

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Educational Administration: Theses, Dissertations,
and Student Research

Educational Administration, Department of


11-2010

7th-12th Grade English/language Arts Teachers and Their Classroom Grading Practices: Investigating the Use of Standards-based Grading in Nebraska's Rural Classrooms

Summer E. Stephens

UNL, summerelizabethstephens@gmail.com

Follow this and additional works at: <http://digitalcommons.unl.edu/cehsedaddiss>

 Part of the [Curriculum and Instruction Commons](#), [Educational Administration and Supervision Commons](#), and the [Other Teacher Education and Professional Development Commons](#)

Stephens, Summer E., "7th-12th Grade English/language Arts Teachers and Their Classroom Grading Practices: Investigating the Use of Standards-based Grading in Nebraska's Rural Classrooms" (2010). *Educational Administration: Theses, Dissertations, and Student Research*. 44.

<http://digitalcommons.unl.edu/cehsedaddiss/44>

This Article is brought to you for free and open access by the Educational Administration, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Educational Administration: Theses, Dissertations, and Student Research by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

7th-12th Grade English/language Arts Teachers and Their Classroom Grading Practices:
Investigating the Use of Standards-based Grading in Nebraska's Rural Classrooms

by

Summer Elizabeth Stephens

A DISSERTATION

Presented to the Faculty of
The Graduate College at the University of Nebraska
In Partial Fulfillment of Requirements
For the Degree of Doctor of Education

Major: Educational Administration

Under the Supervision of Professor Jody Isernhagen

Lincoln, Nebraska

December 2010

7th-12th Grade English/language Arts Teachers and Their Classroom Grading Practices:
Investigating the Use of Standards-based Grading in Nebraska's Rural Classrooms

Summer Elizabeth Stephens, Ed.D.

University of Nebraska, 2010

Adviser: Jody Isernhagen

A problem exists in grading practices accurately measuring student achievement. Both students' academic achievements and nonacademic factors, such as effort, homework completion, and behaviors continue to factor into grades. This combination can lead to inaccurate representation of true academic ability, rendering a grade useless.

While assessment and grading practices continue to be a conversation in most education circles, the implementation of standards-based grading practices in high school English/Language Arts classrooms varies widely. Past studies have shown that grades have provided feedback and been used to motivate and rank students. Future research was needed to address the following research question: How and to what degree are rural 7th-12th grade English/language arts teachers in Nebraska using standards-based grading practices in their classrooms? Four sub-questions focused on teachers' use of learning standards, assessment practices, markers of academic achievement and learner engagement in grading.

The study sample of 636 people included 7th-12th grade English/language arts teachers from Nebraska's Class III rural schools. A quantitative survey using a five-point Likert scale was designed to capture demographic data and the perceptions and

assessment and grading practices of rural 7th-12th grade English/language arts teachers in Nebraska.

It appears some components of standards-based grading are being utilized more than others. Rural 7th-12th grade English/language arts teachers in Nebraska indicated frequent use of standards for their course objectives, although they reported less frequent use of standards when reporting student grades. In addition, a number of participants reported including both formative and summative assessment results in student grades, while responses indicated fewer teachers used zeros and averaging student scores to achieve a final mark. Teachers with various levels of assessment training and educational backgrounds reported using effort as a grading criterion, but gave attendance and behavior less weight when calculating student grades. Finally, the inclusion of students in assessment and grading practices was focused heavily on sharing exemplars with students and not on students actually monitoring their own progress.

Author's Acknowledgements

I am greatly appreciative to the instructors and staff at the University of Nebraska-Lincoln who have been a part of my education over the last seven years, challenging and pushing me to ask more questions and dig deeper to enhance my own learning. Thank you to the members of my doctoral committee as you have been a huge part of my success in this program.

I am also grateful to all of the individuals who took part in my survey, yielding such a great response rate. This speaks volumes about level of professionalism in our Nebraska teaching ranks.

My heartfelt gratitude also goes out to Dr. Jody Isernhagen, who is responsible for helping me successfully complete my dissertation. Her commitment, encouragement, guidance and support helped me greatly in the understanding and writing of the dissertation and in becoming the educator I am today.

Most importantly, my greatest appreciation goes out to my family who have supported me and guided me throughout my education. Thank you to my parents, Wendy and Steve, for always insisting that I be the change I want to see in the world! An immeasurable thank you to my sister, Susan, for being my sounding board throughout my career and for insisting that we think outside of the box to make education better! To my husband Bert, and my three wonderful girls, Elizabeth, Madeline, and Poppy, for helping me make this dream possible and for giving me the motivation to make learning life's most important skill!

Table of Contents

Chapter 1—Introduction	1
Background and Context of the Study	2
Problem Statement	3
Purpose Statement.....	4
Research Objective and Questions.....	5
Research Methodology	5
Research Design.....	5
Population	6
Survey Instrument.....	6
Variables	7
Definition of Terms.....	7
Assumptions.....	10
Delimitations.....	10
Limitations	10
Significance of the Study	11
Summary.....	11
Chapter 2—The Literature Review.....	13
History of Grading	14
Professional Growth and Standards-based Grading	17
Influence of STARS.....	18
Components that Contribute to Standards-based Grading.....	20
Component #1—Learning Standards.....	21

Component #2—Assessment	22
Component #3—Markers of Academic Achievement.....	26
Component #4—Engagement of the Learner	28
Summary	30
Chapter 3—Methodology	31
Introduction.....	31
Research Design.....	31
Research Objective and Questions.....	32
Population	32
Survey Instrument.....	33
Survey Procedures	35
Pilot Survey Procedures	35
Study Survey Procedures	35
Variables	36
Data Analysis	36
Validity of Study.....	38
Reliability of Study.....	39
Protection of Participants’ Right.....	39
Summary	39
Chapter 4—Results	40
Introduction.....	40
Presentation of Descriptive Characteristics of Participants	40
Results by Research Question.....	46

Research Sub-Question #1	47
Item 10	48
Item 11	49
Item 12	50
Item 14	51
Research Sub-Question #2	51
Item 15	53
Item 16	54
Items 17, 18, 19.....	55
Items 20-21	56
Research Sub-Question #3	58
Item 25	58
Items 26-28	60
Items 29-30	62
Item 32	63
Research Sub-Question #4	64
Item 13	65
Item 22	66
Item 23-24.....	67
Item 31	68
Summary	68
Chapter 5—Conclusions, Recommendations, and Future Research	70
Introduction.....	70

Discussion	71
Research Sub-Question #1	71
Research Sub-Question #2	72
Research Sub-Question #3	75
Research Sub-Question #4	76
Recommendations	77
Recommendation One	77
Recommendation Two	78
Recommendation Three	79
Future Research	80
Summary	81
References	83
Appendices	88

List of Tables

Table 1	Demographic Match to Survey Items	34
Table 2	Standards-based Grading Big Ideas Match to Survey Items	34
Table 3	Gender of Participants.....	41
Table 4	Ages of Participants	42
Table 5	Education Levels of Participants	42
Table 6	Years in Education of Participants.....	43
Table 7	Endorsements of Participants.....	44
Table 8	Assessment Professional Development of Participants	45
Table 9	Grades Taught by Participants	45
Table 10	Grade Configuration of Participants' Schools	46
Table 11	Survey Section II—Learning Standards Descriptive Statistics	47
Table 12	Survey Section III—Assessment Practice Descriptive Statistics.....	52
Table 13	Survey Section III—Markers of Academic Achievement Descriptive Statistics—Items 25-30	59
Table 14	Survey Section IV—Student Engagement Descriptive Statistics—Items 13, 22-24, 31	65

List of Appendices

Appendix A	Survey Participants Letter.....	88
Appendix B	Survey Instrument.....	90

Chapter 1

Introduction

Picture these three students and imagine what types of grades they each get in the same English class:

Sally always completes her homework, not always with the right answers, but complete and on time. She sits towards the front of the room, always “pays attention” to the teacher, volunteers when the teacher calls for it, and gets the benefit of the doubt on tests where she only performs middle of the road, because the teacher “believes she knows the right answer but just has a bit of test anxiety.”

Marcus turns in some homework some of the time—partially complete work with a portion of correct responses. He jokes around with others in class, participates in class discussions and projects, and receives average marks on his tests.

Niki rarely, if ever completes homework, sits in the back with her head down during class, scowls when “group work” is assigned, misses class occasionally, but always get 90% or more of the work correct on tests and projects the teacher assigns.

Which student possesses higher academic achievement for the skills measured in that classroom?

In traditional classrooms, the students from the above vignette would earn grades of about an A to A-, C, and D or F respectively based on teacher grading practices which take into account effort, daily work, and behavior. However, when measured in a standards-based classroom that uses true academic achievement with a similar grade scale, those grades would be very different. In that classroom, academic learning standards would be clearly defined and reported, and the students’ grades would appear, possibly as: Sally would earn a B or C, Marcus, a B or C, and Niki would earn an A or B. These are very different pictures of achievement. Consider what Grand Island Public Schools’ Superintendent Stephen Joel ponders: “What is an A? What is a B? Let’s

make sure when we do grades, grades say something” (AP, 2006). Getting closer to creating a truer picture of students’ skills and strengths, the latter would lend itself to that.

Background and Context of the Study

Nebraska is a state that has paved the way for formative classroom assessment and growth models. The state challenged the status quo in terms of its assessment practices. Experts and budding scholars alike have studied Nebraska as the model for the ideal state assessment system. Until recently, local control provided school districts with the opportunity to report student achievement based on local curriculum and local instruction. So, what really takes place in Nebraska’s schools when reporting student achievement? Are state or local standards the basis for the student report cards? Based on the philosophy that Nebraska has embraced to utilize classroom based assessment to measure student achievement, standards should be the critical component used to determine classroom grades.

In a state where assessment literacy is a goal for all teachers, what is the level to which we hold teachers accountable? After schools in the state adopted standards as the basis for what is taught and developed teachers’ assessment literacy, the logical next step is to adapt an age old practice of grading to best reflect student academic achievement. Best practice in assessment would point to the use of standards-based grading as the pinnacle of a high functioning standards-based classroom.

To utilize a system rich in true student achievement data only to revert to a grading system that rewards and punishes and rank orders children doesn’t make sense. By utilizing a standards-based achievement system (one in which standards and clear criteria are evident to define achievement), educators are poised to present to the

stakeholders clear and definable achievement that truly measures what a student can and cannot do at any one point in time.

Not only did this study explore beliefs about grading—it also explored to what extent the grade given by the teacher is aligned with the marks received on the local criterion referenced assessments, both classroom assessments and those used for state reporting purposes.

History has shown that old is not always bad. Grading practices in American schools have changed tremendously over the last 250 years. Gone are the days of measuring learning and providing constructive feedback for improvement. Today, letters and numbers rule the roost in most public schools. Researchers such as Stiggins, Arter, Chappius, and Chappius (2004), O'Connor (2007), and Marzano (2000) have begun to highlight the impact that traditional grading practices have on masking accurate student achievement. Scholars are beginning to pinpoint several critical components that must be present in classrooms to truly measure students' academic achievement and learning, several of which include clear and measurable standards and accurate assessment. The researcher in this study seeks to find out how and to what degree rural teachers of 7th-12th grade English/language arts in Nebraska match the expectations of current researchers and are utilizing standards-based grading in their classrooms.

Problem Statement

A problem exists in grading practices as they relate to accurately measuring student achievement (Carr & Artman, 2002; Reeves, 2002). Both students' academic achievement as well as nonacademic factors, such as effort, homework completion, and behaviors contribute to the determination of grades in many classrooms. This

combination can lead to inaccurate markings resulting in a skewed perception of a student's true academic ability and can render a grade useless in determining academic aptitude in a given subject. "Grades are broken when they mix achievement and non-achievement elements" (O'Connor, 2007, p. 24).

While assessment and grading practices continue to be the topic of conversation in standards and assessment circles, the actual implementation of standards-based grading practices in high school English/Language Arts classrooms varies widely. Past studies have shown that grades have been used for a variety of purposes, including providing feedback, motivating students, ranking students, sorting students, and qualifying students for scholarships and college entry (Stiggins et al., 2004; Guskey & Bailey, 2001; Kohn, 1993; Reeves, 2004). Marzano (2000) argues that the primary and most important purpose of grades should be to provide information or feedback to students and parents, not to rank students. Knowing these various purposes of grades, what is not known is how and to what degree standards-based grading is being implemented in rural 7th-12th grade English/language arts classrooms in Nebraska.

Purpose Statement

The purpose of this quantitative survey study is to determine if rural 7th-12th grade English/Language Arts teachers in Nebraska are using standards-based grading in place of traditional grading procedures. More specifically, to which grading elements are teachers drawn? Underlying this purpose is the intent to investigate the knowledge base of teachers regarding their understanding of the interconnectedness of standards-based assessment, teaching and learning, and standards-based grading. The researcher has served as an English/language arts teacher in Nebraska rural schools, which serves as her

basis for interest in this research. In addition, the researcher has served as both a staff developer of an Educational Service Unit and currently serves as a district Curriculum and Assessment Director.

Research Objective and Questions

The objective of this quantitative survey study was to determine whether or not high school English/Language Arts teachers use standards-based grading in place of traditional grading practices. The overarching research question was: How and to what degree are rural 7th-12th grade English/language arts teachers in Nebraska using standards-based grading practices in their classrooms?

The sub-questions for this study were based on the elements necessary for standards-based grading to be in place in rural 7th-12th grade English/language arts classrooms in Nebraska.

1. What learning standards are included in grading by teachers and how do learning standards support standards-based grading?
2. What methods of assessment and re-assessment are included by teachers and how does assessment support standards-based grading?
3. What markers of academic achievement are included by teachers and how do these markers support standards-based grading?
4. What methods of learner engagement are used by teachers and how does learner engagement support standards-based grading?

Research Methodology

Research design. In quantitative research, the researcher determines what to study, asks specific, focused questions, collects numeric data from participants, analyzes

these numbers using statistics, and conducts the inquiry in an unbiased, objective manner (Creswell, 2005). This type of research emphasized the collection and analysis of information in the form of numbers, the collection of data that measure distinct attributes of individuals, and the procedures of comparing groups and/ or relating factors about individuals or groups in experiments, correlation studies, and surveys (Creswell, 2005). Quantitative research follows the pattern of identifying the research problem, using literature to build a basis for additional research need, followed by the specific research design of data collection, analysis of results, and description of the findings (Creswell, 2005).

Population

The population for this study consisted of 7th-12th grade English/language arts teachers from Nebraska's public schools. The sample of 636 people consisted of teachers from Nebraska's Class III rural schools who teach 7th-12th grade English/language arts courses. This sample was developed utilizing the Nebraska Department of Education online staff database that allows an individual to search for teachers with various characteristics from Nebraska's public schools, including those who teach in Class III schools. Since the database does not list a school's rural/non-rural classification, a list of schools that are considered non-rural for the 2009-2010 school year was identified using data from the United States Census Bureau.

Survey Instrument

A four-section survey was designed for the data collection for this research study. The first section of the instrument focused on the demographics of the participants. The remaining sections of the survey focus on the three big ideas of standards-based grading

obtained from the literature review, including: standards, assessment, and academic achievement markers. A fourth concept that permeates all of the sections is that of student engagement in standards-based grading. Items on the survey were rated using a five-point Likert- scale in order to capture the perceptions and practices of the assessment and grading practices of rural 7th-12th grade English/language arts teachers in Nebraska.

