97	1.05	3.08	1.01
Test made by others	2.76	1.14	3.14	1.21	2.89	1.34	2.67	1.18
Performance	3.79	1.08	1.89	.82	2.69	1.19	3.04	.93
Authentic	2.67	1.21	2.33	1.02	2.98	1.08	2.52	1.04
Publisher or district designed	2.19	1.06	2.13	1.18	2.29	1.09	2.35	1.15

How Assessment and Grading Factors Vary by Content Area

One-way ANOVAs were conducted for each item on the instrument to compare the differences between the core content areas on the various factors teachers use in assessment and grading. Statistically significant differences in means were followed-up with post hoc comparisons using the Tukey HSD test and effect sizes were calculated using Cohen's *d*. Because multiple tests are being conducted, there is an inflation in the Type I error rate. The one-way ANOVA examined the following hypotheses:

H_0: There are no significant differences in means among each of the four core content areas (English, math, science, and social studies).

H_1: There are significant differences in means among the four content areas.

Among achievement factors, see Table 23, there were statistically significant difference in means, at the $p < .05$ level, between the core content areas only in their use of objective mastery [$F(3, 244) = 3.166$, $p = .025$] and homework graded for correctness [$F(3, 244) = 10.207$, $p = .025$] in their grading practices. Post hoc comparisons indicated that math teachers use objective mastery (*Quite a bit*; $M = 4.17$, $SD = 1.03$) less than science teachers (*Quite a bit*; $M = 4.72$, $SD = .91$) with a moderate practical significance ($d = .58$). Among subject area teachers that grade homework for correctness, math did so the least (*Very Little*; $M = 2.47$, $SD = 1.17$) of all four content areas, English (*Some*; $M = 3.60$, $SD = 1.31$, $d = .79$), science (*Some*; $M = 3.26$, $SD = 1.40$, $d = .55$), social studies (*Some*; $M = 3.58$, $SD = 1.35$, $d = .88$), and has a practical significance ranging from moderate to large size. The collective results suggest math teachers grade based on

objective mastery significantly less than science teachers and they also use graded homework the least out of all the core content areas.

Table 23

Results of Analyses of Variance, Post Hoc Analyses, and Effect Size (Cohen's d) for Achievement Factors

Item	*Df*	*Mean Square*	*F*	*p*	*Tukey's Post Hoc*
Percent correct	3, 244	1.396	1.034	.378	
Objective mastery	3, 244	3.398	3.166	.025	Math < Sci (ES = .58)
Academic performance	3, 244	.997	.997	.395	
Homework quality	3, 244	17.457	10.207	<.001	Math < Eng (ES = .79) Math < Sci (ES = .55) Math < SS (ES = .88)

Note. Significant at $p < .05$. Sci = Science, Eng = English, and SS = Social Studies

Among the 15 non-achievement factors, only two were found to have statistically significant differences in means between the content areas. The content area teachers differed in their use of student's work habit in grade calculation [$F(3,244) = 2.924$, $p = .035$] and being required to adhere to a specific policy limiting the number of each letter grades that can be assigned [$F(3,244) = 3.378$, $p = .019$], see Table 24. After post hoc comparison, social studies used work habit (*Very Little*; $M = 2.63$, $SD = 1.31$), as well as being required to use a letter policy (*Not at all*; $M = 1.94$, $SD = 1.41$), more than math (Work Habit – *Very Little*; $M = 2.03$, $SD = 1.01$; Letter Policy – *Not at all*; $M = 1.28$, $SD = .88$) and had a moderate practical significance of .52 and .56 respectively. The results indicate that while the two subject areas vary significantly in their use of these non-achievement factors, they are reported to be used less than 3.0 (*Very Little)*, on average, and likely to have little impact in the final grades.

Table 24

Results of Analyses of Variance, Post Hoc Analyses, and Effect Size (Cohen's d) for Non-Achievement Factors

Item	*df*	*Mean Square*	*F*	*p*	*Tukey's Post Hoc*
Zeros	3, 244	3.299	1.495	.217	
Ability	3, 244	1.151	.747	.525	
Effort	3, 244	2.115	1.348	.259	
Attention or participation	3, 244	3.101	1.775	.152	
Homework completion	3, 244	.963	.541	.655	
Borderline grades	3, 244	2.744	1.454	.228	
Improvement	3, 244	1.360	.765	.515	
Work habit	3, 244	4.230	2.924	.035	Math < SS (ES = .52)
Extra credit academic	3, 244	2.128	1.691	.169	
Performance of current students	3, 244	.898	.737	.531	
Behavior	3, 244	1.206	1.138	.334	
Extra credit non-academic	3, 244	.229	.779	.506	
Letter policy	3, 244	4.567	3.378	.019	Math < SS (ES = .56)
Performance of past students	3, 244	1.928	2.027	.111	
Grades other teachers assign	3, 244	.311	.891	.446	

Note. Significant at $p < .05$. Sci = Science, Eng = English, and SS = Social Studies

The factors in Table 25 examine the beliefs and other factors impacting what final grade a teacher assigns, and there were no statistically significant differences between the subject area mean scores. These results suggest that the factors impact the final grades teachers assign similarly.

Table 25

Results of Analyses of Variance for Final Grade Factors

Item	*df*	*Mean Square*	*F*	*P*
Grades represent achievement only	3, 244	.032	.048	.986
Grade inflation causing grad rate increase	3, 244	.838	.537	.658
Minimum grading policy limits assignment of low grades	3, 244	2.405	.780	.506
Administrators and teachers judge grades assigned	3, 244	.814	.526	.665
Credit recovery encourages grade bumping	3, 244	3.464	1.417	.238

Note. Significant at $p = .05$

The mean differences within the assessment types used were found to be statistically significant in seven out of the 12 items on the instrument (see Table 26 for ANOVA results and Table 27 for Post Hoc analysis). The use of objective assessment varied significantly among the core content area teachers [$F(3,244) = 8.758$, $p < .001$] with science (*Quite a bit*; $M = 4.54$, $SD = .92$, $d = .71$) and social studies (*Quite a bit*; $M = 4.46$, $SD = .83$, $d = .71$) teachers using them more than math teachers (*Some*; $M = 3.72$, $SD = 1.23$), while English teachers (*Quite a bit*; $M = 4.01$, $SD = 1.07$, $d = -.57$) used them less than science teachers. The data suggests English teachers and math teachers use objective assessments similarly while social studies and science teachers used them the most.

Teachers reporting the use of essays on assessments were also found to have a large statistically significant difference with English teachers (*Quite a bit*; $M = 4.13$, $SD = 1.04$) using them more than math (*Very Little*; $M = 2.53$, $SD = 1.54$, $d = -1.22$) and science teachers (*Very Little*; $M = 2.83$, $SD = 1.21$, $d = 1.16$), but also using them similarly to social studies teachers (*Some*; $M = 3.71$, $SD = 1.07$). Additionally, math

teachers (d = -.89) and science teachers (d =-.77) were found to use essays on their exams less than social studies with a large negative practical effect. The data suggests English and social studies teachers use essay exams significantly more than the other core content areas.

The use of performance assessments also varied significantly [$F(3.244) = 39.110$, $p < .000$) with English teachers ($M = 3.79$, $SD = 1.08$) using them more than the other content areas with a moderate practical significance when compared to social studies (*Some*; $M = 3.04$, $SD = .93$, $d = -.75$), and a large practical significance between math (*Not at all*; $M = 1.89$, $SD = .82$, $d = -1.21$) and science (*Very Little*; $M = 2.69$, $SD = 1.19$, $d = -.97$). Also, science (*Very Little*; $M = 2.69$, $SD = 1.19$, $d = .79$) and social studies (*Some*; $M = 3.04$, $SD = .93$, $d = 1.31$) teachers were found to have a large practical effect when compared to math teachers' (*Very Little;* $M = 1.89$, $SD = .82$) use of performance tests.

The use of projects as assessments, both individual [$F(3, 244) = 18.2101$, $p < .000$] and as teams [$F(3.24) = 5.136$, $p = .002$], was found to have statistically significant differences among the content areas. The use of individual projects was implemented the most by English teachers (*Some*; $M = 3.94$, $SD = 1.07$), and the least by math teachers (*Very Little*; $M = 2.56$, $SD = 1.02$, $d = -1.32$) with a very large practically significant difference, while science (*Some*; $M = 3.18$, $SD = 1.17$) and social studies (*Some*; $M = 3.35$, $SD = 1.01$) were statistically similar in their use but also have a moderate negative practical significance in comparison to English teachers. Regarding the use of team projects, teachers of math classes (*Very Little*; $M = 2.48$, $SD = .89$) had a moderate

negative practical effect and used them the least. Overall, in the use of projects as assessments in general, there were similarities as math teachers used them the least among all the core content areas while English teachers used individual projects the most.

The final two assessment types, the use of major exams [$F(3, 244) = 3.809$, p = .011] and the use of authentic assessments [$F(3, 244) = 4.099$, $p = 0.007$], were found to have statically significant differences with a moderate practical effect size for science (*Quite a bit*; $M = 4.02$, $SD = 1.17$, $d = .55$; *Very Little*; $M = 2.98$, $SD = 1.08$, $d = .62$, respectively) when compared to English teachers' use of major exams and math teachers' use of authentic assessments. Science teachers used major exams more than English teachers (*Some*; $M = 3.40$, $SD = 1.07$), while using authentic assessments more than math teachers (*Very Little*; $M = 2.33$, $SD = 1.02$). The results suggest that although science teachers reported using both major exams and authentic assessments more than other content areas, they used major exams similarly to math and social studies and used authentic assessments similarly to English and social studies teachers.

Table 26

Results of Analyses of Variance for Assessment Types

Item	*df*	*Mean Square*	*F*	*p*
Summative	3, 244	.124	.163	.921
Designed by self	3, 244	2.392	2.620	.051
Objective	3, 244	9.296	8.758	<.001
Major exams	3, 244	4.817	3.809	.011
Essay	3, 244	35.852	23.419	<.001
Project individual	3, 244	20.966	18.201	<.000
Formative	3, 244	3.324	2.206	.088
Project team	3, 244	5.402	5.136	.002
Test made by others	3, 244	2.508	1.683	.171
Performance	3, 244	40.565	39.110	<.001
Authentic	3, 244	4.919	4.099	.007
Publisher or district designed	3, 244	.581	.465	.707

Note. Significant at $p < .05$.

