

THE POLITICS OF GRADING: A COMPARATIVE STUDY OF HIGH SCHOOL ENGLISH TEACHERS'

PERSONAL BELIEFS, SELF-REPORTED SYSTEMS, AND ACTUAL PRACTICES

Lisa M. Thibodeaux

Dissertation Prepared for the Degree of

DOCTOR OF EDUCATION

UNIVERSITY OF NORTH TEXAS

May 2013

APPROVED:

Carol Wickstrom, Major Professor

James Laney, Committee Member

Kelley King, Committee Member

Javier Rodriguez, Committee Member

Nancy Nelson, Chair of the Department of  
Teacher Education and Administration

Jerry Thomas, Dean of the College of  
Education

Mark Wardell, Dean of the Toulouse Graduate  
School

Thibodeaux, Lisa M. The Politics of Grading: A Comparative Study of High School English Teachers' Personal Beliefs, Self-Reported Systems, and Actual Practices. Doctor of Education (Curriculum and Instruction), May 2013, 209 pp., 35 tables, 8 figures, references, 88 titles.

The purpose of this study was to attain and analyze data regarding high school English teachers' beliefs about grading practices and their self-reported grading practices, to identify and understand disparities that exist between teachers' beliefs and self-reported practices, to identify discrepancies between the same self-reported practices and evidence of the practices actually utilized, and to consider teachers' perceptions as to the causes for these discrepancies.

Instrumentation for this study included two surveys with both Likert and Likert-like items and an interview/portfolio analysis of teachers' grading systems. A combined total of 204 high school English-language arts teachers representing thirty-eight states and eighty-five schools comprised the sample.

Corresponding pairs of Likert-type items were analyzed using studies of the mode, median, mean rank, and the Mann-Whitney U Test to study a comparison of the medians, and comparisons of true Likert scale item results were completed using studies of the means and an independent samples t-test. Interview/portfolio analysis data were analyzed both descriptively and inferentially including the calculation of 95% confidence intervals for generalizability. All open-ended items were considered qualitatively through a process of identifying and categorizing trends in language and over-arching themes.

Results indicate that the sample finds grading practices recommended by experts in the field to be best grading practices, and respondents generally report the use of these same practices in their own grading systems. The data reveal, however, discrepancies between the

majority of teachers' reported practices and their actual practices. Study participants are likely to place blame for these discrepancies on these sources: campus or district authorities, the limited time available, and the interferences caused by parents.

Copyright 2013

by

Lisa M. Thibodeaux

## ACKNOWLEDGEMENTS

My most sincere gratitude is expressed to Dr. Carol Wickstrom for her wisdom and spirit; her encouragement and guidance were both supportive *and* demanding, and her love for literacy and learning inspired me throughout this process. Likewise, thanks is given to Dr. Leslie Patterson who found a “home” for me and my work when all felt lost. I also wish to acknowledge the guidance of my committee members: Dr. James Laney, Dr. Kelley King, and Dr. Javier Rodriguez. I am grateful for their time, their insights, and their unique and meaningful contributions.

I’d like to also acknowledge the contributions of Dr. Paul Dabbs, whose single question about methodologies prompted a significant change in my work, and Drs. Arthur Applebee and Judith Langer whose presentation at the 2011 National Council of Teachers of English Annual Convention inspired the three-part construct that guided my research.

Endless appreciation is given to Dr. Jim Wussow, a brilliant educator and inspirational leader, and to my professional mentor and dear friend, Susan Modisette, whose unmatched spirit and encouragement contributed to my completion of this program.

I am grateful, beyond words, for the daily love and support I received from my exceptionally talented husband, Troy, and my remarkable daughters, Katherine and Avery, who seemed to believe in me perhaps more than I believed in myself.