There were benefits to using survey methodology to collect data. First, generating the list of teachers from which to collect data was up-to-date with current teaching rosters available from the Nebraska Department of Education. Second, a paper sampling method cut down on the duplication of submissions and allowed the researcher to maintain a clear record of responses. Finally, the use of paper sampling provided a stronger return. Some drawbacks include an increased cost over utilizing a web-based method and possibly less-timely collection of the data.

Variables

The critical dependent variables in this study were the various components of standards-based grading utilized in the classroom. The independent variables that will create significance included the teacher's gender, teacher's age, the size of the school, the length of the teacher's educational career, the teacher's educational background, and the teacher's training in assessment literacy, the gateway for increasing usage of standards-based grading.

Definition of Terms

For the purposes of this study, the following definitions were utilized:

Assessment—Assessments are vehicles for gathering information about students' achievement or behavior (Marzano, 2000).

Criterion-referenced assessment—Criterion-referenced assessments are based on standards, objectives, or benchmarks as the reference points to determining student achievement (Wormeli, 2006).

Formative assessment—Formative assessments are also referred to as “Assessment for Learning” (Stiggins et al., 2004). Formative assessment is a planned process in which assessment-elicited evidence of students’ status is used by teachers to adjust their on-going instructional procedures or by students to adjust their current learning tactics.

- Formative assessment is a process, not a specific test
- Formative assessment is used by both teachers and students
- It takes place during instruction, not after
- The function of the feedback is to help teachers and students make adjustments that will improve students’ achievement (Popham, 2008).

Grading—Grading is the process of reviewing evidence of achievement and determining its value (Davies, 2000).

Marking—Marking is assigning a number, letter, or word to any single student assessment (O’Connor, 2002).

Rural—“rural” consists of all territory, population, and housing units located outside of Urban Areas or Urban Clusters that have core census block groups or blocks that have a population density of at least 1,000 people per square mile and surrounding census blocks that have an overall density of at least 500 people per square mile (U.S. Census Bureau, 2009).

Standards-based Grading—Standards-based grading is assigning grades based on the following criteria:

Learning Standards

- Based on learning targets or outcomes determined by various systems, such as classroom and district developed objectives, and state departments and/or national organization standards.

Assessment

- Based on the opportunities for reassessment and most recent information
- Derived from summative assessment marks

Markers of Academic Achievement

- Academic achievement separated from elements such as attendance, effort, participation, behavior, collaboration, etc.
- Based on individual achievement
- Inclusive of students as participants in their own learning (O’Connor, 2002).

Student (or learner) engagement—Student engagement is on-task behaviors, personal voice, and commitment to personal understanding all of which function as an engine for learning and development (Reeve, 2006).

Summative assessment—Summative assessment is also referred to as “Assessment of Learning” (Stiggins et al., 2004). Summative assessment is designed to provide information about a student’s achievement at the end of the unit of study (O’Connor, 2007).

Traditional grading—Grading practices that are typically norm-referenced, include attitude, effort, attendance, can reflect ambiguous, teacher-developed criteria, and often include averages of student work (Tucker & Coddling, 1998b).

Assumptions

The core assumptions made by the researcher are:

1. The participants are willing participants in the study.
2. The participants will provide honest data on the survey.

A threat to validity may be the various exposure that all teachers have had in regard to standards, assessment, and grading practice trainings. In addition, there is no way to measure the truthfulness of the responses provided by participants.

Delimitations

Delimitations help to define the parameters of the research study (Creswell, 2009). This study will be delimited to the 7th-12th grade English/language arts teachers in Nebraska Class III rural schools. In addition, only rural schools of Nebraska will be reviewed as the researcher has an interest in Nebraska schools and rural education.

Limitations

Creswell (2005) describes limitations as “potential weaknesses or problems with the study” as identified by the researcher (p. 198). The following limitations may be useful for future researchers to consider:

1. The findings of this survey will be based on teacher’s perception versus an outside evaluation.
2. Data collected may be limited by the participants’ personal philosophies and willingness to respond.

Significance of the Study

The importance of this study is multifaceted. The information obtained through the survey will shed light on the current grading practices and reflect how classroom-based assessments are currently being used to report student achievement. Beginning with the 2009-2010 school year, classroom criterion-referenced assessments will no longer serve as a method for reporting student achievement on English/language arts standards in Nebraska. Additionally, the conclusions of the study can begin to illuminate a direction for educators to follow. Student achievement and student grades should mirror one another. “A standards-based grading and reporting system has the potential to provide an honest assessment of the child’s performance and clearly delineate the progress of the student towards the established standards” (Olson, 2005, p. 99).

Nebraska’s School-based, Teacher-led Accountability and Reporting System, or STARS, lent itself to a standards-based grading system; however, changes in Nebraska legislation have threatened this movement. With this in mind, little research has been done to date to measure the degree that student grades in Nebraska schools mirror student achievement on locally developed criterion-referenced assessments.

Summary

This chapter focuses on the purpose for examining questions regarding standard-based grading practices in rural Nebraska’s 7th-12th grade English/language arts classrooms.

Chapter 2 describes the literature review base for the research questions within the study. An investigation of professional growth and its role in standards-based grading, the influence of STARS (School-Based, Teacher-Led Accountability and Reporting

System) on grading practices, and the components of standards-based grading are examined.

Chapter 3 outlines the use of quantitative survey design as the methodology for this study. The role of the researcher, the research design and methodology, and the survey instrument are described. Data collection and analysis procedures are also described in this chapter.

Chapter 2

The Literature Review

In recent years, the movement in education has been toward one of utilizing standards as the basis for measuring student achievement. This follows the outcome-based education movement of the 1980's. With many things in education come questions of best practice approaches to teaching and learning, and most specifically, accurate measurement of student achievement. Literature on the historical context of assessment and grading practices provides an important commentary on the process. Additionally, the researcher reviewed literature on both academic and nonacademic factors as they related to grading practices in a standards-based grading system.

In determining which literature would be reviewed for this study, the researcher relied on the use of various research databases, primary source texts, and reviews of dissertations that had a similar area of focus.

Considering the history of education and the impact of change, one must look closely at the elements of early education that are evident in today's approach to standards-based teaching and learning. Additionally, one must consider change theory and what drives and motivates people to change. When an institution like public education has remained virtually unchanged over the last one hundred years, it is important to consider why. Tyack and Tobin (1994) suggested that periodically, a few innovators or early adopters have challenged the traditional "grammar" (p. 454) of schooling, advocating for ungraded schools, innovative use of time, space and numbers, more flexibility, creating more relevance in subject matter, and creating more teacher collaboration. Many of these innovations have not lasted very long, as is true today.

Many initiatives have arisen to support the implementation of standards and assessment, two areas have seemingly remained stagnant in most schools: grading practices and report cards. Although No Child Left behind does not specifically call for reform in grading practices, it does call for educational reform that impacts the way student achievement is recorded and reported to the public (No Child Left Behind, 2002). Tyack and Tobin (1994) suggested that change has not been sustained for two reasons: reform has not been “required” enough and those earlier reformers cannot sustain the beatings that come along with the change. Guskey and Bailey (2001) suggested five different but interrelated developments in education that have brought increased attention to grading and reporting.

1. The growing emphasis on standards and performance assessments makes current reporting practices inadequate.
2. Parents and community members are demanding more and better information about student learning progress.
3. Advances in technology allow for more efficient reporting of detailed information on student learning.
4. Grading and reporting are recognized as one of educators’ most important responsibilities.
5. There is a growing awareness of the gap between our knowledge base and common practice in grading and reporting. (p. 11)

History of Grading

Along with the growth and change in content and performance standards, so to have changes occurred in the reporting of achievement. The ancient Greeks used assessments in their teaching, but these were not formal evaluations of student achievement. Their purpose was primarily formative. Examinations provided students the opportunity to demonstrate, usually orally, what they had learned. They also gave teachers a clear indication of what topics required additional work or instruction. This

feedback-laden approach was followed into the higher education systems, where grading practices first developed.

In the earliest days of education in the United States, teachers marked student achievement with performance feedback and comments, followed by the administration of examinations ending in degrees, rather than giving something equivalent to today's "grade." M.L. Smallwood researched these phenomena in 1935 in a study of grading practices in higher education (Brookhart, 2004). Faculty presented on-going evaluation of students, making determinations of a student's readiness to take the final examinations toward a degree. This was met with both anxiety by the student and concern for retention by the faculty. By 1775, various grading scales had begun to come into use (Brookhart, 2004). In its inception, Yale University began using a system that was a precursor to the current grading system most traditionally used throughout American education today—the four-point grading scale (Marzano, 2000). Since that time, most schools have used the 100-point scale and the A-F grading scheme (Marzano, 2000). Prior to that in the United States, grading and reporting were virtually unknown in schools. In primary and secondary schools, the teacher reported students' learning progress orally to parents, usually during a home visit.

In the late 1800's, however, formal evaluations of the work of students of all ages began. Teachers would simply write down the skills each student had mastered and those on which additional work was needed. This was done primarily for the students' benefit, since they were not permitted to move on to the next level until they demonstrated their mastery of the current one. It was also the earliest example of a narrative report card.

When compulsory attendance laws at the elementary level began, the number of students entering high schools also increased rapidly. Between 1870 and 1910, the number of public high schools in the United States increased from 500 to 10,000. As a result, subject area instruction became more specific. High school teachers began to employ percentages and other similar markings to certify students' accomplishments rather than written descriptions and narrative reports like the elementary schools were using (Marzano, 2000).

Schools then moved to scales and fewer and larger categories. They began to grade on the curve to "ensure a fairer distribution of grades among teachers and to bring into check the subjective nature of scoring" (Guskey & Bailey, 2001). Grading on the curve was believed to be appropriate at that time because it was well known that the distribution of students' intelligence test scores approximated a normal probability curve. Grading on the curve also relieved teachers of the difficult task of having to identify specific learning criteria (Guskey & Bailey, 2001). With this as the norm, changing the approach to grading has made the journey difficult.

Although a change obviously occurred, "The human tendency to resist change, especially to the institution, is quite strong. So when considering changing the way one grades, as suggested by researchers such as O'Connor and Stiggins, it can alter what people associate with real school" (Marzano, 2000, p. 2). Not only do teachers sometimes struggle with the philosophical shift, so do parents and community members who hold true to the idea "That is not the way I was graded, and I turned out okay."

Guskey (1996) emphasized that difficulties with changing grading practices has been mounting for decades. Grading experts agree that a number of problems plague

current traditional grading practice including teachers considering many factors (not just academic achievement) when assigning grades, teachers weighing assessments differently, and teachers, because of lack of assessment literacy or some other factor, misinterpreting and misrepresenting single scores on individual classroom assessments. Stiggins, Frisbie, and Griswold (1989) put forward three reasons for the disconnect between the recommendation to change and actual practice, including the fact that best practice may be a matter of opinion, that some recommended practices do not take some of the practical aspects of teaching into account, and that even the notion that teachers lack training or expertise in sound practices is difficult to acknowledge.

Professional Growth and Standards-based Grading

Considering Stiggins et al. (1989) reasons from above, one must consider the lack of teacher training and/or expertise in sound assessment and grading practices as a possible reason that standards-based grading has moved so slowly onto the scene.

Guskey and Bailey (2001) described four factors that impact how a teacher determines grades.

First and foremost, they refer to the policies and practices they experienced as students. They do what was done to them. They select from their teacher bank (from those that they had over time), what they believe work best and are most appropriate.” (p. 16)

The second important influence on teachers’ grading and reporting practices is their personal philosophies of teaching and learning. The third source from which teachers draw is state-district-building-department- or grade-level grading policies or lack thereof. Based on 2009 research collected by the Southeast Comprehensive Center (SECC), the SECC found that most state departments of education do not have uniform grading practices. Nine states (18%) currently have statewide uniform grading scales and

policies, while 39 states (78%) do not have statewide uniform grading scales and policies (Southeast Comprehensive Center, 2009). The final factor is what they learned in their undergraduate teacher preparation programs. Missing from the above factors is the factor of teacher training in assessment literacy and grading practices.

Influence of STARS

Keeping the history of grading and professional development in mind, it is critical to examine the power and impact that the Nebraska School-based Teacher-led Assessment and Reporting System (STARS) had upon grading practices. This will allow the researcher to see if the standards and accountability movement are transforming grading practices. School officials must consider to what degree they will improve and transform the measurement of student achievement. In order to do this, the classrooms of Nebraska must be examined to determine the degree to which teachers are using the most effective measurements of student achievement. Should Nebraska classrooms mirror the days of the institution of compulsory education where teachers began to use percentages and other markings in place of written feedback or should it mirror more of the “real school” that so many researchers hold as a vision today? It is important to understand which grading structure is going to encourage the mastery of academic standards and allow for the best assessment to measure achievement for the students within classrooms.

Isernhagen, Dappen, and Mills (2006) found evidence that Nebraska educators were supported through professional development efforts aimed at developing quality classroom assessments, using data and research based teaching strategies over the first five years of the School-based, Teacher-led Accountability and Reporting System in Nebraska (STARS). Teachers attended trainings, worked collaboratively with

Educational Service Unit (ESU) personnel, and participated at the grassroots level of standards and assessment implementation. This structure provided a solid backbone that could lead to the implementation of a new method of reporting student achievement.

However, missing from this context was the conversation of grading student performance and approaching classroom - grading practices differently. Researchers Isernhagen, Florendo, and Guerrero (2009) concluded that rural language arts teachers may not be using principles of sound grading practices as much as their other subject area counterparts are in urban districts. The use of an assessment system is directly linked to the curriculum (and standards) one teaches as well as the ultimate grading of student work. “Standards-based accountability systems are potentially powerful tools for improving student performance” (Olson, 2005, p. 19). Research conducted in 2009 in Nebraska showed a move in the direction to embrace the components of standards-based grading. This is best illustrated by a rural, female middle school language arts teacher about her grading philosophy, “I retest, retest, retest. I give them the higher grade. I don’t average because I want their grade to reflect what they know . . . I give them the grade that they earn” (Isernhagen et al., 2009, p. 78).

Olson (2005) stated that in a classroom focused on performance standards and the use local assessments, the data contained in the report card could be used for accountability “as it creates a truer picture of student achievement at the classroom, school or district level” (p. 19).