Table 27

Results of Post Hoc Analyses and Effect Size (Cohen's d) for Assessment Types

Item	*Tukey's Post Hoc*
Objective	Math < Sci (ES = .76)
	Math < SS (ES = .71)
	Sci > Eng (ES = -.53)
Major exams	Eng < Sci (ES = .55)
Essay	Eng > Math (ES = -1.22)
	Eng > Sci (ES = -1.16)
	SS > Math (ES = -.89)
	SS > Sci (ES = - .77)
Project individual	Eng > Math (ES = -1.32)
	Eng > Sci (ES = -.67)
	Eng > SS (ES = -.57)
	Math < Sci (ES = .57)
	Math < SS (ES = .77)
Project team	Math < Eng (ES = .62)
	Math < Sci (ES = .50)
	Math < SS (ES = .62)
Performance	Eng > Math (ES = -1.98)
	Eng > Sci (ES = -.97)
	Eng > SS (ES = -.75)
	Math < Sci (ES = .79)
	Math < SS (ES = 1.31)
Authentic	Math < Sci (ES =.62)

Note. Includes statistically significant items from Table 26. Sci = Science, Eng = English, and SS = Social Studies

Conclusions

Table 28 provides a summary of practically significant factors impacting grades. The use of factors impacting the final grade, both achievement and non-achievement, as well as other beliefs surrounding its assignment, was found to only have statistically significant results between content areas on four (objective mastery, homework quality, work habit, and letter policy) of the 24 items. The only practically significant difference

among content areas was the lack of use of graded homework by math teachers. These results indicate that teachers grade similarly across the content areas.

Table 28

Summary of Grading Factors with Highest Rating

Item	*Mean*	Percent	Subject Used	Direct Input
Grades represent achievement only (Item 32)	4.54	99.4	All	No
Percent Correct (Item 1)	4.47	95.6	All	Yes
Grade inflation causing grad rate increase (Item 34)	4.47	92.7	All	No
Zeros (Item 4)	4.42	87.6	All	Yes
Objective Mastery (Item 2)	4.41	96.0	All*	Yes
Academic Performance (Item 3)	4.21	96.1	All	Yes
Effort (Item 6)	3.44	81.0	All	Yes
Minimum grading policy limits assignment of low grades (Item 36)	3.34	64.9	All	No
Homework Completion (Item 9)	3.29	72.5	All	Yes
Administrators and teachers judge grades assigned (Item 33)	3.28	71.3	All	No
Borderline grades (Item 10)	3.22	69.7	All	Yes
Homework Quality (Item 7)	3.21	66.5	Not Math*	Yes
Attention or Participation (Item 8)	3.15	68.5	All	Yes
Improvement (Item 11)	3.10	68.1	All	Yes

Note. Items not reported have $M < 3$ and percent reported > 50%. Sci = Science
* significant at $p < .05$.

Teachers agreed that their grades represented achievement, based on percentage correct, and objective mastered, but also indicated other factors, having an average mean of 3.0 (*Some*) or greater, impacted the assignment of their final grades. The most reported non-achievement factor is the use of zeros (*Quite a bit*; $M = 4.42$) for incomplete or missing assignments and followed by student effort, with homework (both graded and ungraded), attention and participation, and improvement being used as a direct input into student grades. These factors were also used together to increase

borderline student grades as part of the teacher's professional judgment. Other external factors impacting final grades are that teachers are limited by minimum grading policies for nine-week reporting and recognize and care that administrators and other teachers judge their professional effectiveness upon the grades they assign. The collective factors found to impact student grades suggest that teachers believe (*Quite a bit*; $M = 4.47$) that grade inflation is contributing to the overall improvement of the graduation rate in North Carolina, is valid.

Regarding assessment use, seven of the 12 items had statistically significant differences between the core content areas, but nine factors were found to be practically significant with means, on average, of 3.0 (*Some*) or more (see Table 29). The majority of high school teachers in the core content areas were found to create their own assessments which were major exams with objective items and summative in nature. Math teachers were found to vary the least in the types of assessments given while English teachers, similar to social studies, were found to use the widest variety.

Table 29

Summary of Assessment Types with Highest Rating

Item	*Mean*	Subject Used ($M \geq 3$)
Summative (Item 21)	4.48	All
Designed by self (Item 22)	4.48	All
Objective (Item 23)	4.17	All; Sci most >Math*
Major exams (Item 27)	3.75	All; Sci > Eng*
Essay (Item 24)	3.29	Eng & SS; Eng & SS most*
Project individual (Item 26)	3.26	Eng, Sci, & SS; Eng most*
Formative (Item 20)	3.12	Eng, Sci, & SS
Project team (Item 29)	2.91	Eng & SS; Math least*
Performance (Item 25)	2.85	Eng & SS; Eng most*
Authentic	2.63	None; Sci *(M* = 2.98) > Math*

Note. Items by subject area reporting $M < 3$ are omitted. Sci = Science, Eng = English, and SS = Social Studies
* significant at $p < .05$.

Overall, grading factors used by high school teachers were found to be used similarly across all content areas, and the results reveal ten factors used as a direct input to determine final grades, while four factors represent teacher beliefs and impact the final grade through policy or perceptions of what others think of the final grades assigned. The assessments used were found to vary significantly among content areas with math differing the most and using only major exams with objective questions that they created. The major findings will be discussed in further detail in the following chapter, and it will also include practical significance, limitations, and suggestions for future research.

\

Chapter 5: Discussion

The current study examined grading practices at the secondary level to address the gap between increasing graduation rates and the preparedness of post-secondary students. The study provides a description of assessment and grading practices within the four core content areas, English, math, science, and social studies, of current classroom teachers and identifies relationships between the core content areas. The following research questions guided the study:

1. What factors do secondary teachers use in determining grades for students in the four core content areas, English, math, science, and social studies?

2. What types of assessments do secondary teachers use in grading students in the four core content areas, English, math, science, and social studies?

3. How do the types of assessment and factors used in grading students vary among English, math, science, and social studies teachers?

The chapter includes a discussion of the major findings, practical significance, limitations, and suggestions for future research.

Major Findings

Grading Factors

Like the findings of Allen (2005), Brookhart (2015), Kunnath (2017), Pilcher (1994) Randall and Engelhard (2009), and Randall and Engelhard (2010), the grades teachers assigned are a conglomerate of various factors and mainly use achievement but

also incorporate non-achievement factors. High school teachers were found to use mainly 10 factors, rank ordered by mean as follows: percent correct, use of zeros, objective mastery, academic performance, effort, homework completion, the adjustment of borderline grades, homework quality, the degree that a student pays attention or participates, and improvement since the beginning of the year, in their grades. These factors were very similar to the factors identified in the literature, but the current study found that zeros were used more frequently than the findings in McMillian (2001) and ability was not found to be a factor impacting the grades of high school students, which is contradictory to the findings of Brookhart (1994), Bursuck et al. (1996), Cross and Frary (1999), Kunnath (2017), McMillian (2001), Pilcher (1994), Randall and Engelhard (2009), and Randall and Engelhard (2010).

In the current study, the use of the factors performance compared to a percentage correct, mastery of learning objectives, the use of academic performance, and the use of zeros in grade calculation for missing and incomplete assignments compared to the findings of Kunnath (2017), Pilcher (1994), and McMillian (2001) suggests that grades mainly represent achievement, but the incorporation of zeros introduce inaccuracies into the final grade. Teachers may believe they are teaching students responsibility by the use of zeros, but they do little to inform the overall progress toward meeting learning objectives or identifying any areas of concern a student may have. Teachers, instead, use zeros to penalize the student for their lack of compliance, preventing the collection and use of data to inform instruction and instead documents their perceived lack of effort or inability to following instructions, which lead to gaps in learning.

Most teachers reported they use the following factors but to a lesser degree: effort, homework completion, adjusting borderline grades, homework quality, attention and participation, and the degree a student improves since the beginning of the year which compare to the findings of Brookhart (1994), Brookhart (2015), Bursuck et al. (1996), Cross and Frary (1999), Duncan and Noonan (2007), Kunnath (2017), McMillian (2001), Pilcher (1994), Randall and Engelhard (2009), and Randall and Engelhard (2010). Nearly all these factors are used to document student behaviors that are not related to achievement. The inclusion of these factors introduces more inaccuracies, which prevent the final grade from being appropriately used in decision making.

The use of homework in the final grades is considered an achievement factor when it is graded and was found to be used in this manner more than including it for completion by Brookhart (1994), Cross and Frary (1999), and McMillian (2001), but is contradictory to the findings of the current study. Homework quality, or the grading of homework, provides relevant information regarding the progress of student learning and was expected by the researcher to be used more than homework completion but findings suggest that the simple act of completing the work is more important than student performance. The findings were even more pronounced for Math teachers reporting minimal grading of homework while simultaneously reporting the use of homework completion the most. The data suggest that the use of homework is more related to student effort or behaviors than accurately reflecting student knowledge and understanding, and as a result, the teachers cannot shape instruction around student needs.

A factor of interest, the adjustment of borderline grades, adds an additional layer to grading by impacting students inconsistently and ultimately creating a grading system that unfairly compares students. The use of behaviors in the final grade, such as effort and paying attention, are used as a direct input to student grades and may impact students equally, but when these factors are combined, essentially using them twice, in borderline cases where students' grades may be *bumped* up based upon teacher professional judgement as suggested by Ornstein (1994). The findings of Cross and Frary (1999) indicate that lower performing students are impacted the most and can be used as a preventive measure to failure, but the example on the instrument, see item 10 in Appendix A, did not include a failing grade suggesting the use of this factor at all degrees of grade assignment. The adjustment of borderline grades for various students, at differing degrees of performance as measured by the current grading scale, create a system where student ability is overrepresented and misinforms those who rely on accurate grading measures, such as for enrollment decisions at post-secondary institutions or in decision-making for student course placement, ultimately, negatively impacting the student.

Further complicating the assignment of grades is the finding, aligning with Kunnath (2017), that teachers agree that effectiveness is judged based upon the grades they assign. This realization can coerce teachers into incorporating grading practices that include factors to increase student passing rate or to produce a grade distribution that satisfies administrators and other teachers within the department. The use of the grading practices revealed by the current study impacts student grades in varying ways dependent

upon student characteristics causing some students who perform well on assessments but, for example, do not turn in homework or participate in class, have a poor grade and the opposite may also be true. It is possible that districts or schools recognize the issues with grading at the high school level which would explain the use of a minimum grading policy found within the current study or they may be more concerned with passing rates and increasing student graduation rate.