And, of course, I am ever-appreciative to my parents, Thurston and Dorothea, who are the most courageous, hard-working, and compassionate people I’ve ever met. I’m grateful to have learned so much from them both about commitment and contribution.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS.....	iii
LIST OF TABLES.....	v
LIST OF FIGURES.....	vii
CONSIDERING THE POLITICS OF GRADING: INTERFERENCES TO INTENDED AND EFFECTIVE PRACTICES.....	1
Introduction.....	1
Methods.....	5
Results.....	14
Discussion.....	35
References.....	40
Appendices	
APPENDIX A: EXTENDED LITERATURE REVIEW.....	43
APPENDIX B: DETAILED METHODOLOGY.....	68
APPENDIX C: COMPLETE/UNABRIDGED RESULTS.....	109
APPENDIX D: SURVEY AND INTERVIEW DOCUMENTS.....	153
APPENDIX E: REQUESTS FOR PARTICIPATION.....	191
COMPREHENSIVE REFERENCE LIST.....	201

LIST OF TABLES

	Page
1.1 Demographic Profiles of Participants vs. US High School English Teachers.....	11
1.2 Participant Knowledge of Grading Experts.....	12
1.3 Best Practice Beliefs in Grading Systems, Likert-type Items .....	15
1.4 Best Practice Beliefs in Grading Systems, Likert Scale (Collective).....	16
1.5 Self-Reported Practices in Grading Systems, Likert-type Items .....	17
1.6 Self-Reported Practices in Grading Systems, Likert Scale (Collective) .....	18
1.7 Comparison of Best Practice Beliefs and Self-Reported Practices .....	21
1.8 Comparison of Best Practice Beliefs and Self-Reported Practices, Likert Scale .....	23
1.9 Interview/Portfolio Assessment Results.....	25
1.10 Identifying Discrepancies.....	31
1.11 Causes for Discrepancies .....	34
B.1 Summary of Research Questions, Strategies of Inquiry, and Instrumentation .....	72
B.2 Design of Survey 1 by Topic/Description and Research Question Addressed.....	74
B.3 Evidence of Expert Status.....	75
B.4 Components of Interview and Portfolio Assessment Survey.....	82
B.5 Design of Survey 2 by Topic/Description and Research Question Addressed.....	86
B.6 Demographic Profiles of Participants vs. U.S. English Teachers .....	93
B.7 Participant Familiarity with Professional Organizations .....	94
B.8 Participant Knowledge of Grading Experts .....	95
B.9 Levels and Courses Represented in Sample.....	97

B.10	Relationship Between Likert Scale Items and Likert-type Items in Survey 1 .....	101
B.11	Analysis Design for Survey 1: Research Questions 1 and 2.....	103
C.12	Best Practice Beliefs in Grading Systems, Likert-type Items .....	112
C.13	Best Practice Beliefs in Grading Systems, Likert Scale .....	113
C.14	Best Practice Beliefs, Assignments Graded, Likert-type Items .....	115
C.15	Self-Reported Practices, Grading Systems, Likert-type Items.....	121
C.16	Self-Reported Practices, Grading Systems, Likert Scale .....	122
C.17	Self-Reported Practices, Assignments Graded, Likert-type Items .....	124
C.18	Comparisons of Research Qs 1 and 2 Results–Likert-type Items, Best Practice .....	128
C.19	Comparisons of Research Qs 1 and 2 Results–Likert-type Items, Graded Assignments .	130
C.20	Survey 1, Part A Results – Self-Reported Practices in Grading Systems, Likert Scale .....	133
C.21	Interview/Portfolio Assessment Results – Questions 6-11 and 12-14.....	136
C.22	Identifying Discrepancies Among Results .....	145
C.23	Perceived Causes for Discrepancies .....	150
C.24	Comparisons of Versions 1.2 and 2.1, Likert Scale.....	152