Standards-based teachers distinguish clearly between teaching activities through which students learn and practice, and summative assessments in which students “perform” and show what they know, understand and can do. They are clear about the purpose of every activity, and grades include only evidence from summative assessments. (O’Connor, 2007, p. 96)

One Nebraska educator declared,

My gut feeling is that I'd like to see our report card the way that we show the community and the patrons, the parents, that we're assessing their children. I'd like to see that become aligned with the assessments. We've got these assessments. They're criterion referenced. They're unique to our district and yet they fulfill the state standards. And then we have this report card that comes out of post-World War II . . . A, B, C, D, and so I guess we need some real training on how to do this. (Isernhagen et al., 2006, p. 30)

This recognition of needed change in reporting first relies on a change in the belief system regarding student failure. A Nebraska superintendent stated,

We have bought into the failure is not an option concept! We've done some things in terms of our grading policy, to enhance student learning. And we're working with our teachers to change a mindset about how students are graded in this district. In other words, I have told the teachers in this district that have given students zeroes for not turning in homework that is not an option. (Isernhagen & Mills, 2007, p. 91)

Components that Contribute to Standards-based Grading

Teachers in the 21st Century struggle to implement a standards-based system in classrooms. According to Tucker and Coddling (1998a), a standards-based education refers to the search for ways of thinking about and operating school systems to ensure that all students achieve well-defined standards of performance. This is a shift in thinking for many teachers. Traditional grading practices were considered by Haladyna (1999) to be steeped in a complicated activity that requires considerable planning and skill. Milton, Pollio, and Eison (1986) likened these methods to that of a definition of grading by Paul Dressel: "An inadequate report of an inaccurate judgment by a biased and variable judge of the extent to which a student has attained an undefined level of mastery of an unknown proportion of an indefinite material" (p. 23). Lack of common grading scales within buildings, districts, or even states can lead to this disconnect between actual performance and teacher judgment. It can be unclear what a particular

mark on a report card means from teacher to teacher, or school to school. According to the Education Commission of the States (ECS), implementation of statewide uniform grading scales has been found to produce several benefits such as comparability of student achievement across districts, more uniformity between high schools for states with merit scholarships, and the same uniformity for students who move between districts (Burke, 2005).

Throughout the literature, four specific components can be identified:

Component #1—Learning Standards; Component #2—Assessment; Component #3—Markers of Academic Achievement; and Component #4—Student Engagement. The next section of the literature review will cover these components in detail.

Component #1—Learning standards. O’Shea (2005) suggested several key ingredients to maintaining a standards-based classroom. Teachers in these classrooms have aligned their content and skills to the state standards and have identified performance levels on the given standards, thus establishing clear learning targets. Guskey (2001) described these targets as “learning goals or standards (that) should stipulate precisely what students should know and be able to do as a result of their learning experiences” (p. 20).

What is the intended learning? That one question should drive all planning and assessment in schools today. Label these learning statements, “content standards, benchmarks, grade level indicators, grade level expectations, essential learnings, learning outcomes, lesson objectives, learning intentions . . .” they all represent learning targets, or statements of intended learning. (Stiggins et al., 2004, p. 54)

DuFour, DuFour, and Eaker (2008) contended that these targets are “Essential knowledge, skills, and dispositions students must acquire” (p. 24).

In regards to these achievement targets, O’Connor (2007) suggested that the key

to success is to utilize “overall and specific performance standards with a limited number of levels, clearly described in the language of the appropriate achievement continuum” (p. 70). As part of that mission to use specific standards with only a few clearly defined levels of performance, one must develop exemplars, or models, of the expectations. Marzano (2000) suggested that grading practices be more reflective of student achievement toward district or state standards and benchmarks.

“Provide an understandable vision of the learning target. Teach students the concepts underpinning quality in your scoring guide . . . share strong and weak student work” (Stiggins et al., 2004, p. 241). Performance standards will be public from the beginning of instruction, depicting the characteristics of the kind of academic achievement being judged. “When such descriptions are accompanied by samples of student work depicting each level of proficiency, we lay a solid foundation for effective judgment of and communication about student achievement” (O’Connor, 2007, p. 65). “If grading and reporting do not relate grades back to standards, they are giving a mixed message. Our grading practices must reflect and illuminate those standards” (Busick, 2000, p. 73).

Component #2—Assessment. Along with clear targets, definitions, and descriptions of student learning comes the use of curriculum pacing guides to ensure critical standards are achieved as evidenced in formative assessment before summative assessments are administered (O’Shea, 2005).

In addition to the alignment of skills and content, teachers must then consider the assessment practices used. In traditional classroom settings, the grading aspect of assessment is overemphasized and the learning or improvement purpose of assessment is

underemphasized (Black & Wiliam, 1998b). It is with quality assessment practices that educators begin to develop their ability to use assessment for learning as well as conduct assessment of learning. Black and Wiliam (1998a) described it as encompassing all those activities done by the teacher and/or the students that provide information to be used as feedback to modify instruction. The way in which the teachers and students use the information determines the formative and summative nature of the assessments.

Stiggins et al. (2004) developed a model of Assessment for Learning and Assessment of Learning that guides the work of the standards-based classroom. The formative assessment is used as feedback for students and their learning and for teachers to monitor and/or adjust instruction.

Assessments for learning happen while learning is still underway. These are the assessments that we conduct throughout teaching and learning to diagnose student needs, plan our next steps in instruction, provide students with feedback they can use to improve the quality of their work, and help students see and feel in control of their journey to success. (Stiggins et al., 2004, p. 31)

DuFour et al. (2008) described this assessment as

a tool used to inform both the teacher and the students about the student's current level of achievement, to guide the teacher's instructional practice, to help the student understand what steps must be taken to further his or her learning, and to motivate the students to take those steps. (p. 202)

Teachers can use assessments to provide grades to students at the end of unit of study. "Assessments of learning are those assessments that happen after learning is supposed to have occurred to determine if it did" (Stiggins et al., 2004, p. 31).

Additionally, Stiggins et al. (2004) suggested that teachers are gathering evidence to determine a student's report card grade.

Standards-based teachers distinguish clearly between teaching activities through which students learn and practice, and summative assessments in which students 'perform' and show what they know, understand and can do. They are clear about

the purpose of every activity, and grades include only evidence from summative assessments.” (O’Connor, 2007, p. 96)

Educators must ask themselves how confident they are that the grades students get in the classroom are reflective of the types of assessments administered and also the grades that the students get accurately reflect the content standards and desired learning outcomes (O’Connor, 2007). With that in mind, however, traditional educational measurement courses teach that only objective, numerical scores should be recorded and used for reporting grades, while researchers and other experts encourage the use of multiple assessment measures, many of which cannot be quantified (Seeley, 1994).

When focusing in on assessments used in standards-based grading, several key ideas must be considered. One such element is the use of the zero. The use of the zero as a mark on a student record is inaccurate, ineffective, even counterproductive, and can even be a distortion of the student’s performance. The use of zero has several fundamental problems. First, the use of the zero denotes that the student has learned nothing. “Zero implies the total absence of learning. Missed tests, scores attained by cheating, or assignments not handed in do not offer data about level of learning” (Stiggins et al., 2004, p. 313). “Zeros give a numerical value to something that has never been assessed and that therefore has no basis in reality” (O’Connor, 2007, p. 26). This would in most cases render the grades ineffective as a communication tool (O’Connor, 2007, p. 86).

In addition to creating the impression that a student has not learned, the zero also does not build motivation. O’Connor (2007) stated that zeros “can actually harm student motivation, and for many students do not result in changes in behavior” (p. 26).

O’Connor goes on to relate that assigning a student a zero can have counterproductive

effects on student motivation. Guskey (2000) contended that, “No studies support low grades or marks as punishments. Instead of prompting great effort, low grades more often cause students to withdraw from learning” (p. 25). Wormeli (2006) asserted that a zero has an “undeserved and devastating influence, so much so that no matter what the student does, the grade distorts the final grade as a true indicator of mastery” (p. 137). This assertion is supported by O’Connor. “Zeros in the record render grades ineffective as communication” (O’Connor, 2007, p. 86).

Not only does the zero misrepresent the student’s learning and create a culture of marginalizing a student’s self-concept, the zero actually involves inappropriate mathematics (O’Connor, 2007, p. 26). O’Connor argued that zeros are “penalties (that) distort the achievement record the grade is intended to communicate” (p. 26).

Several researchers (Carr & Farr, 2000; O’Connor, 2002) suggested that the approach one uses in place of using a zero for missing work might become the impetus that some students need to move forward. In place of a zero, these researchers suggest an incomplete, with the intention of a student either completing the work or some other task that measures the learning of the same standard.

The use of zeros, along with utilizing the mean over other measures of central tendencies or professional judgment, creates a misrepresentation of achievement. “Averaging zeros with other scores to calculate a final grade skews the score and results in an inaccurate picture of student achievement” (Stiggins et al., 2004, p. 313).

Another consideration in line with the use of zeros is the use of the mean or average when calculating grades, “averaging falls far short of providing an accurate description of what students have learned. . . . If the purpose of grading and reporting is

to provide an accurate description of what students have learned, then averaging must be considered inadequate and inappropriate” (Guskey, 1996, p. 21). “The fix for grades broken (by using the mean) is to not use the mean as ‘the measure’ by considering other measures of central tendency, and to recognize that grading should not be merely a numerical, mechanical exercise” (O’Connor, 2007, p. 81). Reeves (2000) implored educators to “abandon the average, or arithmetic mean, as the predominant measurement of student achievement” (p. 10).

Instead, teachers should take grading on as an “exercise in professional judgment.” Grading “involves the collection and evaluation of evidence on students’ achievement or performance over a specified period of time” (Guskey & Bailey, 2001, p. 9). “If more recent information about student achievement shows a new level of attainment, thereby making previous evidence outdated, then the grade should be based on the newer evidence” (Stiggins et al, 2004, p. 311).

Component #3—Markers of academic achievement. In addition overcoming the misguided use of the zero and the averaging of marks, another element that must be used by the teacher in an effort to move toward a standards-based grading model is that of measuring academic performance separately from behavior and other non-academic elements.

“Grades are broken when they mix achievement and non-achievement elements” (O’Connor, 2007, p. 24). Traditionally, teachers consider many factors other than academic achievement when they assign grades. Research shows that many teachers often use factors in grading such as obedience, effort, ability, and motivation. Such factors seem remote from what a grade is supposed to represent, which is the amount of

student learning for a specific period of time. “This conflict does not foster good teaching, effective learning, or a harmonious learning environment” (Haladyna, 1999, p. 3). In a study done by McREL in 1996, 640 teachers in K-12 were asked to identify those skills and abilities in addition to subject matter content that they consider in grading. Thirty-six percent of the teachers of 7th-9th grade use effort, while only 10% (same grades) used behavior (Marzano, 2000). Marzano (2000) pointed out that academic achievement should be the primary factor when developing a grade, but also noted that a compromise could be reached to include behavior as part of the grade such that, “It is appropriate to provide feedback to students on their effort, behavior, and attendance, (and) ideally, this feedback should be kept separate from that provided on academic achievement” (p. 39). “Reporting achievement separately from behaviors means that everyone can know as accurately as possible what a grade means in achievement terms” (O’Connor, 2007, p. 21).

O’Connor (2002) stated, “Strong effort, active participation, and positive attitude are highly valued attributes, but they are reporting variables, not grading variables” (p. 100). He also noted that personal and social characteristics do contribute to achievement, but including attitude in a mark for a product “blurs the assessment of the product and affects the validity and thus the meaning of the grade” (p. 72). In addition, including marks for effort can disproportionately jeopardize the already struggling learner.

Additionally, no common or consistent practices are often established to create a clear understanding of how grading is done. Teachers often weight assessments differently. “Grades have long been recognized in the measurement community as prime

examples of unreliable measurement. What one teacher considers in determining students' grades may differ greatly from the criteria used by another teacher" (Guskey & Bailey, 2001, p. 12).

O'Connor (2007) described the use of extra credit as a distortion of achievement. "Extra credit and bonus points can distort a student's record of achievement—grades are broken as a communication tool if we give points for 'dressing like an Egyptian' when such 'performances' do not demonstrate achievement of specified academic standards" (p. 31). Others agree. "Do not record points for mere completion of extra credit work and consider it achievement information" (Stiggins et al., 2004, p. 313).

Another area that deserves conversation is that of grading of the individual, not the group. Kagan (1995) argued that group grades violate individual accountability. Group grades would fly in the face of the very targets or standards deemed essential for students in the classroom. "Group scores may not accurately reflect the achievement of each student and therefore would be unfair for some members of the group" (O'Connor, 2007, p. 48).

Component #4—Engagement of the learner. The learners themselves actually permeate all of the above components, as they become part of the process every step of the way. O'Connor (2007) suggested that students must be involved in all stages of the assessment process. "Don't leave students out of the grading process. Involve students; they can—and should—play key roles in assessment and grading that promote achievement" (p. 111). The degree to which they are involved in establishing the standards and expectations for learning, the assessment of their own learning, and their understanding and participation in the marking of academic achievement is indicative of

a strong standards-based grading practice. Stiggins et al. (2004) described engaging students in the assessment and grading process.

Whenever students interpret their performance to be below what they want in their record of achievement, they can be given the opportunity to study more, learn more, and retake that assessment. This is especially crucial when the material in question is prerequisite for later learning. If the objective is to bring all students to appropriate levels of mastery of standards, anything we can do to keep students learning and wanting to succeed is worth doing. (p. 325)

McMillan (2009) suggested that students must be aware of the facets of evaluation because they are critical to achievement. “It is unfair for students to not know, at the onset of learning, the basis for grading, the standards on which they will be graded, and the criteria that will be used in the evaluation” (p. 114).

Student self-assessment has also been advocated. Self-assessing occurs as students evaluate the quality of their performance when learning. “Since self-grading is performed in relation to standards of achievement and criteria used to evaluate whether achievement goals have been attained, the pervasiveness of standards-based instruction provides an ideal context in which student self-grading can be fostered” (McMillan, 2009, p. 116). Sadler and Good (2006) described two specific advantages of student self-grading over teacher-grading. There are improvements in general learning when students generate deeper understanding of what is being learning. In addition, students develop meta-cognitive skills to increase their awareness of strengths and gaps in their learning, as they improve their self-monitoring skills. According to several researchers (O’Connor, 2007; Stiggins et al., 2004), having students track their own academic progress and achievement and then communicate about their learning with others is one of the most powerful tools in improving student achievement. “When students know how they will be assessed, and especially when they have been involved in the assessment

decisions, the likelihood of student success is increased greatly” (O’Connor, 2002, p. 117).

Summary

This chapter examined the literature base surrounding standards-based grading. The literature described professional growth and its role in standards-based grading, the influence of STARS (School-Based, Teacher-Led Assessment and Reporting System) on grading practices, and the components of standards-based grading. More specifically, the components of learning standards, assessment, and academic achievement markers were detailed. Based on the literature, a gap still exists between knowledge and adoption of quality standards and assessment practices and transferring this knowledge to reporting of student achievement using standards-based grading.

Chapter 3 discusses the use of the quantitative survey design for this study. The role of the researcher, the research design and methodology, and the survey instrument will be described. Data collection and analysis procedures are also described in this chapter.

Chapter 3

Methodology

Introduction

The purpose of this quantitative survey study was to determine if rural 7th-12th grade English/Language Arts teachers in Nebraska use standards-based grading in place of traditional grading procedures. More specifically, to which grading components are teachers drawn? Underlying this purpose was the intent to investigate the knowledge base of teachers regarding their understanding of the interconnectedness of standards-based assessment, teaching and learning, and standards-based grading. The researcher served as an English/language arts teacher in Nebraska rural schools, which was the basis for her interest. In addition, the researcher spent great amounts of time working with standards and assessment in Nebraska schools as a teacher, staff developer at an educational service unit, and as a district curriculum and assessment director. It was during the course of this work that the researcher noted a disconnect between the use of quality standards and assessment practices with the use of traditional grading practices.