The findings of this study suggest that teachers have a misconception regarding the grades they assign. While teachers agree that the grades assigned by teachers across the state of North Carolina are inflated and contribute to the continually increasing graduation rate, they also believe their grades represent achievement, as opposed to any other factors, but the data suggests that the grades are a *hodgepodge* of factors aligning with the findings of Allen (2005), Brookhart (1994), Cross and Frary (1999), and Guskey (2006). The current grading practices, if they continue, will impact students negatively, especially those having a GPA of 2.8 or higher, by allowing them to be exempt from remedial courses under the College Ready Graduates legislation when they do not possess the skills their GPA suggest. As a result, the high schools of North Carolina will produce students appearing to meet the college readiness requirements who are still underprepared for post-secondary studies.

Assessment Types

The majority of teachers in North Carolina reported, supporting the findings of Duncan and Noonan (2007), Frey and Schmitt (2010) and McMillian (2001), using summative assessments that were objective, and were created by themselves. The current

study also found major exams were frequently used by most teachers, and in combination with the previous statement, it is logical because these types of assessments mirror the End-of-Course Tests and North Carolina Final Exams required to be given in the core content areas. The teachers, by using these assessment types, can allow the students to practice test-taking skills and to become accustomed to taking them with timing constraints. The pitfalls of using these types of assessments that mirror the state exams, are that they limit the ability of teachers to determine student weaknesses in smaller components or topics within each of the core content areas by requiring a more global approach preventing the identification of specific areas of student misunderstandings (Hamilton et al., 2008).

Most teachers also reported the use of essays, individual projects, and formative assessments in their practices at the secondary level. The use of essays and individual projects varied significantly across content areas with English teachers using them the most of all content areas while social studies reported using them similarly, which is comparable to the findings of McMillian (2001). These results are not surprising given the interpretive nature of the English content area emphasizing clear communication, and similarly, social studies requires the analysis of preceding actions or events which are the precursor to some larger outcome, whereas math and science are more objective in nature having a sole answer or explanation.

The use of formative assessments was an interesting revelation because it suggests the NC Falcon initiative beginning in 2010 may have had some influence on teacher assessment practices. Almost 70% of teachers reported using formative

assessments, see Appendix J, resembling the findings of Frey and Schmitt (2010). The results suggest that teachers are using the data to inform instruction, document student progress toward objective mastery, and allow students to struggle with concepts during instruction penalty free. Teachers could improve grading practices while simultaneously informing instruction by using homework for its intended purpose, and encouraging students to practice or work with new information they have not yet mastered (Wormeli, 2006), rather than being a direct input into grades. The penalization for being incorrect and the reward for completion encourages cheating rather than encouraging students to problem solve or critically think.

The use of performance assessments was the least reported assessment type used by the participants in the current study, which was largely different from the findings from the literature. Bol et al. (1998) found that nearly half of teachers reported they used performance assessments at the highest rating, while McMillian (2001) found 39% of teachers used them *Quite a bit* as compared to only 29% of teachers in the present study ($M \geq 4$; See Appendix J). Performance assessments are important not only for examining content areas, but also specific skills such as public speaking, oral presentations, and writing, that are essential to being successful in post-secondary studies.

These results suggest that North Carolina teachers use more traditional assessment methods that could be attributed to the increased demands placed on accountability due to the passage of NCLB in 2001 and the adoption of North Carolina's school accountability model. These legislative requirements emphasize the use of objective assessments and then encourage teachers to replicate these exams allowing students to practice, making it

easier to prepare students for taking the state-mandated assessments. The indirect result of the standards-based movement encourages content mastery but also de-emphasizes the focus on cross-cutting skills, equally important to post-secondary studies, such as writing, presentation skills, and problem-solving. These skills are reserved for use in only a few subjects, even then, the skills are only used some of the time which impacts student performance in high school and post-secondary studies.

Differences Between Core Content Area Teachers

The comparison of the content areas across the 36 items revealed statistically significant results with medium and large effect sizes in 11 of the items. Most of the practically significant findings were in the types of assessments used, only one grading factor (homework quality), and none of the items related to the final grade were statistically significant. The results suggest that the grading factors and beliefs related to the final grade are consistent across the core content areas, and the global descriptive results best represent teachers' use of these factors. The following sections discuss the practically significant results of assessment types by content area.

Math teachers differed the most among all the content areas in the use of grading factors and assessment types, in agreement with Duncan and Noonan (2007) and McMillian (2001). They were most likely to give all assessment types the least among all content areas, but the most likely to give objective assessments when compared to all other assessment types, differing from the findings of Duncan and Noonan (2007). A possible explanation that math teachers were found to vary significantly from the other content areas, but yet use these assessments the most, could be due to a poorly-written

survey item related to objective assessments which did not provide examples that included completion or questions in which students supply a single answer. As a result, the survey item as written may not accurately reflect the use of objective assessments among math teachers and this may suggest there are no differences among the content areas in the use of objective assessments.

Science teachers were found to use major exams more than English teachers and objective assessments more than any other content area, similar to the findings of McMillian (2001). They also reported using authentic assessments significantly more than math teachers and the mean response rate suggest possibly more than any content area. These results are likely due to the science content being more objective than English and social studies, and major exams are used similarly to social studies because the content can be broken down into larger curricular components such as time periods or concepts.

English and social studies teachers were found to be similar in their use of essays. They used them significantly more than math and science teachers, which is similar to the findings of Duncan and Noonan (2007), but differs slightly from McMillian (2001) in that social studies teachers did not use them significantly different from the other content areas. English teachers used performance assessments and individual projects more significantly than any other subject. Even though social studies teachers used them both less than English teachers, they were found to use them more than math teachers. English teachers were also found to use major exams significantly less than science teachers, and the descriptive data suggest they could use them the least. The results indicate that

although English teachers are similar to social studies in their use of assessments, they employ the widest variety of exam types than any other subjects and are less likely to use traditional assessment methods used by the current accountability model.

Overall, the data suggest English and social studies teachers are more interested in the cross-cutting skills measured by these assessment types as well as their overall understanding of the content. They attempt to discern the student's ability to interpret and explain their understanding of smaller components of the curriculum, possibly around specific pieces of literature or eras, which aligns with the content areas being more subjective in their interpretation of meaning. Science teachers, while interested in these skills, are also concerned with authentic assessments that examine real-life problems or experiences to measure students' critical thinking and application. Math, being the significant outlier and focusing solely on objective assessments, may do so because the content area is reliant on basic skill development that builds throughout the year whereas, in the other content areas, the material could allow for a more natural compartmentalization of information into areas such as time periods or events, topics or processes, or an analysis of a piece of literature. In summary, the results suggest English and social studies teachers are the most alike in the use of their assessments, while math teachers varied the least.

Practical Significance

The literature suggests that teachers use four major factors (i.e., achievement, ability, behavior, and effort), which were also found in the present study with the exception that ability was found to be used considerably less. The similarities to the

findings of Brookhart (1994) suggest that grading practices have remained relatively unchanged since 1984 and the transition to a standards-based curriculum has not significantly impacted the way teachers grade. The grading systems used by North Carolina high school core content area teachers do not represent achievement only, they are a *hodgepodge* of factors which do not and cannot in their current reporting methods clearly delineate what students know and what they do not.

The grades teachers assign are based mainly on achievement, but they also use zeros, effort as a direct input, student behavior and improvement to bump up borderline grades, leading to a skewed representation of student performance. For example, the use of zeros is a motivational technique used to punish students (McMillian, 2001) for their lack of organizational skills, effort, or understanding masking the student's true understanding of the content. In addition, the use of homework for completion rather than quality simply rewards compliant behavior, does little to inform instruction, and inflates student performance. Essentially grades are used to motivate and modify behavior, and if students behave, pay attention, work hard, and complete their homework, they should not fail a course (Bursuck et al., 1996, Willingham et al., 2002).

The grades produced as a result of the current grading system do not represent student performance in a class or subject, contradictory to what teachers believe, rather the grades represent how well students met the predetermined criteria established in the syllabus. The effect is compounded when each teacher develops their own grading system absent of any consistency within the school or even department. The use of the grades developed by the current system to inform parents, make educational decisions at

the high school level, or by admission counselors at post-secondary institutions to make placement decisions is inappropriate because the grades do not accurately represent student achievement.

Regarding assessments, teachers generally create their own assessments that mirror the state developed tests students take at the end of the year (Frey & Schmitt, 2010). The assessments are mainly used to measure student learning at the end of a large unit that makes it very difficult, if not impossible, to use this information to guide instruction because this data is collected after the unit has been completely presented. District and building level administrators should work to increase the use of formative assessments in the classroom because increasing their appropriate use and incorporating homework as a tool rather than behavior modification would yield better results on large unit tests. Teachers could use this information to address student weaknesses earlier rather than being reactive by requiring students to *catch-up* while simultaneously learning new material, which can perpetuate learning gaps if skills *previously learned* are foundational in the understanding of a topic.

Because assessments are created by classroom teachers, it is important that they are trained in proper assessment techniques (e.g., the gathering and interpreting of data), and vital they understand the content standards to ensure their instruction is linked to their assessments facilitating student learning (Duncan & Noonan, 2007). In addition, Carter (2007) suggests using the *Total Instructional Alignment* method to ensure that the classroom assessments used are appropriately aligned with the content standard because even with the improvement of teacher assessment practices, they may still fail to properly

assess the appropriate content. The link between standards, assessments, and data analysis ensures the teachers are able to adjust their instruction as necessary to meet their students' needs. In addition, the type of assessment can also be used to measure cross-cutting skills, such as public speaking, writing, and presentations. These skills are important for students preparing for post-secondary studies and should be implemented across all content areas, but current teachers use assessments inconsistently, further blurring the meaning of a grade across content areas. As a result, all teachers would benefit from professional development in the development of assessments to create tests that are better linked to state standards as well as how to properly use the data collected from both formative and summative assessments. Although assessment training may improve instruction and test development, the training has not been found to impact the use of a *hodgepodge* grading system (Cross & Frary, 1999), which requires an alternative approach.