## LIST OF FIGURES

	Page
1.1 Reasons for assigning “0” grades.....	27
1.2 Use of grades for reasons indicated .....	28
1.3 Control over grading decisions .....	29
C.1 Percentage of grade levels and courses in sample .....	135
C.2 Reasons for assigning “0” grades .....	138
C.3 Use of grades for reasons indicated.....	139
C.4 Control over grading decisions.....	141
C.5 Types of assignments .....	142

## CONSIDERING THE POLITICS OF GRADING: INTERFERENCES

### WITH INTENDED, EFFECTIVE PRACTICES

#### Introduction

The language around the concept of learning is rich and lively and inviting; it suggests an aesthetic and energy, an evolvment and a transformation. *Learning* evokes images of inquiry, engaging experiences, drive and purpose, shared critical deliberations, and quiet reflections.

Conversations about school, however, especially secondary school, appear to fail to echo this spirited vernacular. Chat with a high school student about school, and you are likely to hear words more like *assignments* and *tests*; or ask a student how she's "doing" in school, and the most likely response will not be about concepts of learning but about grades as in, "I'm getting a B," or "I have a 75." In truth, our school systems have become so focused on the idea of grades and grading that we have all but abandoned the truest ideal of all – that school is about *learning*.

Formal grading systems as we know them today are a rather recent phenomenon in United States education (Guskey, 2001). The one-room school houses that existed in and before the 18<sup>th</sup> century relied not on "grades" but on a system of feedback by teacher comments reported orally to students and parents (Rich, 2001). By the late 1700s, it is said that word had reached both Harvard and Yale Universities of a professor at Cambridge University in England – one William Farish – who had increased his class size and thusly increased his salary by creating a system of grading that would allow him to by-pass the narrative feedback by adopting the much more efficient system of "grading," much like the one being used in factories to measure the effectiveness of the assembly line workers (Pearson-Casanave & Sosa,

2007). Instructors at Harvard and Yale adapted this new system and respectively implemented a 4-point scale and a 100-point scale by the late 1800s (Marzano, 2000). Shortly after, Mount Holyoke College in Massachusetts converted these numerical systems into letter grades for ease of reporting (Durm, 1993).

Though built under the guise of efficiency, it wasn't long before researchers testing these new systems deemed them to be unreliable; as early as 1911, researchers Daniel Starch and Edward Charles Elliott were arguing that teachers' previously used narrative feedback responses improved student work whereas the numerical and letter grading systems merely served the purpose of sorting students and their abilities (Guskey, 2001).

Nowhere was this truer than in the English classroom where researchers found that teachers were realizing that to give an "accurate" grade, the focus of essay assessment would need to shift from content to the more objective measures such as grammar, neatness, spelling, and form. In 1913 Cornell doctoral student Isidor Edward Finkelstein published a short text entitled *The Marking System in Theory and Practice*, in which he wrote that "school administrators have been using with confidence an absolutely uncalibrated instrument..." (p. 1) and he argued that the "variability in the marks given for the same subject and to the same pupils by different instructors is so great as frequently to work real injustice to the students" (p. 6). Even at this early date, educators were questioning if "marks should indicate performance or ability or accomplishment" (p. 11).

Despite the criticisms of grade use and subsequent failed attempts to systems both efficient and effective, by the 1940s almost all large-scale educational institutions in the United States had begun using either a number or letter grading system; consequently, a 1998 study by

the College Board found that of over 3,000 contemporary U.S. schools surveyed, 91% used some sort of A-F or numeric grading system (Marzano, 2001). It is not uncommon for teachers to use and even recommend instructional methods, like grading practices, solely because such processes were in place while they, themselves, were students (Britzman, 1991). One thing that most American educators today are likely to have in common is that they were students in a system that recorded and reported grades, and they now work in a very similar system, where grades are expected to be calculated meticulously and conveyed to students and parents with confidence and an assurance of accuracy and meaning.