Research Design

In quantitative research, the researcher determines what to study, asks specific, focused questions, collects numeric data from participants, analyzes these numbers using statistics, and conducts the inquiry in an unbiased, objective manner (Creswell, 2005). This type of research emphasizes the collection and analysis of information in the form of numbers, the collection of data that measure distinct attributes of individuals, and the procedures of comparing groups and/ or relating factors about individuals or groups in experiments, correlation studies, and surveys (Creswell, 2005). Quantitative research

follows the pattern of identifying the research problem, using literature to build a basis for additional research need, followed by the specific research design of data collection, analysis of results, and description of the findings (Creswell, 2005).

Research Objective and Questions

The objective of this quantitative survey study was to determine whether or not rural 7th-12th grade English/Language Arts teachers in Nebraska are using standards-based grading in place of traditional grading practices. The overarching research question was: How and to what degree are rural 7th-12th grade English/language arts teachers in Nebraska using standards-based grading practices in their classrooms?

The sub-questions for this study were based on the elements necessary for standards-based grading to be in place in rural 7th-12th grade English/language arts classrooms in Nebraska.

1. What learning standards are included in grading by teachers and how do learning standards support standards-based grading?
2. What methods of assessment and re-assessment are included by teachers and how does assessment support standards-based grading?
3. What markers of academic achievement are included by teachers and how do these markers support standards-based grading?
4. What methods of learner engagement are used by teachers and how does learner engagement support standards-based grading?

Population

The population for this study consisted of all 7th-12th grade English/language arts teachers from Nebraska's rural Class III schools. The target population of 636 people

was based on utilizing the Nebraska Department of Education Online staff database that allows an individual to search for teachers with various characteristics. This directory provides for accurate listings of teachers assigned to specific positions in Nebraska school districts. In the 2009-2010 directory, 636 teachers met the characteristics of the population, including 7th-12th grade English/language arts teachers in rural Class III schools. Since the database did not list a school's rural/non-rural classification, a list of schools that were considered non-rural for the 2009-2010 school year was identified using data from the United States Census Bureau. After sorting and cleaning the data, the researcher created a database with teachers' names and school addresses. Individuals in the sample were randomly assigned a survey identification number.

Survey Instrument

When conducting quantitative studies, researchers often use a combination of categorical (e.g., nominal) and continuous (e.g., interval) scales. The use of a nominal scale allowed the researcher to collect data that describes the participants' traits, attributes, or characteristics, while an ordinal scale provides data on the "extent" to which something is of value or is completed (Creswell, 2005, p. 167). Continuous scales, such as the Likert-scale model, provided response options to questions with assumed equal distances between options.

A cross-sectional survey design using a Likert-scale was utilized for this study. A four-section survey was designed for the data collection for this research study. The first section of the instrument focused on the demographics of the participants and will utilize nominal scales. The remaining sections of the survey focused on the three big ideas of standards-based grading obtained from the literature review: standards, assessment, and

academic achievement markers. A fourth concept that permeated all of the sections is that of student engagement in standards-based grading. Items on these sections of the survey were rated using a five-point Likert-scale in order to capture the perceptions and practices of the assessment and grading practices of rural 7th-12th grade English/language arts teachers in Nebraska. Tables 1 and 2 provide the Demographic and Standards-Based Grading Big Ideas Match to the survey items.

Table 1

Demographic Match to Survey Items

Demographic Match	Survey Item
Characteristics of a 7 th -12 th grade English/language arts teacher in Nebraska Class III rural schools	Item #1-Item #9

Table 2

Standards-based Grading Big Ideas Match to Survey Items

Standards-based Grading Big Ideas Match	Survey Item
Learning Standards: Teachers' use of learning standards as support for standards-based grading	Item# 10-Item #12, Item #14
Assessment: Teachers' methods of assessment as support for standards-based grading	Item #15-Item #21
Markers of Academic Achievement: Teachers' use of markers of academic achievement as support for standards-based grading	Item #25-Item#30, Item #32
Student Engagement in standards-based grading	Item #13, Item #22-Item #24, Item #31

There were several benefits to using survey methodology to collect data. First, the list of teachers from which to collect data was up-to-date with current teaching rosters

available from the Nebraska Department of Education; second, a paper sampling method cut down on the duplication of submissions and allowed the researcher to maintain a clear record of responses while allowing for anonymity of the individuals; finally, the use of paper sampling provided a strong return. Some drawbacks included an increased cost over utilizing a web-based method and possibly less-timely collection of the data.

Survey Procedures

Pilot survey procedures. The survey was piloted with 22 7th-12th grade math and science teachers in the researcher's district that are not teachers of English/Language Arts. The researcher used this group of educators because their training in learning standards and assessment literacy is similar to that of the sample for the study. The pilot participants were asked to review the survey items and provide feedback about the clarity of the questions, which improved the internal consistency of the instrument. The pilot was conducted in January and February 2010. No changes were made to the constructed response items on the survey based on the feedback from the pilot's participants.

Study Survey Procedures

Administration of the paper/pencil survey took place during the second semester of the 2009-2010 school year. After receiving approval to conduct the study from the University of Nebraska Institutional Review Board, the researcher proceeded with the survey. Participants were approached using a mailed survey, asking for their participation. The first mailing included a cover letter describing the research study and its purpose along with the survey and a self-return addressed stamped envelope. The

survey instrument used in the study asked for general demographic information and teacher behavior responses related to classroom grading practices.

After 10 days, a reminder postcard was mailed to participants who had yet to return their survey. At the end of the three weeks, a thank you postcard was mailed to those individuals who responded. The timeline for the implementation of the survey was as follows:

Anticipated Date:	Stage
Monday, February 22, 2009	Letter and Survey mailed
Monday, March 8, 2009	Reminder Postcard mailed
Monday, March 15, 2009	Thank You Postcard mailed

Variables

The critical dependent variables of this study were the various components of standards-based grading utilized in the classroom. The independent variables included the teacher's gender, teacher's age, the size of the school district in which the teacher works, the length of the teacher's educational career, the teacher's educational background, and the teacher's training in assessment literacy, the gateway for increasing usage of standards-based grading.

Data Analysis

Survey designs are procedures in quantitative research in which researchers measure the degree of association (or relationship) between two or more variables using the statistical procedure of correlation analysis (Creswell, 2005). Analysis consists of noting response rates, checking for response bias, conducting descriptive analysis of all items, and then answering descriptive questions. The data were transferred from the

survey instrument and keyed into an Excel document. This document was used for an SPSS analysis, as suggested by the Nebraska Evaluation and Research (NEAR) Center. The researcher organized the collected data into specific and like categories and worked with the NEAR Center on the analysis of the data.

Scoring data means that the researcher assigned a numeric score (or value) to each response category for each question on the instrument used to collect data. When using the categorical scales, the researcher was able to use both nominal and ordinal scales. When using the ordinal scales, the researcher assigned numbers to the responses. For items 15, 20, 21, 26, 27, 28, 29, and 30, the researcher also used reverse-coding of the responses in order to collect consistent data. The researcher developed a code sheet in order to identify how the researcher would code the responses from the survey in order to assist the NEAR Center in the data analysis. For this study, both single-item and summed scores were used. A single-item score is an individual score assigned to each question for each participant in the study. The scores provided a detailed analysis of each person's response to each question on the instrument. Summed scores are the scores of an individual that are added up over several questions and measure the same variable (Creswell, 2005). This allowed the researcher to review the items by the three big ideas of standards-based grading: standards, assessment, and academic achievement markers.

The data was reviewed, cleaned, and assessed for missing data. The researcher considered all data collected in the survey, only excluding data that fell outside the range of the study. Participants with missing scores were still included in the results, if the majority of their surveys were completed (at least 14 items).

The researcher analyzed the data using descriptive statistics that reveal general tendencies in the data and the comparison of how one score relates to the others, including the dependent and independent variables. The researcher measured reliability of the instrument's internal consistency using coefficient alpha.

The researcher also analyzed the data's variability using inferential statistics. The researcher chose to run ANOVA instead of a t-test, because of the multiple factors involved in the research and the desire of the researcher to control for various factors. When conditions warranted, the researcher also conducted post hoc tests using Tukey's honestly significant difference (HSD) test to compare more than two groups within a single independent variable.

Validity of Study

The study was validated through the use of statistical analysis and criterion validity to measure the use of various components of standards-based grading in the 7th-12th grade English/Language Arts classroom. Internal validity was ensured through the standardized directions and follow-up methods for the survey tool. External validity was present in the fact that the survey was designed for all 7th-12th grade English/Language Arts teachers, regardless of gender, age, background, geographic region, or school size. The study might have been impacted by the number of responses received. In addition, there was no way to measure the truthfulness of the responses provided by the participants.

Reliability of Study

The use of an internal consistency reliability coefficient was utilized. The majority of the items were scored as categorical variables (e.g., never to always), thus the alpha provided a coefficient to estimate consistency of scores on the instrument.

Protection of Participants' Rights

Many measures were taken by the researcher to protect the rights of the participants of the study. First, no data were collected until the study was approved by the Institutional Review Board (IRB). An IRB form was filed with the Institutional Review Board at the University of Nebraska-Lincoln, detailing the principal investigator, the number of subjects in the study, and the nature of the research study.

Other than a unique identification number on the survey instrument, no other personal information was connected to the responses. The cover letter attached to the survey stated that their participation was voluntary. Completing and returning the survey implied consent. The participants were assured that their responses are confidential and will not impact their jobs.

Summary

Chapter 3 reviewed the research methodology and design that were employed during the quantitative research study. The instrument for research was a four-section cross-sectional survey. This survey was mailed to all 7th-12th grade English/language arts teachers in Nebraska Class III rural schools. The data analysis for the study included the use of descriptive statistics in order to understand where one score stands in comparison to the others as well as understanding the statistical significance of the data collected.

Chapter 4

Results

Introduction

This study was conducted to explore the use of standards-based grading practices by rural 7th-12th grade English/Language Arts teachers in Nebraska. This chapter is organized around four specific research sub-questions:

1. What learning standards are included in grading by teachers and how do learning standards support standards-based grading?
2. What methods of assessment and re-assessment are included by teachers and how does assessment support standards-based grading?
3. What markers of academic achievement are included by teachers and how do these markers support standards-based grading?
4. What methods of learner engagement are used by teachers and how does learner engagement support standards-based grading?

A 32-item survey instrument was used to collect data from participants. The initial section of the survey identified the independent variables, including the teacher's gender, age, the size of the school district in which the teacher works, the length of the teacher's educational career, the teacher's educational background, and the teacher's training in assessment literacy. The subsequent three parts of the survey addressed the four research sub-questions addressed above.

Presentation of Descriptive Characteristics of Participants

Of the 636 individuals that were mailed surveys, 312 teachers responded to this survey, yielding a 49.1% response rate. Of this group, 19.2% (n = 60) were male and

80.1% (n = 250) were female. Two individuals did not indicate their gender. Table 3 reports the frequencies and percents related to the gender of the participants in the study. This breakdown of responses mirrors the characteristic population of 7th-12th grade English/Language Arts teachers in Nebraska.

Table 3

Gender of Participants

Gender of Participants	Frequency	Percent
Male	60	19.2
Female	250	80.1
Unidentified	2	0.7
Total	312	100.0

Ages of the participants varied. Of the 312 participants, 16% (n = 50) were age 20-29, 19.9% (n = 62) were age 30-39, 21.5% (n = 67) were age 40-49, 28.5% (n = 89) were age 50-59, 12.5% (n = 39) were age 60 or older, and 1.6% (n=5) did not respond. Table 4 reports the frequencies and percents of the ages of the participants in the study. Fifty-eight percent of the participants (n = 179) were 20-49 years of age, while 42% (n = 128) were 50-60+.

Of the 312 participants, 49% (n = 153) possessed education below a master's degree, and 51% (n = 159) possessed a master's degree or higher. Table 5 reports the frequencies and percents related to the highest degree held by participants in the study.

Table 4

Ages of Participants

Age of Participants	Frequency	Percent
20-29	50	16.0
30-39	62	19.9
40-49	67	21.5
50-59	89	28.5
60+	39	12.5
Unidentified	5	1.6
Total	312	100.0

Table 5

Education Levels of Participants

Level of Education	Frequency	Percent
bachelor's degree	30	9.6
bachelor's degree plus graduate hours	123	39.4
master's degree	58	18.6
master's degree plus hours	99	31.7
doctoral degree	2	0.6
Total	312	100.0

Of the 312 participants, 17% (n = 54) reported being in the teaching profession from 1-5 years, 18.6% (n = 58) for 6-10 years, 9.6% (n = 30) for 11-15 years, 12.2% (n = 38) for 16-20 years, 11.2% (n = 35) for 21-25 years, 30.4% (n = 95) for 25 or more years, and (n=2) did not respond. Table 6 reports the frequencies and percents related to the years teachers have been teaching. Thus, 35.6% (n = 112) of the participants reported

Table 6

Years in Education of Participants

Years in Education	Frequency	Percent
1-5	54	17.0
6-10	58	18.6
11-15	30	9.6
16-20	38	12.2
21-25	35	11.2
25+	95	30.4
Unidentified	2	0.6
Total	312	100.0

being in the profession 10 or less years, while 64.4% (n = 198) reported teaching more than 10 years.

Of the 312 participants, the most frequently held endorsement was English, with 47.4% (n = 148). Other endorsements within the subject area reported included 22.1% (n = 69) of participants holding a Language Arts endorsement, while 18.9% (n = 59) reported holding an English/Language Arts endorsement. Only 6% (n = 19) of the participants did not hold an endorsement in the English/Language Arts subject area. Other endorsements listed by 180 participants included Social Sciences, Physical Education/Coaching, Counseling, World Languages, Special Education, Elementary Education, Middle School Endorsements, as well as a variety of other endorsements. Table 7 lists the endorsement reported by the participants.

Table 7

Endorsements of Participants

Endorsement	Frequency	Percent
English	148	47.4
language arts	69	22.1
English/language arts	59	18.9

Of the 312 participants, eight (2.6%) of the participants indicated that they possessed a Nebraska Assessment Endorsement, indicating they had participated in one of the University of Nebraska-Lincoln's Assessment Cohorts. The largest number of participants indicated participating in assessment training provided by an Educational Service Unit (73.4 %, n = 229). A number of participants (n = 218, 69.9%) indicated that they had participated in assessment training provided by a school district. Twenty-eight participants (9%) indicated that they had no formal assessment training. One hundred eighty-seven participants (59.9%) indicated that they had participated in the development of criterion-referenced assessments in their school districts. Table 8 details the assessment professional development of the participants.

Additional demographic data collected in the survey included grades taught by the participants, building configurations of the participants, and size of the school districts of the participants. Grades taught and building configuration data was displayed below for informational purposes. The above variables, however, were not viable data that could be used because the variability of the responses was too great. Of the 312 participants, 208 (66.7%) indicated that they taught Grade 12 students, while 61.5% (n = 192) reported

Table 8

Assessment Professional Development of Participants

Assessment Training	Frequency	Percent
Assessment Endorsement	8	2.6
Educational Service Unit Assessment Training	229	73.4
District Assessment Training	218	69.9
No Formal Training	28	9.0
Criterion-Referenced Test Writing Experience	187	59.9

teaching Grade 11, 60.9% (n = 190) taught Grade 10, 51.6% (n = 161) taught Grade 9, 33% (n = 103) taught Grade 8, and 30.8% (n = 96) taught Grade 7. Table 9 reports the grades taught by the participants.