Randall and Engelhard (2010) suggest implementing an achievement-only grading policy, but they concede that after numerous attempts by grading experts to change grading practices, they remain unchanged. The authors suggest the possible solution is to change reporting methods to address the shortfalls of current grading practices. Brookhart (2015), Brookhart (1994), and Randall and Engelhard (2010) believe the new grading system should report achievement and non-achievement factors separately and 80% of teachers and 70% of students agree (Cross & Frary, 1999). An alternative system, such as standards-based grading, would address the concerns of

accurate grades and document student learning by encompassing the shift to a standardized curriculum while also linking assessment, grading, and instruction.

Standards-Based Grading (SBG) modifies the *hodgepodge* system in place through separating grades into their three components: product, process, and progress, rather than only allowing the reporting of a single grade (Guskey 2006; Guskey, 2001; Muñoz & Guskey, 2015; Scriffiny, 2008; Wormeli, 2006). The SBG method would still use the traditional grade-point average calculation using the letter grade system and percentage grades because it has been traditionally thought of as the measure of achievement, while the non-achievement factors would be included under other categorical representations, such as homework completion, work ethic, and attendance, and reported as a separate grade (Guskey, 2006; Wormeli, 2006). The disaggregation of grades is necessary because "grades should reflect students' performance on specific learning criteria" (Guskey, 2006, p. 672), while still allowing teachers to incorporate the various factors they deem important. This process eliminates the need for grade weights or developing complex point systems, thus clarifying the meaning of a grade, decreasing subjectivity, and making grading easier (Muñoz & Guskey, 2015; Guskey, 2006).

In SBG, assessments are built for a purpose, assessing knowledge and skills, focusing on the specific state standard by which the learning was designed around. Guskey (2001), Guskey (2006), Muñoz and Guskey (2015), and Wormeli (2006) emphasize that teachers must have an intimate understanding of the standard course of study to be able to create specified learning targets, which anchor lessons and student learning activities, because performance is reported as student mastery of content

standards. SBG differs from the grading practices revealed by the current study in which redos, retakes, and late work are all considered an integral part of the learning process and therefore are included in the proficiency rating without penalty (Wormeli, 2006) because student learning is no longer attached to the typical timeline for instruction. In SBG, students can achieve mastery on the previously taught material at any time, and their grades reflect it.

Ultimately, SBG communicates student learning and achievement more accurately and is favored by parents because it provides more comprehensive feedback regarding their child and their performance in each class (Guskey, 2006; Wormeli, 2006). Also, college admission counselors prefer SBG because the transcript becomes a "robust document [that] presents a better and more discerning portrait of students' high school experiences" (Guskey, 2006, p. 674). Overall, this system of reporting provides more accurate and meaningful information regarding student performance which allows grades to be used more appropriately to inform instruction, which will strengthen the relationship between teacher reported grades and student performance on standardized assessments (Guskey, 2006; Muñoz & Guskey, 2015; Wormeli, 2006). "The end result is the transformation of the traditional approach into a standards-based report card that creates a straightforward link between curriculum and assessment" (Muñoz & Guskey, 2015, p. 66).

Limitations

The present study focused on eight of the 115 school districts throughout North Carolina, representing approximately seven percent of the state's districts and total high

schools. The demographics of the participants also vary significantly from the reported demographics of the state's teachers, with 93.1% of the participants being white as opposed to 49.5% represented statewide (Division of School Business, Financial and Business Services & North Carolina Department of Public Instruction, 2016, January). Taking these factors into consideration, the results discussed may not accurately represent North Carolina's teachers use of assessment and grading factors.

In addition, the survey relies on self-reported assessment and grading practices occurring during the spring semester of 2018 and may result in teachers reporting socially acceptable practices rather than those implemented, even after the assurance of confidentiality. Also, all items in the survey were required to be responded to which could have resulted in a lower return rate should a participant not be willing to answer a single question. Lastly, because the survey was distributed prior to the end of the semester, it may not accurately represent the assessment and grading practices used to determine the final grades after submission.

Suggestions for Future Research

The current study examining assessment and grading practices of North Carolina high school teachers would benefit from replication due to the sample size being too small, geographically isolated, and, overall, not representative of the state's teaching force. The sample included a disproportionate representation of grading practices due to nearly all of the reporting participants being white, and through the inclusion of other racial groups, would provide a more accurate description of assessment and grading practices. In addition, the participants were in close geographic proximity and as

neighboring districts; they may adopt similar practices due to close working relationships, and when in combination with the small sample size, would not reveal differences in grading factors used in districts from other geographic regions.

The variation among the content areas and their use of assessment types suggests that the nature of the specific subject may influence the types used by teachers. Future researchers would also find it worthwhile to examine the link between the assessments used and the skills taught within each content area to determine if cross-cutting skills such as writing, speaking, and presentation skills are taught in isolation or are relatively widespread across all content areas. Should these skills be taught by all teachers, then the results may suggest that alternative assessment types are not used due to the traditional objective assessments being used in the accountability models to determine student proficiency and growth, but rather based in the beliefs of assessments of teachers. This would be important in determining if the emphasis is on core content alone or the development of skills that would be beneficial after high school.

Another interesting aspect of practices becoming popular is the use of credit recovery in high schools. Course credit recovery is defined in North Carolina State Board of Education policy CCRE-001, as

> a block of instruction that is less than the entirety of the Standard Course of Study for that course. Credit recovery delivers a subset of the Standard Course of Study or blueprint of the original course in order to specifically address deficiencies in a student's mastery of the course and target specific components of a course necessary for completion.

The relatively new practices is an area of importance when examining the disconnect between the increasing high school graduation rate and the preparedness for post-secondary studies because credit recovery allows students to earn credit, after failing a class, by taking an abbreviated course facilitated via an online program, a cluster of assignments related to the content the student did not master, or something similarly developed by the school. The future success of these students hinges on what they learn in high school, but the use of credit recovery may not address weaknesses, simply act as an avenue to increase graduation rates, and leaves concerns regarding it being completed to fidelity. In addition, the use of credit recovery by potential student athletes prevents them from pursuing division I or II athletics because it fails to meet the requirements of the NCAA clearing house course requirements and thus limits opportunities for students. It is necessary to study the impact of credit recovery at the high school because the improved graduation rates increase school accountability grades but may leave students unprepared for their future.

Finally, in search of support for the change to SBG, there are few empirical studies examining its effectiveness. Upon initial investigation, it appears that standards-based progress reports correlate higher with achievement on standardized assessments than traditionally assigned grades (Welsh, D'Agostino, and Kaniskan, 2013). Grade reformers would benefit from studying the impacts of SBG on student achievement because the findings of Welsh et al. (2013) suggest that, should teachers adhere to the tenets of SBG and break away from the *hodgepodge grading system*, they will be able to more accurately report student achievement as measured by the standardized assessments used in the accountability models and college placement exams.

Concluding Thoughts

Grades throughout history have been plagued with concerns regarding subjectivity, but even with the pressures of student, teacher, and school accountability, grades still do not represent their traditional purpose. Nevertheless, the traditional meaning grades denote are communicated regularly even when it is evident non-achievement factors are incorporated and skew student GPA. Until current grading practices are addressed through achievement-only practices, or alternative grading processes reporting non-achievement factors independently of a final grade, such as in standards-based grading, students will continue to graduate unprepared for post-secondary work.

References

ACT. (2017). *The condition of college & career readiness 2017: North Carolina key findings.* ACT, Inc. Retrieved from https://www.act.org/content/dam/act/unsecured/documents/cccr2017/North-Carolina-CCCR-2017-Final.pdf

Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House, 78*(5), 218-223.

Anderson, R. S. (1998). Why talk about different ways to grade? The shift from traditional assessment to alternative assessment. *New Directions for Teaching and Learning*, *1998*(74), 5–16.

Black, P., & Wiliam, D. (1998). Inside the black box, raising standards through classroom assessment. *Phi Delta Kappa,* 80(2) 139-44.

Black, P., & Wiliam, D. (2003), 'In praise of educational research': formative assessment. British Educational Research Journal, 29, 623–637. doi:10.1080/0141192032000133721

Bol, L., Stephenson, P.L., O'Connell, A.A., & Nunnery, J. A. (1998). Influence of experience, grade level, and subject area of teacher's assessment practices. *The Journal of Educational Research, 91*(6), 323-330.

Brookhart, S. M. (1994). Teachers' grading: Practice and theory. *Applied Measurement in Education, 7*(4), 279-31.

Brookhart, S. M. (2015). Graded achievement, tested achievement, and validity. *Educational Assessment, 20*(4), 268-296.

Bursuck, W., Polloway, E. A., Plante, L. Epstein, M. H., Jayanthi, M., & McConeghy, J. (1996). Report card grading and adaptations: A national survey of classroom practices. *Exceptional Children, 62*(4), 301-318.

Carter, L. (2007). *Total instructional alignment*. Bloomington, IN: Solution Tree.

Center for Community College Student Engagement. (2016). *Expectations meet reality: The underprepared student and community colleges*. Austin, TX: The University of Texas at Austin, College of Education, Department of Educational Administration, Program in Higher Education Leadership.

Chen, X., & Simone, S. (2016, September). Remedial Course taking at U.S. Public 2- and 4-Year Institutions: Scope, Experiences, and Outcomes(Statistical Analysis Report). Washington, DC:U.S. Department of Education Web Publishing. Retrieved from: https://nces.ed.gov/pubs2016/2016405.pdf

Clinedinst, M., Koranteng, A., & Nicola, T. (2015). State of College Admission. *National Association of College Admission Counseling.* Retrieved from http://www.nacacnet.org/studentinfo/articles/Pages/Factors-in-the-Admission-Decision.aspx

Clotfelter, C., Ladd, H., Muschkin, C., & Vigdor, J. (2015). Developmental Education in North Carolina Community Colleges. Educational Evaluation and Policy Analysis, 37(3), 354–375.

Cross, L. H., & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education. 12:1*, 53-72.

Curreton, L. W. (1971). The history of grading practices. *National Council of Measurement in Education,* 2(4), 1-9.

Curry, K. A., Mwavita, M., Holter, A., & Harris, E. (2016). Getting assessment right at the classroom level: using formative assessment for decision making. *Educational, Assessment, Evaluation, and Accountability, 28*, 89–104.

Cyrenne, P., & Chan, A. (2012). High school grades and university performance: a case study. *Economics of Education Review, 31*, 524—542.