Though a great deal of research and theory has been published on grades and grading systems, the implementation of these recommendations is sometimes limited or altogether restricted by other power structures in the school environment, including administrator limitations and guidelines, pressures from parents, competition organization rules, state requirements, university admission procedures, and even teachers' own histories or erroneous understandings about their roles and responsibilities (Guskey, 2010). These authority constructs are part of the politics of grading. But without a better understanding of how we are actually using grades and grading, why we accept and advance certain theories of the meaning of grades, and what chasms exist between our beliefs and our practices, we cannot, as a profession, lay claim to grades as a meaningful communication of student performance or to grading practices as reliable or valid.

### *Grading in the English Classroom*

Grading in the English classroom is difficult, to say the least. Where there are occasional aims that can be taught explicitly and measured objectively, the real work of the high school

English teacher is centered around students' attitudes and dispositions toward literacy, including an appreciation for the aesthetics of writing, an ability to notice and emulate the writer's craft, the development of compelling and insightful arguments and interpretations, and an ability to make perceptive and original connections within and among texts. The act of accurately quantifying these achievements lies somewhere between demanding and incomprehensible, and yet the English teacher is asked to do this some thousand times over the course of a single school year.

Few studies exist that address grading practices in English-language arts classes, nor are studies readily found that address the politics – the authority, power, and social relations – that affect high school teachers' decisions about what, how, when, and why to grade. As grades and grading can have significant impact on students' current and future academic lives, these topics are worthy of consideration.

The intention of this work is to invite high school English teachers into a conversation around the role that grades and grading play in their classrooms and why these systems exist and are used thusly. The study is designed to (a) invite reflection among respondents regarding the grading practices they support and use, (b) to attain and analyze data about high school English teachers' beliefs and practices around grading, (c) to identify and understand disparities that may exist between teachers' beliefs and self-reported practices, and (d) to identify discrepancies between the same self-reported practices and evidence of the practices actually utilized.

A research study designed to generate data used to enhance this work was conducted around these specific research questions:

- 1 – What do high school English-language arts teachers believe a grade should measure?
- 2 – What do high school English-language arts teachers self-report their grades to measure?
- 3 – What do high school English-language arts teachers' grades actually measure?
- 4 – what do teachers perceive to be the causes for the discrepancies between what high school ELA teachers say a grade should measure and what the grades actually measure?

The theory of the social construction of knowledge exists throughout this study – specifically, the belief that teachers seeking to understand their own practices and belief systems can construct valid interpretations of the political and power situations that affect their day to day decision making. And, to an extent, an advocacy stance in this work exists that supports the rights of students to have an educational experience where grading practices at minimum do not interfere with the learning, and at best have a positive influence on the classroom experience.

## Methods

To determine the relationship among teachers' beliefs about grading, self-reported practices, and actual practices, this study was conducted using questionnaire and interview surveys collected from a sample of high school English-language arts teachers currently certified and teaching at least one high school English-language arts (required, non-elective) course.

The sequence of study was such: (1) First a determination was made as to the degree to which the sample believes that certain grading practices are, indeed, best practices, (2) concurrently, the extent to which these same members of the population reported using these ideal practices in their own classrooms was analyzed; (3) next it was determined the extent to which teachers' grading records supported these self-reports on classroom grading practice; and (4) finally, as discrepancies were found between beliefs, self-reported practices, and actual practices, a study as to teachers' perceived causes for these discrepancies was performed.

### *Instrumentation*

Because this study is grounded in matters that are directly relevant to the participants in their daily professional lives – i.e., their professional beliefs about best practice and their self-reported and actual practices – the challenge regarding the design of instrumentation was to construct a method of soliciting responses that capture participants' most honest and complete ideas while not suggesting judgment or causes feelings of distress.