Table 9

Grades Taught by Participants

Grade Taught	Frequency	Percent
7	96	30.8
8	103	33.0
9	161	51.6
10	190	60.9
11	192	61.5
12	208	66.7

Of the 312 participants, 101 reported teaching in K-12 buildings (32.4%). One hundred one participants reported teaching in a 7-12 building, 80 reported teaching in a

9-12 building, and 30 reported teaching in a building with some other grade configurations. Table 10 provides the grade configurations of the participants' schools.

Table 10

Grade Configurations of Participants' Schools

Grade Configurations of Participants' Schools	Frequency	Percent
K-12	101	32.4
7-12	101	32.4
9-12	80	25.6
Other configurations	30	9.6
Total	312	100.0

Results by Research Question

Based on a survey assessing the use of the four components of standards-based grading practices (learning standards, various assessment practices, academic achievement markers, and involving students in grading), teachers were asked to respond to survey items using a Likert-scale 1-5 response, with 1 = Never, 2 = Infrequently, 3 = Sometimes, 4 = Frequently, and 5 = All of the time. The results will be displayed by research sub-question.

Due to the reverse polarity of survey-item wording, several items were reverse-coded in addition to running analysis on the original responses. Specifically, items 15, 20, 21, 26, 27, 28, 29, and 30 from the primary survey were reverse-coded. Reverse coding was conducted in order to allow for the final numeric response to represent a higher value to indicate a supported use (or non-use) of an attribute of standards-based grading. In

representing the results below, however, both the original and reverse coding were displayed.

Research Sub-Question #1: What learning standards are included in grading by teachers and how do learning standards support standards-based grading? The first research sub-question addressed in this study sought to identify how often learning standards were used by classroom teachers in terms of their course objectives, lesson planning, and grading practices. In this category (Survey Items 10-13), the average of the scores for all participants was 3.70, indicating these teachers included learning standards in grading “Sometimes” to “Frequently.” Table 11 shows the descriptive statistical breakdown for the four survey items that addressed Research Sub-question #1.

Table 11

Survey Section II—Learning Standards Descriptive Statistics

Survey Question	N	Mean	Mode	SD	Variance
10. To what extent do you use local or state standards as a basis for your course objectives?	311	4.17	4	.847	.718
11. To what extent do you use local or state standards as a basis for your lessons?	311	4.03	4	.825	.681
12. To what extent do you use local or state standards as a basis for your students' grades?	311	3.33	4	1.032	1.066
13. To what extent are models of students work (exemplars) shared with and evaluated by students?	311	3.29	3	.856	.732

Note: Learning Standards Likert Scale: 1 = Never, 2 = Infrequently, 3 = Sometimes, 4 = Frequently, and 5 = All of the time

The item rated strongest within this category was item 10, “To what extent do you use local or state standards as a basis for your course objectives?”, with a mean of 4.17. Item 13, “To what extent are models of students’ work (exemplars) shared with and evaluated by students?”, had a mean score of 3.29. The survey reliability statistic (Cronbach’s Alpha) for items 10-13 in this section was .785. Discussion of item 13 can be found under the fourth sub-question data analysis—Student Involvement.

Item 10: “To what extent do you use local or state standards as a basis for your course objectives?” This item was rated by 311 participants, and the mean of the scores was 4.17, indicating between “Frequently” and “All of the time.” When ANOVA was run, no statistical significance was identified in the independent variables.

Males reported scores with a mean of 4.07, and females reported scores with a mean of 4.18. Teachers age 20-29 reported scores with a mean of 4.07, while teachers age 30-39 reported scores with a mean of 4.24.

Teachers who had taught for 1-5 years reported scores with a mean of 4.02, while teachers who had taught for 6-10 years reported scores with a mean of 4.31. Teachers with master’s degrees or higher reported scores with a mean of 4.20, while teachers with degrees and coursework below the master’s level reported scores with a mean of 4.11.

Teachers holding an assessment endorsement reported scores on item 10 with a mean of 4.25, while teachers who had participated in Education Service Unit training on assessment reported scores with a mean of 4.22. Participants who indicated having participated in district level training reported scores with a mean of 4.18, while those who reported having worked on assessments at the district level and those that reported never

having received assessment training reported scores with a mean of 4.17 and 4.14, respectively.

Item 11: “To what extent do you use local or state standards as a basis for your lessons?” This item was rated by 311 participants, and the mean of the scores was 4.03, indicating “Frequently” being used. A one-way ANOVA was used to test for a difference between males’ and females’ use of standards as the basis for lesson planning. Use of standards differed significantly between the two genders, $F(1, 308) = 4.189$, $p = .042$. Males reported scores with a mean of 3.82, and females reported scores with a mean of 4.05. No other statistical significance was found when comparing other independent variables.

Teachers age 50-59 reported scores with a mean of 4.07, while teachers age 30-39 and 60+ reported scores with a mean of 4.05. Teachers age 40-49 reported scores with a mean of 3.97, while teachers in the youngest age group, 20-29, reported scores with a mean of 3.86. Teachers who had taught for 6-10 years reported scores with a mean of 4.17, while teachers who had taught 25+ years reported scores with a mean of 4.05.

Teachers who had taught 1-5 years reported scores with a mean of 3.85. Teachers with master’s degrees or higher reported scores with a mean of 4.04, while teachers with degrees and coursework below the master’s level reported scores with a mean of 3.97.

Teachers holding an assessment endorsement reported scores with a mean of 4.13, while teachers who had participated in Education Service Unit training on assessment reported scores with a mean of 4.07. Teachers who indicated having participated in district level training and those who assisted in writing district level assessments reported

scores with a mean of 4.04, while the 27 participants that reported never receiving assessment training reported scores with a mean of 4.14.

Item 12: “To what extent do you use local or state standards as a basis for your students’ grades?” This item was rated by 311 participants, and the mean of the scores was 3.29, indicated “Sometimes” to “Frequently” being used. When ANOVA was run, no statistical significance was identified in the independent variables.

Males reported scores with a mean of 3.20, and females reported scores with a mean of 3.44. Teachers age 30-39 reported scores with a mean of 3.47, while teachers age 50-59 reported scores with a mean of 3.39. Teachers age 40-49 reported scores with a mean of 3.34, while teachers in the oldest age group, 60+, and those in the youngest age group, 20-29, reported scores with means of 3.15 and 3.16 respectively.

Teachers who had taught for 6-10 years reported scores with a mean of 3.53, while teachers who had taught 25+ years reported scores with a mean of 3.33. Teachers who had taught 1-5 years reported scores with a mean of 3.15. Teachers with master’s degrees or higher reported scores with a mean of 3.38, while teachers with degrees and coursework below the master’s level reported scores with a mean of 3.25.

Teachers holding an assessment endorsement reported scores with a mean of 3.75, while those teachers who indicated having participated in Education Service Unit training on assessment reported scores with a mean of 4.42. Participants who indicated having participated in district level training reported scores with a mean of 3.38, while teachers that reported never having received assessment training and those who reported assisting in writing district level assessments reported scores with a mean of 3.29.

Item 14: “Does your current report card reflect student performance on local or state standards?” In this constructed response item, participants were asked to identify if their school’s current report card reported results of student performance on local or state standards. Thirty-three participants did not respond to this item (10.58%), while 78.53% (n = 245) of the participants indicated that their report card did not reflect performance on local or state standards. In addition, 10.89% (n = 34) of participants indicated that in some form, their school’s report card indicated student performance on local or state standards.

Research Sub-Question 2: “What methods of assessment and re-assessment are included by teachers and how does assessment support standards-based grading?” The second research sub-question addressed in this study sought to identify what methods of assessment and measurement were included by teachers in determining grades and how assessment practices support standards-based grading. Items 17, 18, and 19, presented as a cluster, had a reliability of .746 and scores with a mean 3.91. Items 15 and 16 did not align with other survey items, so no group reliability number was available. Items 20 and 21 lacked variance in responses, resulting in a coefficient alpha of .083. The lack of variance in responses can be accounted for in the particular nature of the item.

The item rated strongest within the category of assessment practices was item 18 (mean 3.93), “To what extent are skills taught and retaught to mastery?” The item with the scores with the lowest mean on the original responses was item 20, “To what extent do you assign zeros for assigned work? Table 12 shows the descriptive statistical

breakdown for the seven survey items that addressed Research Sub-question #2, including those items that were also reverse-coded.

Table 12

Survey Section III—Assessment Practice Descriptive Statistics—Items 15-21

Survey Question	N	Mean	Mode	SD	Variance
15. To what extent is formative assessment, such as homework or practice, used for the basis of grades?	310	3.84	4	2.412	.816
15.* To what extent is formative assessment, such as homework or practice, used for the basis of grades?	310	2.28	2	.917	.841
16. To what extent are summative assessments used for the basis of grades?	309	3.86	4	.836	.841
17. To what extent are concepts taught and retaught to mastery?	309	3.88	4	.610	.698
18. To what extent are skills taught and retaught to mastery?	311	3.93	4	.608	.372
19. To what extent are student assessment results used to adjust, improve, or support instruction?	312	3.92	4	.765	.585
20. To what extent do you assign zeros for assigned work?	307	2.65	2	.983	.965
20.* To what extent do you assign zeros for assigned work?	307	3.35	3	.983	.965
21. To what extent do you average scores to assign a grade?	306	3.20	5	1.464	2.142
21.* To what extent do you average scores to assign a grade?	306	2.80	3	1.464	2.143

Notes: Assessment Practice Likert Scale: 1 = Never, 2 = Infrequently, 3 = Sometimes, 4 = Frequently, and 5 = All of the time

** = Reverse-coding item*

Item 15: “To what extent is formative assessment, such as homework or practice, used for the basis of grades?” This item was rated by 310 participants, and the mean of the scores was 3.84, equating to “Sometimes” to “Frequently” on the Likert-scale (mean was 2.28 when reverse-coding was applied). Reverse coding was conducted in accordance with making the higher numeric response to an item indicate a positive attribute, indicating a stronger connection to standards-based grading. Original results depicted that participants marked the item with a mean of 3.84, indicating that they used formative assessment results for the basis of grades between “Sometimes” and “Frequently,” which is considered a negative attribute in standards-based grading. Thus in this instance, the higher the score, the less supportive the practice is with the tenets of standards-based grading. In addition to running analysis to delineate the degree to which the tenets of standards-based grading are being followed, the researcher also ran ANOVA, identifying no statistical significance in the independent variables.

When using the original survey responses, males reported scores with a mean of 3.63, while females reported scores with a mean of 3.74. Teachers age 50-59 reported scores with a mean of 3.58, while teachers age 20-29 reported scores with a mean of 3.96. Teachers age 40-49 reported scores with a mean of 3.62, teachers age 60+ reported scores with a mean of 3.64, and teachers age 30-39 reported scores with a mean of 3.84. Teachers who had taught for 21-25 years and 25+ years reported scores with means of 3.63 and 3.62, respectively, while teachers who had taught 1-5 years reported scores with a mean of 3.94. Teachers with master’s degrees or higher reported scores with a mean of 3.64, while teachers with degrees and coursework below the master’s level reported scores with a mean of 3.79.

Teachers holding an assessment endorsement reported scores with a mean of 3.63, while teachers who participated in Education Service Unit training on assessment reported scores with a mean of 3.75. Participants who indicated having participated in district level training reported scores with a mean of 3.38, while those participants that reported never receiving assessment training and those who reported assisting in writing district level assessments reported scores with means of 3.78 and 3.70, respectively.

Item 16: “To what extent are summative assessments used for the basis of grades?” This item was rated by 309 participants, and the mean of the scores was 3.86 (“Sometimes” to “Frequently”). An ANOVA was run, and no statistical significance was identified in the independent variables.

Males reported scores with a mean of 3.76, while females reported scores with a mean of 3.89. Teachers age 60+ reported scores with a mean of 3.71, while teachers age 30-39 reported scores with a mean of 3.98. Teachers age 40-49 reported scores with a mean of 3.83, teachers age 20-29 reported scores with a mean of 3.96, and teachers age 50-59 reported scores with a mean of 3.80. Teachers who had taught for 11-15 years and 21-25 years reported scores with the means of 3.77 and 3.63, respectively, while teachers who had taught 16-20 years reported scores with a mean of 4.16. Teachers with master’s degrees or higher reported scores with a mean of 3.96, while teachers with degrees and coursework below the master’s level reported scores with a mean of 3.76.

Teachers holding an assessment endorsement reported scores with a mean of 4.00, while teachers who participated in Education Service Unit training on assessment reported scores with a mean of 3.84. Participants who indicated having participated in district level training reported scores with a mean of 3.85, while those participants that

reported never having received assessment training and those who reported assisting in writing district level assessments rated the item 4.00 and 3.84, respectively.

Items 17, 18, 19: “To what extent are concepts taught and retaught to mastery? To what extent are skills taught and retaught to mastery?” and “To what extent are student assessment results used to adjust, improve, or support instruction?” These items focused on using mastery of learning as a focus for instruction and grading. Additionally, these items also focused on the teaching and re-teaching of concepts and skills to mastery and the use of the assessment results to adjust, improve, or support instruction. On item 17, 309 participants responded, while 311 and 312 participants responded to items 18 and 19, respectively. The three items’ mean response was 3.91 (“Sometimes” to “Frequently”). When ANOVA was run, no statistical significance was identified in the independent variables.

On the combined items (17-19), males reported scores with a mean of 3.83, while females reported scores with a mean of 3.93. Teachers age 30-39 reported scores with a mean of 3.97, while teachers age 20-29 reported scores with a mean of 3.80. Teachers age 40-49, 50-59, and 60+ reported scores with the means of 3.87, 3.92, and 3.91, respectively. Teachers who had taught for 16-20 years reported scores with the mean of 4.06, while those who had taught for 21-25 years reported scores with the mean of 3.74.

Teachers with less than a master’s degree reported scores with the mean of 3.94, while those with master’s degrees or higher reported scores with the mean of 3.88. Teachers who reported having some assessment training (either through the Assessment Cohort, ESU trainings, or district-level training), reported an average of 3.93 on items 17-19, while those that did not have any assessment training reported an average of 3.69.

Items 20-21: “To what extent do you assign zeros for assigned work?” and “To what extent do you average scores to assign a grade?” Items in this category reported very low reliability, but the construct of the survey, along with the participant group characteristics, may account for the low variance of responses. Reverse coding was conducted in accordance with making the higher numeric response to an item indicate a positive attribute, indicating a stronger connection to standards-based grading. Reverse-coded results for item 20 presented a mean of 3.345, which actually provided a more negative connection to standards-based grading by detracting from the original scores mean of 2.65 (indicating that they assigned zeros to assigned work “Infrequently” to “Sometimes”); however, reverse-coded results for item 21 presented a mean of 2.80, which indicates that original results depicted that participants reported scores with a mean of 3.20 (indicating they average student work “Sometimes”).

When ANOVA was run on item 20, both as stated and reverse-coded, no statistical significance was identified in the independent variables. Using the reverse coded item 21, a one-way ANOVA was used to test for differences among ages of teachers in their response to averaging of student scores. Use of averaging differed significantly across the age groups, $F(4, 296) = 5.121, p = .001$. Tukey’s honestly significant difference (HSD) post-hoc comparison of teachers age 20-29 and teachers age 60+ indicates that the older teachers ($M = 2.03, 95\% \text{ CI } [1.61, 2.44]$) reported more frequent use of averaging to figure student grades than the teachers age 20-29 ($M = 3.32, 95\% \text{ CI } [2.88, 3.76]$), $p = .000$. Comparisons between the other age groups were not statistically significant at $p < .05$.