Dillman, D. A., (2000). Mail and Internet Surveys: The Tailored Design Method. New York, New York. John Wiley & Sons, Inc.

Dillman, D. A., & Smyth, J. D., (2007). Design effects in the transition to web-based surveys. *American Journal of Preventative Medicine, 32*(5S), S90-S96.

Dillman, D. A., Smyth, J. D., & Christian, L.M. (2014). Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Designed Method. New York, New York. John Wiley & Sons, Inc.

Division of School Business, Financial and Business Services & North Carolina Department of Public Instruction. (2016, January). *Facts and Figures 2015-2016* [Data File]. Retrieved from: http://www.dpi.state.nc.us/docs/fbs/resources/data/factsfigures/2015-16figures.pdf

Duncan, C. R., Noonan, B. (2007). Factors affecting teachers' grading and assessment practices. *Alberta Journal of Educational Research, 53*(1), 1-21.

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

Education Week. (1994). Summary of the Improving America's Schools Act. *Education Week*, *14(10),* 18-19.

Frey, B. B., & Schmitt, V. L. (2010). Teachers' classroom assessment practices. *Middle Grades Research Journal*, *5*(3), 107–117.

Gall, M. D., Gall, J. P., & Borg, W. R., (2010). Applying Education Research (6th ed.). Boston, MA: Pearson.

Guskey, T. R. (2001). Helping standards make the grade. *Educational Leadership, 59(1),* 20-27.

Guskey, T. R. (2006). Making high school grades meaningful. *Phi Delta Kappan, 87(9).* 670-675.

Hadley, A. T. (1902). The meaning and purpose of secondary education. *The School Review, 10*(10) 729-741.

Hamilton, L. S., Stecher, B. M., & Yuan, K. (2008, December). Standards-Based Reform in the United States: History, Research, and Future Directions. Retrieved from: http://www.rand.org/content/dam/rand/pubs/reprints/2009/RAND_RP1384.pdf

Hongli, L. (2016) How is formative assessment related to students' reading achievement? Findings from PISA 2009, *Assessment in Education: Principles, Policy & Practice, 23*(4), 473-494.

Jaggars, S., & Stacey, G.W. (2014, January). *What we know about developmental education outcomes.* Retrieved from Community College Research Center website: http://ccrc.tc.columbia.edu/media/k2/attachments/what-we-know-about-developmental-education-outcomes.pdf

Kittleson, M. (1997). Determining effective follow-up of e-mail surveys. *American Journal of Health Behavior, 21*(3), 193–196.

Kunnath, J. P. (2017). Teacher grading decisions: influences, rationale, and practices. *American Secondary Education, 45*(3), 68-88.

Madaus, G. F., & O'Dwyer, L. M. (1999). A short history of performance assessment: Lessons learned. *Phi Delta Kappa International, 80*(9), 688-695.

McMillian, J. H., (2001). Secondary teachers' classroom assessment and grading practices. *Educational Measurement: Issues and Practice, 20*(1), 20-32.

Mertler, C. A. (2000). Assessing student performance: A descriptive study of the classroom assessment practices of Ohio teachers. *Education, 120*(2), 285-296.

Muñoz, 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 Association of College Admission Counseling. (2015). *Factors in the admission decision: Your high school grades are the #1 factor in the college admission decision.* Retrieved from http://www.nacacnet.org/studentinfo/articles/Pages/Factors-in-the-Admission-Decision.aspx

National Center for Education Statistics. (2015). [Data table illustrating high school graduates, by sex and control of school: Selected years 1869-70 through 2025-26]. Digest of Education Statistics. Retrieved from https://nces.ed.gov/programs/digest/d15/tables/dt15_219.10.asp

National Center for Public Policy and Higher Education and the Southern Regional Education Board. (2010). Beyond the rhetoric: Improving college readiness through coherent state policy. *National Center for Public Policy and Higher Education.* Retrieved from http://www.highereducation.org/reports/college_readiness/CollegeReadiness.pdf

National Commission for Excellence in Education. (1983). *A Nation at Risk: The Imperative for Educational Reform.* United States Department of Education.

Nord, C., Roey, S., Perkins, R., Lyons, M., Lemanski, N., Tamier,…Herrold, K. (2011). *The Nation's Report Card: America's High School Graduates* (NCES 2011-462). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

North Carolina State Board of Education. (2014, October). From the Board Room: Activities of the NC Board of Education. Retrieved from: http://stateboard.ncpublicschools.gov/sbe-newsletter/2014-newsletter/10newsletter.pdf/view

North Carolina Department of Public Instruction. (2015, January). 2014 Ready Accountability Background Brief Supplement: North Carolina Performance Grades. Retrieved from:http://www.ncpublicschools.org/docs/accountability/reporting/spgbckgrndpack15.pdf

North Carolina Department of Public Instruction. (2010, May). Implementation Guide: NC FALCON, North Carolina's Formative Assessment Learning Community's Online Network. Retrieved from: http://www.ncpublicschools.org/docs/accountability/educators/falconguide.pdf

North Carolina Department of Public Instruction. (2016). [Data tables illustrating specific 4-year cohort graduation] Cohort Graduation Rates. Retrieved from http://www.ncpublicschools.org/accountability/reporting/cohortgradrate

Ornstein, A.C. (1994). Grading practices and policies: An overview and some suggestions. *NASSP Bulletin*. 55-64.

Pilcher, J. K. (1994). The value-driven meaning of grades. *Educational Assessment, 2*(1), 69-88.

Randall, J., & Engelhard, Jr. G. (2009). Examining teacher grades using Rasch measurement theory. *Journal of Educational Measurement, 46*(1), 1-18.

Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education, 26,* 1372-1380.

Rattray, J., & Jones, M., C. (2007). Essential elements of questionnaire design and development. *Journal of Clinical Nursing, 16,* 234-243.

Riley, R. W. (1995, September). Archived Information. *The Improving America's Schools Act of 1994*. U.S. Department of Education. Retrieved from: https://www2.ed.gov/offices/OESE/archives/legislation/ESEA/brochure/iasa-bro.html

Rossides, D. W. (1984). What is the purpose of education? The worthless debate continues. *Change: The Magazine of Higher Learning,* 16(3), 14-46.

Sánchez-Fernández, J. , Muñox-Leiva, F., & Monotor-Rios, F. H., (2011). Improving retention rate and response quality in web-based surveys. *Computers in Human Behavior, 28,* 507- 514.

Schneider, J., & Hutt, E. (2014). Making the grade: a history of the A-F marking scheme. *Journal of Curriculum Studies. 46*(2), 201-224.

Scriffiny, P. L. (2008). Seven reasons for standards-based grading. *Educational Leadership, 66(2), 70-74.*

SPSS Inc. (2009). PASW Statistics, Student Version 18.0 [Computer Software]. Chicago: SPSS Inc.

State Board of Community Colleges. (2015, February 20) *Multiple Measures of Placement. [Policy Attachment]*. Retrieved from http://www.nccommunitycolleges.edu/sites/default/files/state-board/program/prog_04_multiple_measures_2-12-15.pdf

State Board of Community Colleges and the State Board of Education. (2018, March). Progress on the implementation of the career and college ready graduates program. [A report to the: Joint Legislative Committee], p. 1-15.

Stiggins, R. J. (1991). Facing challenges of a new era of educational assessment. *Applied Measurement in Education, 4*(4), 263-273.

Stiggins, R. J. (2017). The perfect assessment system. ASCD. Alexandria, VA. P. 1-120.

SuccessNC. (2013). NC community colleges final report. Raleigh, NC: US. NC Community College System.

University of North Carolina General Administration. (2015, February). The University of NC remedial and development activities report 2014. Retrieved from https://www.northcarolina.edu/sites/default/files/unc_remedial_developmental_activities_report_-_2014.pdf

University of North Carolina System. (2013). UNC Institutional Summary [Data set]. Retrieved from http://analytics.northcarolina.edu/aa/fallenr/sdf.instsumm.2013.pdf

Weimiao, F., & Yan, Z., (2009). Factors affecting response rates of web survey: a systematic review. *Computers in Human Behavior, 26*, 132-139.

Welsch, 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 and Practice, 32*(2). 26-36. DOI: 10.1111/emip.12009

Wiliam, D. (2011). *Formative assessment: definitions and relationships* [PowerPoint slides]. Retrieved from: http://www.dylanwiliam.org/Dylan_Wiliams_website/Presentations_files/What%20is%20formative%20assessment%3F.pptx

Willingham, W. W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, 1–37.

Wormeli, R. (2006). *Fair isn't always equal: Assessing and grading in the differentiated classroom.* Portland, ME: Stenhouse.

Appendix A

The Instrument

Directions: Please select one class that you are currently teaching and identify the following information based upon your selection.

Please identify the subject area that represents the class you have selected as directed above.

□ English

□ Math

□ Science

□ Social Studies

Please identify the class ability level that represents the class you have selected as directed above.

□ Remedial

□ Regular/College Prep

□ Honors

□ Advanced Placement

Please identify the grade level that best describes that best represents the class you have selected as directed above.

□ 9

□ 10

□ 11

□ 12

Directions: Please provide your rating on the following items numbered 1-32 based on the following statement: "To what extent will the final grades of students in your class, as described above, be based on..."

Factors Used in Determining Grades

1. Performance compared to a set scale of percentage correct (e.g. 90 = A)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

2. Specific learning objectives mastered

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

3. Academic performance as opposed to other factors (e.g. using achievement versus effort or ability)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

4. Inclusion of zeros for incomplete or missing assignments in the determination of the final grade

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

5. Ability levels of the students (e.g. Student's grade is increased if performance doesn't match ability)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

6. Student effort (how much the student tried to learn)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

7. Quality of completed homework (graded for correctness)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

8. Degree to which the students pay attention and/or participates in class

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

9. Completion of homework (not graded for correctness)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

10. Effort, improvement, behavior, and other "non-test" indicators for students with borderline grades (e.g. a grade of a 79 is adjusted up to an 80)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

11. Improvement of performance since the beginning of the year

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

12. Work habits and neatness

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

13. Extra credit for academic performance (completing extra assignments)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

14. Performance compared to other students in the class (e.g grading on a bell curve, or adjusting grades, such as adding points to bring up the class average)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

15. Disruptive student behavior

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

16. Extra credit for nonacademic performance (e.g., bringing in items for food drive)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

17. Formal or informal school or district policy of the percentage of students who may obtain As, Bs, Cs, Ds, Fs

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

18. Performance compared to students from previous years (e.g. grading by using previous years as a baseline)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

19. Grade distributions of other teachers (e.g. adjusting grades to match another teacher's grades)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

Directions: Please provide your rating on the following items numbered 20-31 based on the following statement: "To what extent will the final grades of students in your class, as described above, be based on..."