Survey 1 called, Survey on High School English Teachers' Grading Practices, was designed to capture both teachers' beliefs about best grading practices and teachers' self-reports on their own grading practices. The best grading practice statements developed for this instrument were founded on recommendations drawn from these five expert sources: *Assessment for Learning* by Black, Harrison, Lee, Marshall, and Wiliam (2003); "Seven Practices for Effective Learning" by McTighe and O'Connor (2005); "Effective Grading" by Reeves (2008); "Grading for Success" by Tomlinson (2001), and *A Repair Kit for Grading: 15 Fixes for Broken Grades* by O'Connor (2007).

From these recommendations, the following fourteen general statements on grading systems were developed:

- A teacher should adjust a student's report card grade if the final computation does not accurately reflect the teacher's personal knowledge of that student's success in the class.
- Before an assignment is graded, students should have an opportunity to practice and master the skill that is being measured.
- Before beginning an assignment, a student should be shown a well done model (e.g., a sample essay or project) in order to gain a better understanding of what is expected of him.
- Formative assessments – those assignments used to let the teacher know how the student is performing during the instruction and practice portion of the learning process – should be recorded but not factored into the final grade.
- It is more important to know how well a student mastered the designated content than when; therefore, new evidence of understandings should replace old evidence in the grade book.
- Only what has been taught in class should be graded.
- Student tardies, mis-behaviors, absences, or other conduct issues should be considered when determining a student's academic grade.
- Students benefit more from meaningful feedback (written or oral) than from a single letter or number grade.
- Students should be evaluated more heavily on their work completed at the end of a unit of study because it reflects their eventual understandings.
- Students should be given an opportunity to self-assess and revise their work using a criterion or grading guide before the teacher grades it.
- Students who are given options for demonstrating their understandings perform better than students who are not.
- The best consequence for students who do not complete an assignment is to require them to complete the assignment.

- The purpose of a grade is to communicate a student's level of academic achievement to students, parents, and others.
- The teacher should be ultimately in control of the grading decisions.

Respondents were asked to indicate the degree to which they believed these statements to be best practice by selecting a level of agreement or disagreement on a Likert scale with 4 points – 3 representing *complete agreement*, 2 representing *general agreement but with some reservations*, 1 representing *general disagreement but with some reservations*, and "0" representing *complete disagreement*.

For this and in all questions on all survey instruments in this research study, these particular ranges of options were intended to capture the intensity of the respondents' agreement or disagreement but without the option for neutrality; the intentional omission of neutrality is fitting to this study as much as the responses to these items are not hypothetical to the respondents – i.e., they are considering statements that have immediate connections to their current professional lives. Researchers from the University of Toronto and University of Denver (Lamb, Allen, & Green, 2010) found that "informed respondents" (p. 2) – those with professional experience in the field related to the statements of study – had a lesser need for neutral responses, and Krosnick et al. reported (2002) that the neutral option can "discourage some respondents from doing the cognitive work necessary to report" (p. 371) truthfully or to the best of their abilities their attitudes toward the topic at hand.

A second instrument, designed for the purpose of directing the interview and grade book analysis process, called Interview/Portfolio Analysis: Teachers' Grading Practices in High School English, was developed using both open and selected response items. This specific

purpose of this instrument is to determine the extent to which grading beliefs and reported practices are, indeed, practiced.

And a final instrument, Survey 2 called Survey on Causes for Discrepancies between High School English Teacher Beliefs and Practices, was used to capture teachers' perceptions of the reasons for discrepancies found to exist between teachers' beliefs about best practices, their self-reported practices, and evidence of actual practice. To that end, the development of this survey was directed by results gathered from both Survey 1 and the interview/portfolio analysis instrument.

All instruments were tested for reliability and construct validity using one of the following methods: a test/re-test model to measure consistency (with results analyzed by way of a bivariate correlation), a pilot sample survey of content validity using directed and open-ended responses, or a quasi-pilot study using early participants' post-survey feedback. Survey 1 and Survey 2 were administered through [www.surveymonkey.com](http://www.surveymonkey.com), and employed various logic features that enhanced their validity.