In addition on the reverse coded item 21, Tukey's HSD comparison of teachers with 25+ years experience and teachers with 1-5 and 6-10 years experiences indicates that teachers with 25+ years experience ($M = 2.28$, 95% CI [1.99, 2.58]) reported more frequent use of averaging to figure student grades than the teachers with experience of 1-5 years ($M = 3.49$, 95% CI [3.09, 3.89]), $p = .000$ and 6-10 years ($M = 3.12$, 95% CI [2.76, 3.48]), $p = .006$. No other statistical significance was found when comparing other independent variables.

The following analysis of items 20 and 21 was run using the original survey responses of 307 and 306 participants, respectively, in order to maintain fidelity with the remainder of the research sub-question. When using the original survey responses, males reported scores on item 20 with a mean of 2.76, while females reported scores same item with a mean of 2.63. On item 21, males reported scores with a mean of 3.31, while females reported scores on that same item with a mean of 3.17. On item 20, teachers age 20-29 reported scores with a mean of 2.92, while teachers age 50-59 reported scores with a mean of 2.48.

On item 20, teachers who had taught 21-25 years experience reported scores with a mean of 2.35, while teachers who had taught between 16-20 years reported scores with a mean of 2.79. On item 20, teachers with less than a master's degree reported scores with a mean of 2.67, while those with a master's degree or higher reported scores with a mean of 2.64. On item 21, teachers with less than a master's degree reported scores with an mean of 3.17, while teachers with a master's degree or higher reported scores with a mean of 3.23.

Scores for teachers who participated in assessment training through the Assessment Cohort, ESU training, or district-level training had a mean of 2.75, 2.59, and 2.67 respectively on item 20 and a mean of 3.29, 3.17, and 3.27 respectively on item 21.

Research Sub-Question 3: “What markers of academic achievement are included by teachers and how do these markers support standards-based grading?”

The question was addressed using the third section of the survey and included indicators that were both academic and non-academic markers. The survey items dealt with academic performance as indicators of student achievement, as well as the use of effort, behavior, and attendance in the grading of student performance. Items 29-30 focused on the use of extra credit and assigning group grades to students respectively. The mean on the original responses was 2.14 and 2.40, respectively, ranging from responses of “Never” to “Frequently.” The means of items 29-30, when recoded, become 3.859 and 3.596 respectively. Table 13 shows the descriptive statistical breakdown for the six survey items that addressed Research Sub-question #3.

Item 25: “To what extent do you include academic achievement as the basis for grades?” This item was rated by 307 participants, and the mean of the scores was 4.01 (“Frequently”). When ANOVA was run, no statistical significance was identified in the independent variables.

Males reported scores with a mean of 3.80, while females reported scores with a mean of 4.07. Teachers age 40-49 reported scores with a mean of 4.09, while teachers age 30-39 reported scores with a mean of 3.90. Teachers who had taught 1-5 years reported scores with a mean of 3.87, while teachers who taught 16-20 years and more than 25 years of experience reported scores with means of 4.16 and 4.10, respectively.

Table 13

*Survey Section III—Markers of Academic Achievement Descriptive Statistics—**Items 25-30*

Survey Question	N	Mean	Mode	SD	Variance
25. To what extent do you include academic achievement as the basis for grade?	307	4.01	4	.893	.797
26. To what extent do you include effort as the basis for grades?	310	3.08	3	.882	.777
26.* To what extent do you include effort as the basis for grades?	310	2.92	3	.882	.777
27. To what extent do you include behavior as the basis for grades?	312	1.87	1	.907	.822
27.* To what extent do you include behavior as the basis for grades?	312	4.13	4	.907	.822
28. To what extent do you include attendance as the basis for grades?	311	1.62	1	.893	.797
28.* To what extent do you include attendance as the basis for grades?	311	4.38	5	.893	.797
29. To what extent do you include extra credit as the basis for grades?	312	2.14	2	.841	.707
29.* To what extent do you include extra credit as the basis for grades?	312	3.86	4	.841	.707
30. To what extent do you give group grades for assignments?	312	2.40	2	.733	.537
30.* To what extent do you give group grades for assignments?	312	3.60	4	.733	.537

Note: Markers of Academic Achievement Likert Scale: 1 = Never, 2 = Infrequently, 3 = Sometimes, 4 = Frequently, and 5 = All of the time

** = Reverse-coding*

Teachers with a master's degree or higher and those with a degree lower than a master's degree reported scores with similar means 4.01 and 4.00, respectively. Teachers holding assessment endorsements reported scores with a mean of 3.50, while teachers with ESU

and/or district assessment training reported scores with means of 3.03 and 4.02, respectively. Teachers who indicated having no formal assessment training reported scores with a mean of 4.07.

Items 26-28: “To what extent do you include effort (behavior/attendance) as the basis for grades?” On the original survey items, 310, 311, and 312 participants reported scores on items 26, 27, and 28 with means of 3.08, 1.87, and 1.62, respectively. These means equate to “Sometimes” on item 26 and between “Never” and “Infrequently” on items 27 and 28. The means of items 26-28, when reverse-coded became 2.92, 4.13, and 4.28 respectively. The following analysis of items 26-28 was run using the original survey responses in order to maintain fidelity with the remainder of the research sub-question.

When ANOVA was run on item 26 (both the original and reverse-coded responses, no statistical significance was identified in the independent variables. Using the reverse coded item 27, a one-way ANOVA was used to test for differences in the use of including behavior in student grades. Use of behavior in figuring student grades differed significantly in teachers who had taught different lengths of time, $F(5,303) = 7.87, p = .005$. Tukey’s HSD post-hoc comparison of teachers with 25+ years experience and 1-10 years experience indicates that the more experienced teachers with 25+ years experience ($M = 4.49, 95\% \text{ CI } [4.35, 4.63]$) reported less frequent use of behavior in figuring student grades than the teachers with experience of 1-5 years ($M = 3.85, 95\% \text{ CI } [3.57, 4.13]$), $p = .000$ and 6-10 years ($M = 3.71, 95\% \text{ CI } [3.44, 3.97]$), $p = .000$.

Comparisons between the other age groups were not statistically significant at $p < .05$. When ANOVA was run on item 28 (both the original and reverse-coded responses), no statistical significance was identified.

When using the original survey responses, males reported scores with means of 2.83, 1.82, and 1.55 on items 26, 27, and 28, respectively, while females reported scores with means of 3.14, 1.87, and 1.64 on those same items.

On item 26, teachers age 30-39 reported scores with a mean of 3.19, while teachers age 40-49 reported scores with a mean of 2.95. On item 27, teachers age 20-29 reported scores with a mean of 2.20, while teachers age 60 or older reported scores with a mean of 1.67.

On item 26, teachers who had taught 16-20 years reported scores with a mean of 2.92, while teachers who had taught for 6-10 years reported scores with a mean of 3.29. On item 27, teachers who had taught 25+ years reported scores with a mean of 1.51, while teachers who had taught 1-5 years and 6-10 years reported scores with a mean of 2.15 and 2.29 respectively. Teachers who had taught for 11-15 years reported scores with a mean of 1.97, while teachers who had taught 16-20 years reported scores with a mean of 1.76.

On items 26-28, teachers with degrees below master's level reported scores with means of 3.11, 1.97, and 1.70, respectively, while teachers with master's degrees or higher reported scores with means of 3.05, 1.77, and 1.55, respectively. These results would equate to "Sometimes" on item 26 and "Never" to "Infrequently" on items 27 and 28.

Items 29-30: “To what extent do you include extra credit as the basis for grades?” and “To what extent do you give group grades for assignments?” These items asked participants about their use of scores that are not obtained as part of individual academic achievement (extra credit and group grades). These two items were rated by 312 participants, and the mean of the scores for item 29 was 2.14 (use of extra credit) and the mean of the scores for item 30 was 2.40 (giving group grades). These means would equate to “Infrequently” on the Likert-scale. When ANOVA was run, no statistical significance was identified in the independent variables.

On item 29, males reported scores with a mean of 2.15, while females reported scores with a mean of 2.14. On item 30, males reported scores with a mean of 2.35, while females reported scores with a mean of 2.43. On item 29, teachers age 40-49 and age 60+ reported scores with means of 1.96 and 2.00, respectively, while teachers age 20-29 and age 50-59 reported scores with means of 2.24 and 2.30, respectively. On item 30, teachers age 20-29 reported scored with a mean of 2.68, while teachers age 40-49 and those age 60+ reported scores with means of 2.24 and 2.26, respectively.

On both items 29 and 30, teachers who had taught 1-5 years reported scores with means of 2.25 and 2.72, respectively, while teachers who had taught 16-20 years and 20-25 years of experience reported means of 1.89 and 2.32 (16-20 years) and 1.97 and 2.26 (20-25 years), respectively. Teachers with a master’s degree or higher reported scores on items 29 and 30 with means of 2.11 and 2.30, respectively, while teachers with a degree lower than a master’s degree reported scores with means of 2.18 and 2.52 on items 29 and 30.

On item 29, teachers with no specific assessment training reported scores with a mean of 2.54, while teachers holding an assessment endorsement reported scores with a mean of 2.50. Those with ESU and/or district assessment training reported scores with means of 2.11 and 2.12, respectively. On item 30, teachers who indicated having no formal assessment training reported scores with a mean of 2.43, while teachers that had ESU training, district assessment training, and teachers holding an assessment endorsement reported scores with means of 2.39, 2.39, and 2.75, respectively.

Item 32: “Does your school have a common set of grading criteria that applies to all subjects, such as a common grading scale?” Participants provided a variety of answers to this item, with 79.5% (n = 248) of the participants indicating that they had a common set of criteria, although the common criteria were extremely varied among those 248 responses.

Within those responses indicating a common set of criteria existed, responses represent a number of uncommon criteria, which indicates the variability that existed among the definition of common criteria. One common measure is a percentage/letter grade scale; however, the range of the scales included a five, seven, eight, and ten point scales. In addition to the percentage/letter grade scale, letter grade scales were also reported, as were number scales (1-4 and 1-5). Even those teachers who indicated that common criteria existed in their schools identified that common criteria can still be muddled with different expectations by teachers. An 18 year veteran teacher described his school’s common criteria like this: “Yes we are given a common scale, but the criteria is developed by each teacher individually.” A first year, male teacher described his school’s common criteria as follows: “Percentages for letter grades are established by

the school, but teachers decide how to set up their homework, quizzes, and tests to assign those percentages.” Another example of the concerns with the common criteria is exemplified by this six-year veteran, “We have a common scale, but some teachers adjust accordingly to the sport season! (So they can be eligible to play).”

Some of the common criteria described by participants indicated that they utilized both a traditional and a weighted scale for honors courses as well as dual credit courses. While most of the positive responses included letters and percentages, some of those who answered affirmatively indicated that the common criteria they had included not letters and numbers, but rather practices, such as the use of no zeros, limiting the amount of weight various assignments or assessments can have, as well as a no-failure policy for specific core courses.

Research Sub-Question 4: “What methods of learner engagement are used by teachers and how does learner engagement support standards-based grading?”

The question was answered using items that were key components of the first three sections of the survey. The connection of students to their learning is notable in the research about standards-based grading, including awareness of learning targets, participation in viewing exemplars and evaluating their own work, and in tracking their own progress.

Items that connected to these sections were imbedded within the three sections of the survey and included items 13, 22, 23, 24, and 31. Items 23 and 24 aligned and had a reliability of .641 and scores with a two-item mean 2.686 (“Infrequently” to “Sometimes”). Table 14 shows the descriptive statistical breakdown for the five survey items that addressed Research Sub-question #4.

Item 13: “To what extent are models of students work (exemplars) shared with and evaluated by students?” This item was rated by 311 participants, and the mean of the scores was 3.29 (“Sometimes” to “Frequently”). When ANOVA was run, no statistical significance was identified in the independent variables.

Males reported scores with a mean of 3.25, while females reported scores with a mean of 3.29. Teachers age 40-49 reported scores with a mean of 3.43. Teachers age 50- 59

Table 14

Survey Section IV—Student Engagement Descriptive Statistics—Items 13, 22-24, 31

Survey Question	N	Mean	Mode	SD	Variance
13. To what extent are models of students work (exemplars) shared with and evaluated by students?	311	3.29	3	.856	.732
22. To what extent are scoring criteria shared with students?	311	4.53	5	.626	.392
23. To what extent are students involved with the creation of scoring criteria for their work?	312	2.43	2	.912	.831
24. To what extent are students involved with the scoring and evaluation of their own work?	312	2.94	3	.756	.572
31. To what extent are students responsible for tracking their own performance?	310	3.2	3	1.146	1.313

Note: Student Engagement Likert Scale: 1 = Never, 2 = Infrequently, 3 = Sometimes, 4 = Frequently, and 5 = All of the time

reported scores with a mean of 3.28, while teachers age 30-39 and 60+ reported scores with a mean of 3.26. Those in the youngest age group, 20-29, reported scores with a mean of 3.10.

Teachers who had taught for 11-20 years reported scores with a mean of 3.45, while teachers who had taught 6-10 years reported scores with a mean of 3.29. Teachers had taught 1-5 years reported scores with a mean of 3.04. Teachers with master's degrees or higher reported scores with a mean of 3.36, while teachers with degrees and coursework below the master's level reported scores with a mean of 3.19.

Teachers holding an assessment endorsement reported scores with a mean of 3.50, while teachers who participated in Education Service Unit training on assessment and those who participated in district level training reported scores with a mean of 3.36. Teachers who indicated they assisted in writing district level assessments reported scores with a mean of 3.29, while those who indicated receiving no assessment level training reported scores with a mean of 3.00.

Item 22: "To what extent are scoring criteria shared with students?" This item was rated by 311 participants, and the mean of the scores was 4.53 ("Frequently" to "All of the time"). When ANOVA was run, no statistical significance was identified in the independent variables.

Males reported scores with a mean of 4.47, while females reported scores with a mean of 4.54. Teachers age 20-29 and 50-59 reported scores with a mean of 4.57, while teachers age 30-39, 40-49, and 60+ reported scores with means of 4.52, 4.51, and 4.46, respectively.

Teachers who had taught 16-20 years reported scores with a mean of 4.68, while teachers who had taught over 25 years reported scores with a mean of 4.59. Teachers

who had taught 6-10 years reported scores with a mean of 4.36. Teachers with a master's degree or higher reported scores with a mean of 4.56, while those teachers with degrees below a master's degree reported scores with a mean of 4.50.

Teachers holding assessment endorsements reported scores with a mean of 4.50, while those with ESU training and those with district training reported scores with a mean of 4.52 and 4.59 respectively. Those with no formal training reported scores with a mean of 4.39, while those that participated in writing assessments at their district level reported scores with a mean of 4.53.

Items 23-24: “To what extent are students involved with the creation of scoring criteria for their work?” and “To what extent are students involved with the scoring and evaluation of their own work?” These two items were grouped together to describe the behaviors of teachers regarding student involvement in creating and using evaluation criteria for their work. These two items had a coefficient alpha of .641. These items were rated by 312 participants, and the two item mean of the scores was 2.69 (“Infrequently” to “Sometimes”). When ANOVA was run, no statistical significance was identified in the independent variables.