Types of Assessments Used

20. Assessments given during instruction (e.g. Formative-not included in the final grade)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

21. Assessments given after instruction (e.g summative)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

22. Assessments designed primarily by yourself

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

23. Objective assessments (e.g., multiple choice, matching, short answer)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

24. Essay-type questions

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

25. Performance assessments (e.g., structured teacher observations or ratings of performance such as a speech, paper, or oral presentation)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

26. Projects completed by individual students

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

27. Major exams (e.g. midterms, quarterly, large unit exams)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

28. Authentic assessments (e.g., "real world" performance tasks, such as case studies or performing a task such as a student designed experiment)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

29. Projects completed by teams of students

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

30. Assessments provided by publishers or supplied to the teacher (e.g., in instructional guides or manuals)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

31. Assessments created by other teachers (e.g. sharing among departmental colleagues or obtained through an internet search)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

Directions: Please provide your rating on items numbered 32-36.

The Final Grade

32. To what degree do you feel that your final grades represent student achievement in your classes?

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

33. How strongly do you agree with the statement: "Teachers recognize and care that administrators and other teachers judge their professional effectiveness upon the grades they assign."

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

34. How strongly do you agree that grade inflation contributes to the overall improvement in graduation rate in North Carolina? (Grade inflation is defined as the awarding of grades that are higher than the degree of student achievement, or learning, as measured against the content standards.)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

35. How strongly do you agree with the following statement: "The availability of credit recovery for students encourages me to adjust end-of-course grades." (e.g. 59 is rounded up to a 60 to prevent course failure)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

36. How strongly do you agree with the following statement: "Minimum grading policy prevents the assignment of significantly low grades." (e.g. cannot give a student a 30% in the first 9 weeks of a semester due to school or district policy)

□ Completely
□ Extensively
□ Quite a bit
□ Some
□ Very Little
□ Not at all

Directions: Please provide the following demographic information:

Please specify your gender.

□ Male

□ Female

Please specify your ethnicity or race.

□ Asian or Pacific Islander

□ Black or African American

□ Hispanic or Latino

□ Native American or American Indian

□ White

□ Multiracial

Please select the appropriate range that matches your years of experience as a high school teacher.

□ 0-5

□ 6-10

□ 11-15

□ 16-20

□ 20-25

□ 26+

Drawing Entry

Four $25 VISA Gift Cards

If you would like to enter to win one of four $25 VISA Gift Cards, please enter your email below. Winners will be notified via email by June 15th to obtain an address for the delivery.

_____

Appendix B

Introductory Email

Department of Educational Leadership

9201 University City Boulevard, Charlotte, NC 28223-0001

Dear XX,

My name is Nathan McLean, and I am a doctoral student at the University of North Carolina at Charlotte, I am working towards the completion of my dissertation, and XX schools has given me permission to contact you. It is my understanding that you are a current high school teacher who teaches English, math, science, or social studies. I am contacting teachers, such as yourself, to inquire about their assessment and grading practices and to identify areas of commonalities among the content areas to address the gap between increasing graduation rates and the preparedness of incoming post-secondary students.

The results from this study will help parents, teachers, administrators, and college/university counselors better understand how grades vary across the content areas and how to interpret their meaning. Through understanding the nuances of assessment and grading practices among high school teachers, those making decisions based on teacher assigned grades will be more informed and able to make more appropriate decisions regarding policies and practices impacting students.

As part of your agreement to participate, I will be offering a drawing for one of four $25 VISA gift that can be used as desired for your classroom. To participate, you will simply need to submit your email address at the end of the survey so that I may contact you should you be selected in the drawing. Your answers will be kept completely

confidential. Please consider taking a few minutes to complete this survey by clicking the hyperlink below:

Proceed to the Informed Consent

I would be more than happy to respond to any questions or comments about this study. Please contact me at nmclean5@uncc.edu.

Thank you very much for helping in the study and the completion of the requirements for my doctoral degree.

Sincerely,

Nathan E. McLean
Doctoral Candidate

Dr. Claudia Flowers
Professor
cpflowers@uncc.edu

P.S. If for some reason I have made a mistake, and you are not a high school teacher currently teaching English, math, science, or social studies, or you would like to opt out of the study, please respond to the email above, and I will remove your email from future contacts. Again, thank you very much.

Appendix C

First Reminder Email

Department of Educational Leadership

9201 University City Boulevard, Charlotte, NC 28223-0001

Dear XX,

Last week I emailed you a questionnaire seeking information regarding the assessment and grading practices you implement in your classroom. If you have already completed the survey, please accept my most gracious thanks. If you have not done so, please consider taking 6 to 8 minutes to provide your information for my dissertation research. I am especially thankful for your help, and without your feedback, the true meaning of grades and how they vary between the content areas would remain unclear. As a reminder, I am offering four $25 VISA gift cards for those who complete the survey and provide their email address so that I may contact the individual who was drawn. To complete this survey, click the hyperlink below:

Proceed to the Informed Consent

Thank you for your participation,

Nathan E. McLean
Doctoral Candidate
Nmclean5@uncc.edu

Dr. Claudia Flowers
Professor
cpflowers@uncc.edu

Appendix D

Second Reminder Email

Department of Educational Leadership

9201 University City Boulevard, Charlotte, NC 28223-0001

Dear XX,

About two weeks ago, I sent you a survey asking about the assessment and grading practices you use in your classroom. If you have already completed the survey, I give you many thanks. I also apologize for sending you another email and this is the last time I will contact you, but because identifying information was not collected, I am unable to remove those who have previously responded. If you have not done so, please do so today, because time is running out. As a reminder, I am offering four $25 VISA gift cards for those who complete the survey and provide their email address so that I may contact the winners.

The assessment and grading practices of other teachers within your school, district, and surrounding areas have been submitted. The responses provided very interesting preliminary data, and I believe the data is going to be very useful.

I am writing again to emphasize the importance that the information you provide will help to get accurate results. Although the questionnaires were only sent to a few districts in North Carolina, it is only through hearing from nearly everyone that the data can be generalizable to your district and the surrounding area.

Regarding survey procedures, there will be no identifying information collected, so there is no way to connect your responses with you. Protecting the confidentiality of all

participants is very important to me. As a reminder, participation in this study is completely voluntary.

I hope that you will take a few minutes to complete the survey. If you have already submitted your questionnaire and/or wish to opt out of future contacts, please email me at nmclean5@uncc.edu.

CLICK HERE TO PROCEED TO THE INFORMED CONSENT

Sincerely,

Nathan E. McLean
Doctoral Candidate

Dr. Claudia Flowers
Professor
cpflowers@uncc.edu

P.S. If you have any questions or comments, please feel free to contact me at nmclean5@uncc.edu. Again, thank you very much.

Appendix E

Frequency by Percentage, Mean, Median, and Standard Deviation of Grading Factors Used to Determine Final Grades for All Participants

Item	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	*Mean*	*Median*	*SD*
Grading factors									
Percent Correct (Item 1)	2.0	2.4	14.9	28.2	31.9	20.6	4.47	5.00	1.16
Objective Mastery (Item 2)	0.8	3.2	14.1	31.9	35.9	14.1	4.41	4.50	1.05
Academic Performance (Item 3)	0.8	3.2	19.0	36.3	32.7	8.1	4.21	4.00	1.00
Zeros (Item 4)	4.0	8.5	15.3	19.4	19.4	33.5	4.42	5.00	1.49
Ability (Item 5)	23.8	23.4	29.4	16.5	5.6	1.2	2.60	3.00	1.24
Effort (Item 6)	8.1	10.9	36.3	23.4	16.5	4.8	3.44	3.00	1.23
Homework Quality (Item 7)	11.7	21.8	25.4	19.8	17.3	4.0	3.21	3.00	1.38
Attention or Participation (Item 8)	13.7	17.7	28.6	22.2	15.3	2.4	3.15	3.00	1.33
Homework Completion (Item 9)	10.9	16.5	29.4	22.6	16.5	4.0	3.29	3.00	1.33
Borderline grades (Item 10)	11.3	19.0	33.5	14.5	16.1	5.6	3.22	3.00	1.38
Improvement (Item 11)	14.9	16.9	31.0	20.6	13.7	2.8	3.10	3.00	1.33
Work habit (Item 12)	28.2	29.4	22.2	15.3	3.6	1.2	2.40	2.00	1.22
Extra credit academic (Item 13)	40.7	28.2	21.4	5.6	3.2	0.8	2.05	2.00	1.13
Performance of current students (Item 14)	48.4	27.0	16.1	4.8	2.8.	0.8	1.89	2.00	1.10
Behavior (Item 15)	59.3	20.6	14.9	2.4	2.0	0.8	1.70	1.00	1.03

Appendix E (Continued)

Item	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Extra credit non-academic (Item 16)	83.1	12.9	3.6	0	0.4	0	1.22	1.00	.54
Letter policy (Item 17)	77.0	7.3	7.3	3.6	2.4	2.4	1.54	1.00	1.18
Performance of past students (Item 18)	71.4	13.3	10.1	2.0	3.2	0	1.52	1.00	.98
Grades other teachers assign (Item 19)	87.5	8.5	2.1	0.8	0.8	0	1.19	1.00	.59
The Final Grade									
Grades represent achievement only (Item 32)	0	0.8	8.9	34.7	46.9	8.9	4.54	5.00	.81
Administrators and teachers judge grades assigned (Item 33)	6.0	22.6	30.2	24.2	12.5	4.4	3.28	3.00	1.24
Grade inflation causing grad rate increase (Item 34)	1.6	5.6	14.5	24.6	29.8	23.8	4.47	5.00	1.25
Credit recovery encourages grade bumping (Item 35)	27.0	24.2	21.0	10.9	8.9	8.1	2.75	2.00	1.57
Minimum grading policy limits assignment of low grades (Item 36)	21.4	13.7	22.2	11.7	14.5	16.5	3.34	3.00	1.75

Note. Item numbers refer to the instrument, see Appendix A.