### *Participants*

This study was conducted with regards to the population of U.S. high school teachers who currently teach at least one English-language arts class required for graduation, in any Grade 9-12, and who are duly certified to teach that class according to the requirements established in their state. Excluded from the targeted population, for the purposes of this work, are non-certified teachers or teachers of English elective courses only such as creative writing and journalism. Additionally, because of problems with generalizability, teachers at campuses with atypical populations were not considered to be the population for this study. For the

purposes of this study, schools with atypical populations include: schools not deemed “Regular” through United States Census categorization, alternative campuses, vocational schools, special education centers, charter schools, magnet schools, online/virtual schools, correspondence schools, hospital schools, juvenile/detention schools, adult only campuses, and night only campuses.

Non-probability sampling procedures were applied – specifically, a convenience sample was used with the understanding that a sampling bias likely exists against those in the population not easily accessible and against those unlikely to volunteer to participate in a study. The parameters of the convenience sample included all those high school English teachers considered part of the population who were within ease of access through direct contact, indirect contact via association with others personally known to the researcher, or indirect contact via social networking sites Facebook, Twitter, and NCTE’s Teaching and Learning Forum.

#### *Survey 1 Sample*

One hundred and thirty-six members of the population completed Survey 1. Thirty-eight states and 85 schools are represented in the total sample. Demographics were collected and are presented and compared to the population demographic in Table 1.1.

Data regarding the level at which participants had familiarity with educators and researchers widely published on the topic of grading is reported in Table 1.2. This list is inclusive of the grading experts whose grading recommendations were used to construct the items in Survey 1.

Table 1.1

*Demographic Profiles of Participants vs. U.S. High School English Teachers*

Characteristic	Population %*	Sample %
<b>Age</b>		
≤29	21	10
30-39	27	30
40-49	22	30
50+	31	30
<b>Gender</b>		
Male	16	11
Female	84	89
<b>Years of Teaching Experience</b>		
1-5	26	15
6-9	16	16
10-14	16	21
15-24	23	33
25+	17	15
<b>Types of School Community</b>		
Rural	42	25
Town	18	12
City	16	14
Suburb	24	49

\*Feistritzer (2011) and U.S. Department of Education’s National Center for Education Statistics (2012).

It should be noted that in addition to writing on grading and grading practices, several of these researchers listed have also written extensively on other topics including classroom instruction, classroom management, differentiation, and curriculum design. These data are not intended to suggest that respondents were aware of these grading experts because of the experts’ work on grading alone.

These additional characteristics of the sample were derived from additional data collected in Survey 1:

- 85% of respondents were either members of or very familiar with the work of NCTE.

- 36% of respondents were either members of or very familiar with the work of ASCD.
- Current teaching assignments of respondents were as follows: 38.8% taught 9th grade, 52.2% taught 10<sup>th</sup> grade, 41.3% taught 11<sup>th</sup> grade, 37.1% taught 12<sup>th</sup> grade, with many respondents reporting teaching assignments in more than one grade level.
- In addition to currently teaching high school, at least 28.9% of respondents had previously taught middle school.

On-level, honors, and advance placement teachers were represented in the sample.

Table 1.2

*Participant Knowledge of Grading Experts*

Grading Expert	I have an extensive understanding of his/her work.	I have a limited understanding of his/her work.	I know the name but am not otherwise familiar with his/her work.	I do not recognize the name.
Alfie Kohn	12.6%	21.0%	13.4%	52.9%
Carol Ann Tomlinson	22.7%	16.0%	20.2%	41.2%
Douglas Reeves	7.6%	7.6%	11.9%	72.9%
Grant Wiggins	22.0%	16.9%	11.0%	50.0%
Jay Mctighe	18.5%	16.0%	13.4%	52.1%
Ken O'Connor	5.1%	11.9%	11.9%	71.2%
Ric Stiggins	10.1%	12.6%	11.8%	65.5%
Robert Marzano	48.3%	25.0%	8.3%	18.3%
Susan Brookhart	3.4%	5.0%	12.6%	79.0%
Thomas Guskey	4.3%	7.7%	9.4%	78.6%

-

### *Interview/Portfolio Analysis Sample*

Thirty-nine respondents to Survey 1 also participated in the subsequent interview/portfolio analysis through a volunteering process initiated in the final item of Survey 1. Each volunteer was contacted using a self-provided email address or telephone number. Separate demographics and characteristic information were not captured for this sub-group.