On items 23-24, males reported scores with a mean of 2.38, while females reported scores with a mean of 2.90. Teachers age 50-59 reported scores with a mean of 2.54, while teachers age 30-39 reported scores with a mean of 2.81. Teachers who had taught 6-10 years reported scores with a mean of 2.92, while teachers who had taught 25+ years reported scores with a mean of 2.52.

On items 23-24, teachers with a master's degree or higher reported scores with a mean of 2.69, while teachers with degrees below a master's degree reported scores with a

mean of 2.68. Teachers who indicated having some assessment training (whether through the Assessment cohort, ESUs, or district training) reported scores with a mean of 2.71, while those that had no formal training reported scores with a mean of 2.46.

Item 31: “To what extent are students responsible for tracking their own performance?” This item was rated by 310 participants, and the mean of the scores was 3.20 (“Sometimes” to “Frequently”). This item provided information that teachers surveyed have varying levels of use of students’ tracking their own performance. When ANOVA was run, no statistical significance was identified in the independent variables.

Males reported scores with a mean of 3.33, while females reported scores with a mean of 3.16. Teachers age 20-29 reported scores with a mean of 2.96, while teachers age 60+ reported scores with a mean of 3.46. Teachers who had taught 16-20 years reported scores with a mean of 3.53, while teachers who had taught 6-10 years reported scores with a mean of 2.86. Teachers with degrees lower than a master’s degree reported scores with a mean of 3.22, while teacher with a master’s degree or higher reported scores with a mean of 3.18.

Teachers holding an assessment endorsement reported scores with a mean of 2.38, while teachers with ESU training reported scores with a mean of 3.23. Teachers with district level assessment training reported scores with a mean of 3.24, while teachers that indicated no formal assessment training reported scores with a mean of 4-.14.

Summary

Chapter 4 presented the results of the study, addressing all four research sub-questions. Demographic data as well as descriptive statistics were used to analyze the responses of 312 teacher participants to the quantitative survey conducted in March 2010.

The researcher focused specifically on the results reported by 7th -12th grade English/Language Arts teachers in rural Nebraska classrooms. An unexpected finding of the study was that teachers who are younger and those with less experience do not, as a general rule, utilize the components of standards-based grading any more than those teachers who are older or have more experience. The data shows that there are a number of components considered in the grading practices of rural 7th-12th grade English/Language Arts teachers in Nebraska. From the four components of standards-based grading, it appears that there are some components being utilized more than others. Teachers in rural Nebraska 7th-12th grade English/language arts classrooms indicated frequent use of standards in the development of their course objectives, but reported less frequent use of standards when actually reporting student grades. In addition, a number of participants reported including both formative and summative assessment results in student grades, while results indicated fewer teachers still using zeros and averaging student scores to achieve a final mark. In addition, teachers with various levels of training in assessment literacy and educational backgrounds reported results ranging from “Infrequently” to “Frequently” when considering the use of effort as a grading criteria, while attendance and behavior were found across all participants to hold less weight when calculating student grades. Finally, the inclusion of students in assessment and grading practices is focused heavily on sharing exemplars with students with less emphasis on students actually monitoring their own progress by evaluating and tracking results.

CHAPTER 5

Conclusions, Recommendations, and Future Research

Introduction

The purpose of this quantitative survey study was to determine whether or not high school English/Language Arts teachers use standards-based grading in place of traditional grading practices. The overarching research question was: How and to what degree are rural 7th-12th grade English/language arts teachers in Nebraska using standards-based grading practices in their classrooms?

The sub-questions for this study were based on the elements necessary for standards-based grading to be in place in rural 7th-12th grade English/language arts classrooms in Nebraska.

1. What learning standards are included in grading by teachers and how do learning standards support standards-based grading?
2. What methods of assessment and re-assessment are included by teachers and how does assessment support standards-based grading?
3. What markers of academic achievement are included by teachers and how do these markers support standards-based grading?
4. What methods of learner engagement are used by teachers and how does learner engagement support standards-based grading?

The findings of this study provide a snapshot of the grading practices and behaviors of 7-12th grade English/Language Arts classroom teachers in rural Nebraska. The components included in determining a student's grade were investigated in three categories including the use of clear learning standards, the various aspects of assessment

and measurement, and the use of academic versus non-academic factors. Student involvement in standards-based grading permeates all three of these components. The discussion will be presented by research question.

Discussion

Research Sub-Question #1: “What learning standards are included in grading by teachers and how do learning standards support standards-based grading?” This research question was addressed by five specific items on the quantitative survey. Teachers were asked about their use of standards to guide their course objectives, daily lessons, student grades, and their use of standards to inform students of clear targets. In addition, participants were asked to identify if their current report card provided information about student performance on state or local standards. The average response for the item regarding the use of standards to determine course objectives was “Frequently.” This would align with the work that the Nebraska teachers have completed over the past ten years in the work with the STARS assessment system. More interestingly, however, was that the average began to decrease as the use of standards became more “real” or “accountable.” Moving from listing a standard as a key course objective to using it on a daily basis to hold students accountable, the average did decrease slightly, but maintained a status of “Frequently” being used. When asked if standards were used to determine students’ grades, the mean again decreased to 3.33, indicating that the mean answer was “Sometimes,” although the mode response was “Frequently.” This would be reflective of the results of the constructed response item #14, where nearly 80% of the survey participants indicated that student performance on standards was not part of their school’s report card.

Finally, participants were asked “To what extent are models of students work (exemplars) shared with and evaluated by students?” The mean of 3.29 and mode of 3 reflects most people responded “Sometimes” on this item. The results of this section of items indicates that teachers know the standards, have begun to use the standards to guide their lessons and expectations for students, but do not always use those standards or targets as the basis for student grading and reporting results on student achievement in their classrooms.

On this research sub-question as a whole, no specific subgroups reported scores that had statistical significance over other subgroups. Females did report scores that averaged “Frequently,” while males reported scores that averaged “Sometimes” to “Frequently.” One group to note for this research question is the group having taught for 6-10 years. These teachers reported scores with high means on using standards for their course objectives, their daily lesson planning, and ultimately figuring students’ grades. Research by Stiggins et al. (1989) suggested that the variance denoted here could be for the very reasons shifting to standards-based grading is so arduous—because best practice may be a matter of opinion, that some recommended practices do not take some of the practical aspects of teaching into account, and that even the notion that teachers lack training or expertise in sound practices is difficult to acknowledge.

Research Sub-Question 2: “What methods of assessment and re-assessment are included by teachers and how does assessment support standards-based grading?” This research question was addressed by seven specific items on the quantitative survey. Teachers were asked about their use of formative and summative assessments in determining grades, as well as their use of the assessment information to

teach and reteach skills and content to mastery and to adjust instruction. In addition to the above focuses, participants were also asked to identify the frequency with which they average scores to determine a grade as well as the extent to which they shared their scoring criteria with students.

In responding to several items in this section, participants provided responses that did not align with the expectations of standards-based grading. Average responses were “Sometimes” to “Frequently” regarding the use of formative assessments in grading, while responses for the use of summative assessment as a component of grading more closely aligned with expectations for standards-based grading with the average nearing “Frequently.”

In results similar to those of sub-question one, younger teachers reported scores with higher means for the inclusion of formative assessment in a student’s grade, while older teachers and those with more experience and higher education reported scores with lower means for the inclusion of formative assessment results in a student’s grade. Results were fairly comparable among all of the groups who experienced various assessment trainings or no formal training at all, with all of those groups reporting that they “Sometimes” to “Frequently” included formative assessment in students’ grades. Research indicates that formative assessment is still being used, but rationale for its use is not as understood. In traditional classroom settings, the grading aspect of assessment is overemphasized and the learning or improvement purpose of assessment is underemphasized (Black & Wiliam, 1998b).

All of the demographic sub-groups marked “Sometimes” to “Frequently” as the average result for using summative assessment results in students’ grades. Although not

statistically significant, teachers with master's degrees reported scores with higher means than those reported by teachers without a master's degree. One may argue that additional training can provide the background and understanding needed to shift grading practices from using formative, or practice, work for grading to that of assessment of learning to mark a student's grade. Again, although not statistically significant, the responses of teachers who reported having no formal assessment training had a mean score of "Frequently" on this item, comparable only to those who possessed an Assessment Endorsement, refuting the theory that one would have to have additional education or training to fully internalize the shift in the use of formative and summative assessments in grading. In terms of adjusting instruction based on student assessment results, results were varied, leading the researcher to conclude that no one subgroup stood out as expert in the practice of teaching and reteaching to mastery. Those with less formal graduate work indicated that they focused more on teaching and reteaching to mastery, while teachers with 21-25 years of experience indicated that they only "Sometimes" to "Frequently" taught and retaught skills/concepts and adjusted instruction based on assessment results.

In addition to the use of formative and summative assessment to reteach and refocus instruction, the final components of sub-question 2 included the use of zeros and averaging student performance to arrive at a student's grade. Overall, participants indicated that they "Infrequently" to "Sometimes" used zeros and "Sometimes" to "Frequently" used averaging scores to determine a grade, both of which elements are considered a negative attribute in standards-based grading. The youngest teachers reported scores that reflected less use of averaging student grades as compared to older

teachers. Older teachers reported scores that could be considered “Frequently” averaging of grades, while teachers under 30 indicated that they “Infrequently” to “Sometimes” average scores to determine grades. This seems to indicate to the researcher that the grading practices of a classroom teacher may be, in fact, directly connected as research suggests, to the practices to which teachers were subjected to. Guskey and Bailey (2001) described four factors that impact how a teacher determines grades. First and foremost, they refer to the policies and practices they experienced as students. They do what was done to them. If averaging has been a standard practice for the last century, then it would stand to reason that these results are accurate to current reality.

Research Sub-Question 3: “What markers of academic achievement are included by teachers and how do these markers support standards-based grading?”

This research question was addressed in seven specific items on the quantitative survey. Teachers were asked to what extent they included academic achievement, effort, behavior, and attendance in their calculation of students’ grades. In addition, teachers were asked to what extent they included extra credit in figuring a student’s grade as well as the extent to which they graded students as a group on group work. Finally, participants were asked to describe their school’s common set of grading criteria, if one in fact was in place.

Teachers from all of the subgroups reported that they used academic achievement “Frequently.” Although not statistically significant, those that had no formal assessment training actually reported scores that indicated more frequent use of academic achievement in determining grades than did those holding an assessment endorsement.

Participants indicated largely that they “Sometimes” included effort as a grading criterion, while the use of behavior results and attendance were less likely to become part of a student’s grade. This finding is supported by a 1996 McREL study cited by Marzano (2000). In that study, thirty-six percent of teachers of 7th-9th grade used effort, while only 10% (same grades) used behavior.

Teachers from various subgroups reported using effort, attendance, and behavior as elements in figuring a student’s grade to various degrees. Most participants reported using extra credit and group grades “Infrequently.”

Research Sub-Question 4: “What methods of learner engagement are used by teachers and how does learner engagement support standards-based grading?”

This research question was addressed by five specific items on the quantitative survey. Teachers were asked about the extent to which they shared the learning targets and the scoring criteria with students, the extent that they involved students in the creation of and execution of those scoring criteria, and the degree to which they held students accountable for tracking their own academic performance.

Participants indicated that they “Sometimes” shared exemplars with students, with the teachers with the least experience reporting that they used this practice less than those who had been teaching longer. Sharing the scoring criteria with students was reported as occurring between “Frequently” and “All of the time,” which may indicate that the practice of using the statewide writing assessment rubric, for example, may be spilling over into general practice. Teachers are sharing criteria for the work that is expected of the students. However, students are not the ones that appear to be establishing the criteria and/or evaluating their own work, which is counter to best practice. McMillan (2009)

and Sadler and Good (2006) suggested that self-evaluation can, in fact, create deeper understanding of the material being learned as well as foster learning independence. Results show that most of the work done by students for the development of and use of student-developed criteria occurs “Infrequently” to “Sometimes.” In addition to participating in the development and scoring of their own work, scores indicated that students are also only “Sometimes” held accountable for tracking their own performance, which again is counter to best practice.

Recommendations

Results of this quantitative study may be of particular interest to classroom teachers, building-level leaders, curriculum and assessment directors, as well as professors of higher education and other educational leaders interested in grading methodology as steps should be taken to use standards as the basis for grading.

Recommendation One. Based on the results of the study, it appears evident that the issuing of grades is currently based on teacher professional judgment in most cases, regardless of training, age, experience, or gender. Even in cases where the school holds a common set of grading criteria, it appears that the criteria for one school is not necessarily the same for another school. The assessment literacy work of a school district should begin to include the discussions of what it means to achieve certain grades/marks in a given school district. Part of the work of professional learning communities is to provide the professional development necessary to embark on the journey toward standards-based grading, including identifying the clear targets, developing quality assessments, and then determining what criteria or components are utilized when determining a student’s grade. Walker (2006) suggested that faculty should hold critical

conversations and come to consensus about the purpose of grades, factors to include when grading, the role of zeros in grading, ways to provide meaningful feedback to students, meaningful assessment practices, and intervention for struggling students. Developing common criteria for grading will not completely eliminate teacher subjectivity or bias in grading, but it will support consistency from teacher to teacher, course to course, and subject to subject. This legitimizes the credibility of the school district's rigor and underscores true academic achievement.

Beyond individual school district's determining common sets of grading criteria, the state could play off of this idea and work to determine some common reporting measures that are easily accessible to colleges and universities to level the playing field of what grades mean. As suggested by Burke (2005), benefits of having uniform grading scales locally as well as at the state level include comparability of student achievement across districts, providing equal access and opportunity when awarding various scholarships across the state, and allowing students who move between district to have a common expectation from numeric grades. The research conducted by the Southeast Comprehensive Center further supports this divide between and among states' uniformity when it comes to student grading practices (2009).

Recommendation Two. Based on the results of the study, it is apparent that additional time and support must be allocated to the work of standards, assessment, and accountability in pre-service teacher education. Institutions of higher education must develop additional course requirements for pre-service and even graduate students in education to develop stronger assessment literacy as well as create more opportunities for these students to learn about standards-based grading practices and procedures. Methods

courses must focus on components of standards-based learning in order to produce graduates that can implement quality grading practices in their classrooms, and possibly even challenge the status quo of the districts' grading practices when they become teachers in a district. The research of Stiggins et al. (1989) supported the notion that teachers lack training and/or expertise in sound assessment and grading practices as a possible reason for this lethargic movement to embrace standards-based grading in practice.

Recommendation Three. A final recommendation would be to focus attention on training teachers on ways to include students in the grading process, from identifying clear targets and exemplars to developing the criteria for measuring their own work. Numerous researchers and educational experts have suggested that including students in the process of grading is critical to students actually learning and internalizing the work that schools have them do. When students understand what is expected of them and have opportunities to measure their own performance against those standards, true learning occurs. According to several researchers (McMillan, 2009; O'Connor, 2007; Sadler and Good, 2006; Stiggins et al., 2004), providing students with clear objectives and expectations, having students evaluate and track their own academic progress and achievement and then communicate about their learning with others are some of the most powerful tools in improving student achievement. It resonates with the practices of the ancient Greeks, who provided feedback to young Olympians expecting them to learn from the feedback and change their performance to mirror that of excellence. Additional training and support for teachers will make this paradigm shift more of a reality.