Appendix F

Frequency by Percentage, Mean, Median, and Standard Deviation of Grading Factors Used to Determine Final Grades by English Teachers

Item	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Grading factors									
Percent Correct (Item 1)	1.5	0	16.4	23.9	32.8	25.4	4.63	5.00	1.13
Objective Mastery (Item 2)	1.5	3.0	14.9	31.3	35.8	13.4	4.37	4.00	1.09
Academic Performance (Item 3)	1.5	3.0	25.4	34.3	29.9	6.0	4.06	4.00	1.03
Zeros (Item 4)	6.0	6.0	20.9	16.4	17.9	32.8	4.33	5.00	1.55
Ability (Item 5)	22.4	17.9	29.9	23.9	3.0	3.0	2.76	3.00	1.29
Effort (Item 6)	9.0	4.5	32.8	22.4	28.4	3.0	3.66	4.00	1.27
Homework Quality (Item 7)	4.5	16.4	29.9	20.9	20.9	7.5	3.60	3.00	1.32
Attention or Participation (Item 8)	9.0	13.4	31.4	25.4	17.9	3.0	3.39	3.00	1.27
Homework Completion (Item 9)	10.4	17.9	32.8	17.9	13.4	7.5	3.28	3.00	1.39
Borderline grades (Item 10)	6.0	19.4	31.3	19.4	14.9	9.0	3.45	3.00	1.36
Improvement (Item 11)	9.0	13.4	37.3	20.9	17.9	1.5	3.30	3.00	1.22
Work habit (Item 12)	28.4	25.4	22.4	17.9	4.5	1.5	2.49	2.00	1.28
Extra credit academic (Item 13)	44.8	31.3	17.9	3.0	3.0	0	1.88	2.00	1.01
Performance of current students (Item 14)	53.7	23.9	13.4	6.0	3.0	0	1.81	1.00	1.08
Behavior (Item 15)	64.2	14.9	17.9	3.0	0	0	1.60	1.00	.89

Appendix F (Continued)

Item	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Extra credit non-academic (Item 16)	83.6	10.4	6.0	0	0	0	1.22	1.00	.546
Letter policy (Item 17)	80.6	7.5	4.5	4.5	1.5	1.5	1.43	1.00	1.05
Performance of past students (Item 18)	74.6	9.0	11.9	1.5	3.0	0	1.49	1.00	.98
Grades other teachers assign (Item 19)	86.6	6.0	6.0	0	1.5	0	1.24	1.00	.70
The Final Grade									
Grades represent achievement only (Item 32)	0	0	11.9	32.8	44.8	10.4	4.54	5.00	.84
Administrators and teachers judge grades assigned (Item 33)	7.5	20.9	26.9	31.3	9.0	4.5	3.27	3.00	1.24
Grade inflation causing grad rate increase (Item 34)	1.5	7.5	13.4	22.4	28.4	26.9	4.49	5.00	1.31
Credit recovery encourages grade bumping (Item 35)	31.3	20.9	16.4	11.9	11.9	7.4	2.75	2.00	1.65
Minimum grading policy limits assignment of low grades (Item 36)	19.4	19.4	20.9	10.4	16.4	16.4	3.37	3.00	1.76

Note. Item numbers refer to the instrument, see Appendix A.

Appendix G

Frequency by Percentage, Mean, Median, and Standard Deviation of Grading Factors Used to Determine Final Grades by Math Teachers

Item	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Grading factors									
Percent Correct (Item 1)	1.6	3.1	12.5	37.5	32.8	12.5	4.34	4.00	1.06
Objective Mastery (Item 2)	1.6	4.7	14.1	42.2	29.7	7.8	4.17	4.00	1.03
Academic Performance (Item 3)	0	0	17.2	40.6	37.5	4.7	4.30	4.00	.81
Zeros (Item 4)	1.6	12.5	15.6	28.1	20.3	21.9	4.19	4.00	1.37
Ability (Item 5)	18.8	31.3	31.3	17.2	1.6	0	2.52	2.50	1.04
Effort (Item 6)	6.3	9.4	48.4	20.3	12.5	3.1	3.33	3.00	1.11
Homework Quality (Item 7)	21.9	35.9	21.9	14.1	6.3	0	2.47	2.00	1.17
Attention or Participation (Item 8)	14.1	21.9	34.3	20.3	6.3	3.1	2.92	3.00	1.24
Homework Completion (Item 9)	10.9	12.5	21.9	32.8	17.2	4.7	3.47	4.00	1.35
Borderline grades (Item 10)	15.6	21.9	31.3	4.7	20.3	6.3	3.11	3.00	1.51
Improvement (Item 11)	15.6	12.5	39.1	20.3	10.9	1.6	3.03	3.00	1.25
Work habit (Item 12)	37.5	31.3	23.4	6.3	1.6	0	2.03	2.00	1.01
Extra credit academic (Item 13)	39.1	28.1	21.9	7.8	3.1	0	2.08	2.00	1.10
Performance of current students (Item 14)	51.6	26.6	15.6	3.1	1.6	1.6	1.81	1.00	1.08
Behavior (Item 15)	57.8	20.3	18.8	1.6	1.6	0	1.69	1.00	.94

Appendix G (Continued)

Item	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Extra credit non-academic (Item 16)	85.9	12.5	1.6	0	0	0	1.16	1.00	.41
Letter policy (Item 17)	87.5	4.7	3.1	3.1	0	1.6	1.28	1.00	.88
Performance of past students (Item 18)	78.1	12.5	7.8	0	1.6	0	1.34	1.00	.76
Grades other teachers assign (Item 19)	87.5	10.9	1.6	0	0	0	1.14	1.00	.39
The Final Grade									
Grades represent achievement only (Item 32)	0	1.6	6.3	35.9	50.0	6.3	4.53	5.00	.78
Administrators and teachers judge grades assigned (Item 33)	4.7	26.6	40.6	10.9	14.1	3.1	3.13	3.00	1.19
Grade inflation causing grad rate increase (Item 34)	0	6.3	12.5	23.4	29.7	28.1	4.61	5.00	1.20
Credit recovery encourages grade bumping (Item 35)	21.9	31.3	18.8	12.5	6.3	9.4	2.78	2.00	1.55
Minimum grading policy limits assignment of low grades (Item 36)	20.3	12.5	21.9	14.1	6.3	25.0	3.48	3.00	1.84

Note. Item numbers refer to the instrument, see Appendix A.

Appendix H

Frequency by Percentage, Mean, Median, and Standard Deviation of Grading Factors Used to Determine Final Grades by Science Teachers

Item	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Grading factors									
Percent Correct (Item 1)	3.1	1.5	13.8	24.6	32.3	24.6	4.55	5.00	1.23
Objective Mastery (Item 2)	0	0	10.8	26.2	43.1	20.0	4.72	5.00	.91
Academic Performance (Item 3)	1.5	3.1	16.9	29.2	38.5	10.8	4.32	4.00	1.08
Zeros (Item 4)	6.2	10.8	10.8	12.3	21.5	38.5	4.48	5.00	1.63
Ability (Item 5)	27.7	23.1	32.3	9.2	6.2	1.5	2.48	2.00	1.25
Effort (Item 6)	7.7	15.4	40.0	221.5	10.8	4.6	3.26	3.00	1.22
Homework Quality (Item 7)	12.3	20.0	23.1	21.5	20.0	3.1	3.26	3.00	1.40
Attention or Participation (Item 8)	16.9	20.0	26.2	20.0	15.4	1.5	3.02	3.00	1.36
Homework Completion (Item 9)	9.2	18.5	32.3	24.6	13.8	1.5	3.20	3.00	1.21
Borderline grades (Item 10)	13.8	18.5	38.5	15.4	10.8	3.1	3.00	3.00	1.28
Improvement (Item 11)	15.4	23.1	27.7	21.5	7.7	4.6	2.97	3.00	1.35
Work habit (Item 12)	20.0	40.0	18.5	15.4	4.6	1.5	2.49	2.00	1.20
Extra credit academic (Item 13)	40.0	32.3	21.5	4.6	0	1.5	1.97	2.00	1.03
Performance of current students (Item 14)	49.2	21.5	21.5	6.2	0	1.5	1.91	2.00	1.10
Behavior (Item 15)	60.0	27.7	6.2	3.1	1.5	1.5	1.63	1.00	1.02

Appendix H (Continued)

Item	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Extra credit non-academic (Item 16)	83.1	13.8	3.1	0	0	0	1.20	1.00	.47
Letter policy (Item 17)	76.8	6.2	7.7	1.5	4.6	3.1	1.60	1.00	1.30
Performance of past students (Item 18)	67.7	20.0	7.7	1.5	3.1	0	1.52	1.00	.94
Grades other teachers assign (Item 19)	89.2	9.2	1.5	0	0	0	1.12	1.00	.38
The Final Grade									
Grades represent achievement only (Item 32)	0	0	7.7	40.0	44.6	7.7	4.52	5.00	.75
Administrators and teachers judge grades assigned (Item 33)	3.1	24.6	24.6	33.8	7.7	6.2	3.37	3.00	1.21
Grade inflation causing grad rate increase (Item 34)	3.1	4.6	15.4	27.7	30.8	18.5	4.34	4.00	1.25
Credit recovery encourages grade bumping (Item 35)	30.8	26.2	26.2	4.6	7.7	4.6	2.46	2.00	1.42
Minimum grading policy limits assignment of low grades (Item 36)	21.5	20.0	18.5	15.4	20.0	4.6	3.06	3.00	1.57

Note. Item numbers refer to the instrument, see Appendix A.