Future Research

The results of this study yielded a variety of key findings that are crucial to the work of classrooms, schools, districts, educational service units, post-secondary teacher preparation programs, and the state. In addition, teachers from rural 7-12th grade classrooms involved in the study were given an opportunity to reflect on their current grading practices. Although findings of this study indicated that teachers consider and use a variety of components when determining student grades, there is still need for future research in several areas.

Although there have been studies conducted to identify statewide scoring criteria, further study is needed in Nebraska, across the United States, and even internationally. The same study conducted by this researcher could be conducted in other states to see if there are differences in grading practices across the nation, which this researcher doubts.

It is recommended that a study be conducted in Nebraska's urban schools using a similar instrument to survey 7th-12th grade English/Language Arts teachers in those classrooms. This type of study might yield different results than those of the rural districts' classrooms described in this study. Research conducted by Isernhagen et al. (2009) concluded that rural language arts teachers may not be using principles of sound grading practices as much as their other subject area counterparts are in urban districts. Further investigation into this phenomenon is strongly encouraged.

An additional study could also be conducted using qualitative research methods. This method would allow teachers to have conversations about their own grading practices and allow the researcher to probe deeply into the beliefs about grading rather than the quantitative methodology used in this study. This may allow the researcher to

identify the professional development needs of the teachers in an attempt to develop stronger standards-based grading practices throughout all classrooms. A study like this could also collect more specific data related to common grading scales used within schools, researching the development and actual implementation of and accountability for using those scales.

Another focus for future study could be to investigate the preparation and training of pre-service teachers in the areas of standards, assessment and grading. Little research has been conducted focusing on the programming and outcomes for pre-service teachers on the use of standards in lesson preparation, the development of appropriate assessment practices, and the accurate and effective communication of student achievement through grades. An extension of this study could provide additional insight into the actual professional development opportunities provided to not only pre-service teachers, but also practicing classroom teachers.

A final recommendation for future study revolves around the shift in testing focus in the state of Nebraska. As the state takes on a common, statewide test in the area of reading, researchers could examine the results of the state test in comparison to the grades students achieve in the classroom to look for a correlation. Beginning in 2010-2011, students' grades in the core subject of English/Language Arts at the high school level will be reported through the Nebraska Student and Staff Record System (NSSRS), as will the results of the Nebraska State Accountability test for Reading (NeSA-R).

Summary

Grading practices of secondary teachers have been written about, studied, critiqued and reviewed for as long as a public education system has been alive in the

United States. Based on the research, it is clear that teacher judgment has long been the primary focus of the criticism and concern with grading. Based on the work of Stiggins (1989, 2004), O'Connor (2002, 2007), Guskey (1996, 2000, 2001), Reeves (2002, 2004) and others, it is clear that a new set of criteria for best representing student achievement is available to teachers, but the best method to delivering that message to the teaching ranks is yet to be determined. Old habits die hard, so assisting every educator in the transition from traditional grading practices to those of a standards-based system will take a concerted effort with clear and effective professional development at the ready. The state of Nebraska is ripe for this transition, with its history of formative and summative assessment development and the increasing implementation of professional learning communities in its districts. This study revealed that a number of teachers are already meeting many of the criteria of standards-based grading head-on. Addressing best practice in grading and creating systems that allow for standards-based grading will continue to drive this movement.

References

- Associated Press. "Grand Island schools consider grades based on proficiency." *Lincoln Journal Star* (Lincoln, NE), Oct. 1, 2006.
- Black, P. & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-75.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 10, 139-148.
- Brookhart, S. M. (2004). *Grading*. Upper Saddle River, NJ: Pearson Education, Inc.
- Burke, M. (2005). *ECS StateNotes: Student achievement: Statewide uniform grading scales*. Lincoln, IL. Retrieved August 14, 2010, from <http://www.ecs.org/clearinghouse/64/12/6412.pdf>.
- Busick, K. (2000). Grading and standards-based assessment. In E. Trubmull & B. Farr (Eds.), *Grading and reporting student progress in an age of standards* (pp. 71-86). Norwood, MA: Christopher Gordon.
- Carr, J., & Artman, E. M. (2002). *The bottom-up simple approach to school accountability and improvement*. Norwood, MA: Christopher-Gordon.
- Carr, J., & Farr, B. (2000). Taking steps toward standards-based report cards. In *Grading and reporting student progress in an age of standards* (pp. 185-208). Norwood, MA: Christopher-Gordon.
- Creswell, J. W. (2005). *Educational research: Planning, conducting, and evaluation quantitative and qualitative research* (2nd ed.). Upper Saddle River, NJ: Pearson Prentice Hall.

- Creswell, J. W. (2009). *Research design: Qualitative, quantitative, and mixed methods approaches*. Thousand Oaks, CA: Sage.
- Davies, A. (2000). *Making classroom assessment work*. British Columbia: Connections.
- DuFour, R., DuFour, R., & Eaker, R. (2008). *Revisiting professional learning communities at work*. Bloomington, IN: Solution Tree.
- Guskey, T. R. (Ed.). (1996). *Communicating student learning: The ASCD yearbook 1996*. Alexandria, VA: ASCD.
- Guskey, T. R. (2000). Grading policies that work against standards . . . and how to fix them. *NASSP Bulletin*, 84(620), 20-29.
- Guskey, T. R. (2001). Helping standards make the grade. *Educational Leadership*, 59(1), 20-27.
- Guskey, T. R., & Bailey, J. M. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin.
- Haladyna, T. M. (1999). *A complete guide to student grading*. Needham Heights, MA: Allyn & Bacon.
- Isernhagen, J., Dappen, L., & Mills, S. (Eds.). (2006). *Charting STARS: Portraits of excellence. Year Five Report Comprehensive Evaluation of Nebraska's School-based Teacher-led Assessment and Reporting System*. Lincoln, NE: University of Nebraska-Lincoln.
- Isernhagen, J., & Mills, S. (Eds.). (2007). *Charting STARS: Pathways to success. Year Six Report Comprehensive Evaluation of Nebraska's School-based Teacher-led*

Assessment and Reporting System. Lincoln, NE: University of Nebraska-Lincoln.

Isernhagen, J., Florendo, J., & Guerrero, C. (2009). Administrator and teacher perceptions of a balanced assessment system. In J. Isernhagen & S. Mills (Eds.), *Transitions . . . A journey to a balanced assessment system. Year Eight Report Comprehensive Evaluation of Nebraska's School-based Teacher-led Assessment and Reporting System* (pp. 47-92). Lincoln, NE: University of Nebraska-Lincoln.

Kagan, S. (1995). Group grades miss the mark. *Educational Leadership*, 52(8), 68-71.

Kohn, A. (1993). *Punished by rewards: The trouble with gold stars, incentive plans, A's, praise, and other bribes*. Boston: Houghton Mifflin.

Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: ASCD.

McMillan, J. H. (2009). Synthesis of issues and implications for practice. In T. R. Guskey (Ed.), *Practical solutions for serious problems in standards-based grading* (pp. 105-117). Thousand Oaks, CA: Corwin.

Milton, O., Pollio, H. R., & Eison, J. A. (1986). *Making sense of college grades*. San Francisco: Jossey-Bass.

No Child Left Behind Act, 20 U.S.C. 107-110 (2002).

Olson, K. E. (2005). Standards-based report cards: Accountability at the classroom level. (Doctoral dissertation, University of Southern California, California.)
Dissertations & Theses: A&I database. (Publication No. AAT 3196865).

O'Connor, K. (2002). *How to grade for learning: Linking grades to standards*. Thousand Oaks, CA: Sage.

- O'Connor, K. (2007). *A repair kit for grading: 15 fixes for broken grades*. Portland, OR: Educational Testing Service.
- O'Shea, M. (2005). *From standards to success*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Popham, J. W. (2008). *Transformative assessment*. Virginia: ASCD.
- Reeve, J. (2006). Extrinsic rewards and inner motivation. In C. Evertson, C. M. Weinstein, & C.S. Weinstein (Eds.), *Handbook of classroom management: Research, practice, and contemporary issues* (pp. 645-664). Mahwah, NJ: Erlbaum.
- Reeves, D. (2000). Standards are not enough: Essential transformations for school success. *NASSP Bulletin*, 84(620), 5-19.
- Reeves, D. (2002). *Holistic accountability: Serving students, schools, and community*. Alexandria, VA: ASCD.
- Reeves, D. (2004). *101 more questions about standards, assessment, and accountability*. Englewood, CO: Advanced Learning.
- Roorda, N. L. (2008). The impact of professional development on assessment and grading practices for secondary teachers. (Doctoral dissertation, Walden University.) Dissertations & Theses: A&I database. (Publication No. AAT 3342447).
- Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment*, 11(1), 1-31.
- Seeley, M. M. (1994). The mismatch between assessment and grading. *Educational Leadership*, 52(2), 4-6.

- Southeast Comprehensive Center. (2009). *Rapid response—Statewide uniform grading practices*. Metairie, LA: Southeast Comprehensive Center. Retrieved on August 14, 2010, from http://secc.sedl.org/orc/rr/secc_rr_00094.pdf
- Stiggins, R., Arter, J., Chappius, J., & Chappius, S. (2004). *Classroom assessment for student learning*. Portland, OR: ATI.
- Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school: Building a research agenda. *Educational Measurement: Issues and Practice*, 8, 5-14.
- Tucker, M., & Coddling, J. (1998a). Raising our standards for the standards movement. *Education Week*, 17(23), 56-58.
- Tucker, M., & Coddling, J. (1998b). *Standards for our schools: How to set them, measure them, and reach them*. San Francisco: Jossey-Bass.
- Tyack, D. & Tobin, W. (1994). The "Grammar" of schooling: Why has it been so hard to change? *American Educational Research Journal*, 31 (3), 453-79.
- United States Census Bureau. (n.d.). *Urban and rural classification*. Retrieved on November 23, 2009, from http://www.census.gov/geo/www/ua/ua_2k.html
- Walker, K. (2006). Role of zero in grading. Omaha, NE: The Principals Partnership, a service of Union Pacific Foundation. Retrieved on September 12, 2010 from <http://www.principalspartnership.com/zerograding.pdf>.
- Wormeli, R. (2006). *Fair isn't always equal: Assessing and grading in the differentiated classroom*. Portland, ME: Stenhouse.

Appendix A

Survey Participant Letter



COLLEGE OF EDUCATION AND HUMAN SCIENCES
Department of Educational Administration

Dear Educator,

I am writing to ask your help in a research study being conducted of 7th-12th Grade English/Language Arts teachers in rural Nebraska. This study is part of an effort to learn more about grading practices including: teachers' familiarity with different assessment practices, familiarity with standards-based grading, and implementation of standards-based.

You were selected as someone who currently teaches English/language arts to students in grades ranging 7th-12th. You are being asked to complete the attached survey, which should not take more than 15-20 minutes of your time. Please return the survey in the envelope provided. Results from this survey will be used in a dissertation examining the use of standards-based grading in Nebraska English/Language Arts classrooms and will further the study of Nebraska's balanced assessment system.

Your participation in this study is vital. Therefore, we want you to know that your information will remain anonymous, and your contact information will be kept private and secure. Survey data will be presented in aggregate form or will be shared with no personal identifying information attached; your individual responses will never be revealed to anyone but the researcher. Your participation is completely voluntary. You are free to decide not to participate in this study. You can also withdraw at any time without harming your relationship with the researchers or the University of Nebraska-Lincoln. There are no known risks associated with this research.

If you agree to participate in this study, please complete the enclosed survey and return to the researcher. Returning a completed survey implies consent to participate. If you have questions about this study, you may contact me, Summer Stephens at (402) 770-1781, or my advisor, Jody Isernhagen at (402) 472-1088. If you have questions concerning your rights as a research subject that have not been answered by the investigators or to report any concerns about the study, you may contact the University of Nebraska-Lincoln Institutional Review Board at (402) 472-6965.

Thank you very much for helping with this important study.

Summer Stephens, Ed.D. Candidate
Primary Research Investigator
1225 Woodland Ave., Beatrice, NE 68310
Cell: (402) 770-1781
Email: bsstephens@windstream.net

Jody Isernhagen, Ed.D
Secondary Investigator
132 TEAC Hall, Lincoln, NE 68588-0360
Office: (402) 472-1088
Email: jisernhagen3@unl.edu

Appendix B

Survey Instrument

Grading Practices of 7th-12th Grade English/language arts teachers in rural Nebraska

Overview

The purpose of this survey is to capture practices and perceptions of 7th-12th grade English/language arts teachers relevant to grading of student achievement.

Section I: Demographics

1. Please circle your gender: Male Female

2. Please indicate your age: _____

3. Please indicate the number of years you've been a classroom teacher of grades ranging from 7th-12th: _____

4. Please indicate your highest level of education:
 ____ Bachelor's degree ____ Bachelor's plus graduate level coursework
 ____ Master's degree ____ Master's degree plus other graduate level coursework
 ____ Doctorate degree

5. Please list the endorsement(s) on your Nebraska teaching and/or administrative certificate:

6. Which of the following describes your current level of assessment literacy? Check all that apply:
 _____ I have a Nebraska Assessment Endorsement
 _____ I have attended Educational Service Unit trainings on assessment
 _____ I have attended district level training on assessment
 _____ I do not have any assessment training
 _____ I have participated in developing criterion-referenced assessments in a school district

7. Please indicate the grades that you currently teach. Check all that apply:
 _____ 7 _____ 8 _____ 9 _____ 10 _____ 11 _____ 12

8. Please indicate the number of students in your school district: _____

9. Please indicate the grade configuration that best describes your school building:
 _____ K-12 _____ 7-12 _____ 9-12 _____ 6-8 Other: _____

Section II: Learning Standards

10. To what extent do you use local or state standards as a basis for your course objectives?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

11. To what extent do you use local or state standards as a basis for your lessons?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

12. To what extent do you use local or state standards as a basis for your students' grades?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

13. To what extent are models of students work (exemplars) shared with and evaluated by students?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

14. Does your current report card reflect student performance on local or state standards? If so, please describe the format (use the back page if needed)._____

Section III: Assessment Practices

15. To what extent is formative assessment, such as homework or practice, used for the basis of grades?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

16. To what extent are summative assessments used for the basis of grades?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

17. To what extent are concepts taught and retaught to mastery?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

18. To what extent are skills taught and retaught to mastery?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

19. To what extent are student assessment results used to adjust, improve, or support instruction?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

20. To what extent do you assign zeros for assigned work?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

Section III: Assessment Practices (cont.)

21. To what extent do you average scores to assign a grade?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

22. To what extent are scoring criteria shared with students?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

23. To what extent are students involved with the creation of scoring criteria for their work?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

24. To what extent are students involved with the scoring and evaluation of their own work?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

Section IV: Markers of Academic Achievement

25. To what extent do you include academic achievement as the basis for grades?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

26. To what extent do you include effort as the basis for grades?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

27. To what extent do you include behavior as the basis for grades?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

28. To what extent do you include attendance as the basis for grades?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

29. To what extent do you include extra credit as the basis for grades?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

30. To what extent do you give group grades for assignments?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

31. To what extent are students responsible for tracking their own performance?

Never	Infrequently	Sometimes	Frequently	All of the time
1	2	3	4	5

32. Does your school have a common set of grading criteria that applies to all subjects, such as a common grading scale? If so, please describe the criteria and/or the common grading scale (use the back as needed)