Appendix I

Frequency by Percentage, Mean, Median, and Standard Deviation of Grading Factors Used to Determine Final Grades by Social Studies Teachers

Item	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Grading factors									
Percent Correct (Item 1)	1.9	5.8	17.3	26.9	28.8	19.2	4.33	4.00	1.25
Objective Mastery (Item 2)	5.8	17.3	26.9	34.6	15.4	15.4	4.37	4.50	1.12
Academic Performance (Item 3)	0	7.7	15.4	42.3	23.1	11.5	4.15	4.00	1.07
Zeros (Item 4)	1.9	3.8	13.5	21.2	17.3	42.3	4.75	5.00	1.34
Ability (Item 5)	26.9	21.2	23.1	15.4	13.5	0	2.67	3.00	1.38
Effort (Item 6)	9.6	15.4	21.2	30.8	13.5	9.6	3.52	4.00	1.42
Homework Quality (Item 7)	7.7	13.5	26.9	23.1	23.1	5.8	3.58	4.00	1.35
Attention or Participation (Item 8)	15.4	15.4	21.2	23.1	23.1	1.9	3.29	3.00	1.43
Homework Completion (Item 9)	13.5	17.3	30.8	13.5	23.1	1.9	3.21	3.00	1.39
Borderline grades (Item 10)	9.6	15.4	32.7	19.2	19.2	3.8	3.35	3.00	1.33
Improvement (Item 11)	21.2	19.2	17.3	19.2	19.2	3.8	3.08	3.00	1.55
Work habit (Item 12)	26.9	19.2	25.0	23.1	3.8	19.9	2.63	3.00	1.31
Extra credit academic (Item 13)	38.5	19.2	25.0	7.7	7.7	1.9	2.33	2.00	1.37
Performance of current students (Item 14)	36.5	38.5	13.5	3.8	7.7	0	2.08	2.00	1.17
Behavior (Item 15)	53.8	19.2	17.3	1.9	5.8	1.9	1.92	1.00	1.28

Appendix I (Continued)

Item	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Extra credit non-academic (Item 16)	78.8	15.4	3.8	0	1.9	0	1.31	1.00	.73
Letter policy (Item 17)	59.6	11.5	15.4	5.8	3.8	3.8	1.94	1.00	1.41
Performance of past students (Item 18)	63.5	11.5	13.5	5.8	5.8	0	1.79	1.00	1.23
Grades other teachers assign (Item 19)	86.5	7.7	0	3.8	1.9	0	1.27	1.00	.82
The Final Grade									
Grades represent achievement only (Item 32)	0	1.9	9.6	28.8	48.1	11.5	4.58	5.00	.89
Administrators and teachers judge grades assigned (Item 33)	9.6	17.3	28.8	19.2	21.2	3.8	3.37	3.00	1.36
Grade inflation causing grad rate increase (Item 34)	1.9	3.8	17.3	25.0	30.8	21.2	4.42	5.00	1.23
Credit recovery encourages grade bumping (Item 35)	23.1	17.3	23.1	15.4	9.6	11.5	3.06	3.00	1.65
Minimum grading policy limits assignment of low grades (Item 36)	25.0	3.8	28.8	5.8	15.4	21.2	3.46	3.00	1.86

Note. Item numbers refer to the instrument, see Appendix A.

Appendix J

Frequency by Percentage, Mean, Median, and Standard Deviation of Assessments Used by All Core Content Areas

Type	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Formative (Item 20)	11.3	20.2	29.0	26.6	10.9	2.0	3.12	3.00	1.24
Summative (Item 21)	0.8	1.2	8.9	35.5	46.0	7.7	4.48	5.00	.87
Designed by self (Item 22)	0.4	2.0	14.1	27.0	45.2	11.3	4.48	5.00	.97
Objective (Item 23)	0.8	6.9	19.4	26.6	40.3	6.0	4.17	4.00	1.08
Essay (Item 24)	13.7	17.3	20.6	25.8	19.8	2.8	3.29	3.00	1.40
Performance (Item 25)	15.7	23.8	31.5	18.5	9.3	1.2	2.85	3.00	1.23
Project individual (Item 26)	6.0	19.8	35.9	20.6	15.7	2.0	3.26	3.00	1.18
Major exams (Item 27)	0.8	12.5	30.6	31.5	16.5	8.1	3.75	4.00	1.14
Authentic (Item 28)	14.9	33.1	33.5	12.5	4.4	1.6	2.63	3.00	1.12
Project team (Item 29)	8.9	24.2	41.9	19.0	4.4	1.6	2.91	3.00	1.05
Publisher or district designed (Item 30)	31.0	32.7	21.8	10.9	3.6	0	2.23	2.00	1.11
Test made by others (Item 31)	14.5	23.4	35.9	13.7	11.3	1.2	2.88	3.00	1.23

Note. Item numbers refer to the instrument, see Appendix A.

Appendix K

Frequency by Percentage, Mean, Median, and Standard Deviation of Assessments Used by English Teachers

Type	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Formative (Item 20)	10.4	19.4	25.4	22.4	19.4	3.0	3.30	3.00	1.35
Summative (Item 21)	0	1.5	7.5	41.8	41.8	5	4.46	4.00	.80
Designed by self (Item 22)	0	0	14.9	25.4	46.3	13.4	4.58	5.00	.91
Objective (Item 23)	7.5	7.5	26.9	28.4	31.3	6.0	4.01	4.00	1.07
Essay (Item 24)	3.0	3.0	16.4	37.3	35.8	4.5	4.13	4.00	1.04
Performance (Item 25)	4.5	3.0	31.3	34.3	23.9	3.0	3.79	4.00	1.08
Project individual (Item 26)	0	9.0	26.9	31.3	26.9	6.0	3.94	4.00	1.07
Major exams (Item 27)	0	20.9	37.3	26.9	10.4	4.5	3.40	3.00	1.07
Authentic (Item 28)	17.9	29.9	28.4	16.4	6.0	1.5	2.67	3.00	1.21
Project team (Item 29)	7.5	19.4	40.3	22.4	7.5	3.0	3.12	3.00	1.14
Publisher or district designed (Item 30)	29.9	35.8	22.4	9.0	3.0	0	2.19	2.00	1.06
Test made by others (Item 31)	13.4	26.9	40.3	10.4	7.5	1.5	2.76	3.00	1.14

Note. Item numbers refer to the instrument, see Appendix A.

Appendix L

Frequency by Percentage, Mean, Median, and Standard Deviation of Assessments Used by Math Teachers

Type	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Formative (Item 20)	17.2	20.3	37.5	17.2	6.3	1.6	2.80	3.00	1.20
Summative (Item 21)	1.6	0	7.8	40.6	43.8	6.3	4.44	4.50	.85
Designed by self (Item 22)	0	1.6	6.3	32.8	48.4	10.9	4.61	5.00	.83
Objective (Item 23)	3.1	15.6	23.4	25.0	29.7	3.1	3.72	4.00	1.23
Essay (Item 24)	34.4	25.0	12.5	14.1	9.4	4.7	2.53	2.00	1.54
Performance (Item 25)	34.4	46.9	14.1	4.7	0	0	1.89	2.00	.82
Project individual (Item 26)	15.6	31.3	39.1	9.4	4.7	0	2.56	3.00	1.02
Major exams (Item 27)	1.6	12.5	28.1	37.5	14.1	6.3	3.69	4.00	1.11
Authentic (Item 28)	20.3	42.4	26.6	6.3	4.7	0	2.33	2.00	1.02
Project team (Item 29)	12.5	37.5	42.4	4.7	3.1	0	2.48	2.50	.89
Publisher or district designed (Item 30)	40.6	26.6	14.1	17.2	1.6	0	2.13	2.00	1.18
Test made by others (Item 31)	10.9	15.6	37.5	21.9	12.5	1.6	3.14	3.00	1.21

Note. Item numbers refer to the instrument, see Appendix A.

Appendix M

Frequency by Percentage, Mean, Median, and Standard Deviation of Assessments Used by Science Teachers

Type	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Formative (Item 20)	10.8	23.1	21.5	35.4	6.2	3.1	3.12	3.00	1.24
Summative (Item 21)	1.5	1.5	9.2	29.2	46.2	12.3	4.54	5.00	.99
Designed by self (Item 22)	1.5	4.6	21.5	26.2	36.9	9.2	4.20	4.00	1.12
Objective (Item 23)	0	1.5	13.8	24.6	49.2	10.8	4.54	5.00	.92
Essay (Item 24)	15.4	26.2	27.7	21.5	9.2	0	2.83	3.00	1.21
Performance (Item 25)	18.5	23.1	38.5	12.3	6.2	1.5	2.69	3.00	1.19
Project individual (Item 26)	6.2	20.0	43.1	12.3	16.9	1.5	3.18	3.00	1.17
Major exams (Item 27)	0	6.2	33.8	26.2	20.0	13.8	4.02	4.00	1.17
Authentic (Item 28)	6.2	26.2	41.5	18.5	4.6	3.1	2.98	3.00	1.08
Project team (Item 29)	9.2	20.0	41.5	24.6	3.1	1.5	2.97	3.00	1.05
Publisher or district designed (Item 30)	26.2	35.4	26.2	7.7	4.6	0	2.29	2.00	1.09
Test made by others (Item 31)	18.5	20.0	32.3	13.8	13.8	1.5	2.89	3.00	1.34

Note. Item numbers refer to the instrument, see Appendix A.

Appendix N

Frequency by Percentage, Mean, Median, and Standard Deviation of Assessments Used by Social Studies Teachers

Type	1 Not at all	2 Very Little	3 Some	4 Quite a bit	5 Extensively	6 Completely	Mean	Median	SD
Formative (Item 20)	5.8	17.3	32.7	32.7	11.5	0	3.27	3.00	1.07
Summative (Item 21)	0	1.9	11.5	28.8	53.8	3.8	4.46	5.00	.83
Designed by self (Item 22)	0	1.9	13.5	23.1	50.0	11.5	4.56	5.00	.94
Objective (Item 23)	0	1.9	11.5	28.8	53.8	3.8	4.46	5.00	.83
Essay (Item 24)	0	15.4	26.9	30.8	25.0	1.9	3.71	4.00	1.07
Performance (Item 25)	3.8	23.1	44.2	23.1	5.8	0	3.04	3.00	.93
Project individual (Item 26)	1.9	19.2	34.6	30.8	13.5	0	3.35	3.00	1.01
Major exams (Item 27)	1.9	9.6	21.2	36.5	23.1	7.7	3.92	4.00	1.15
Authentic (Item 28)	15.4	34.6	38.5	7.7	1.9	1.9	2.52	2.50	1.04
Project team (Item 29)	5.8	19.2	44.2	25.0	3.8	1.9	3.08	3.00	1.01
Publisher or district designed (Item 30)	26.9	32.7	25.0	9.6	5.8	0	2.35	2.00	1.15
Test made by others (Item 31)	15.4	32.7	32.7	7.7	11.5	0	2.67	3.00	1.18

Note. Item numbers refer to the instrument, see Appendix A.