### *Sample for Survey 2*

The convenience sample for Survey 2 was derived using the same methods as those used for Survey 1, including direct contact, indirect contact via association with others personally known to the researcher, or indirect contact via social networking sites Facebook, Twitter, and NCTE's Teaching and Learning Forum. Ninety-nine respondents began the survey, and 68 completed it. One hundred percent of the participants in this survey were members of the intended population; additional demographic information was not collected.

## Results

### *Teacher Beliefs about Grading Practices from Survey 1*

The strategy of inquiry employed for the exploration of teacher beliefs about best grading practices was quantitative by way of a sample, cross-sectional survey in the form of an on-line questionnaire. These instructions were provided to respondents: *Please answer this question with regards to your personal beliefs about best grading practices that should be implemented by high school English teachers in general.* These instructions were followed by this question: *To what extent do you agree with each of the following statements?*

Results presented in Table 1.3 represent a treatment of these items as Likert-type in that they are functioning here as independent of each other. Consequently, they were analyzed using a methodology appropriate to an ordinal scale (Boone & Boone, 2012; Clason & Dormody, 1994; Jamison, 2004; Kuzon, Urbanchek, & McCabe, 1996). These item results are reported descriptively, with frequency reported in percentages and mode reported to indicate the central tendency. (For conciseness in reporting, the best practice topics listed in Table 1.3 are abbreviations of the complete best practice statements used on the survey.)

In eight of the 14 items for this question, the response with the highest frequency was found in the category of more extreme agreement – *Completely agree* – and in one instance, had shared bi-modality with the next level of agreement. In the remaining five of the 14 items, the mode was found in the category representing the next level of agreement – *Mostly agree, but with some reservations*.

Though limitations in the ability to interpret ordinal data expressed only in descriptive statistics exist, it can be concluded that the respondents to this survey agreed that each of the 14 grading system practices listed here were indeed best practices.

In addition to these results determined by the Likert-type treatment of these items, the data were analyzed with regards to how each of the 14 items works together as a single Likert scale item. Once responses provided for the 14 items related to beliefs about best grading practices were summed, they were analyzed using a methodology appropriate to authentic Likert items (Boone & Boone, 2012; Likert, 1932), including both descriptive statistics (measures of the mean, median, and mode, and standard deviations) and inferential statistics (the

standard error of the mean and upper and lower limits of confidence intervals). These results are presented in Table 1.4.

Table 1.3

*Best Practice Beliefs in Grading Systems, Likert-type Items*

Best Practice / Topic	Agreement		Disagreement	
	Completely agree (3)	Mostly agree, but with some reservations (2)	Mostly disagree, but with some reservations (1)	Completely disagree (0)
Selective adjustment of a student's final grade	21.1%	46.9%	22.7%	9.4%
Opportunities for practice and mastery before grading	72.8%	23.3%	3.9%	0.0%
Use of exemplars for understanding expectations	57.4%	37.2%	5.4%	0.0%
Avoiding calculating formative assessments into final grades	13.2%	43.4%	38.8%	4.6%
Replacing old evidence with new evidence in grade book	30.4%	37.5%	22.7%	9.4%
Grading only on what was taught in class	36.4%	42.6%	14.7%	6.2%
Separation between behavior grades and academic grades	41.4%	35.2%	14.8%	8.6%
Emphasis on feedback rather than grades	79.1%	19.4%	1.5%	0.0%
Weighting end of unit grades more than early evidence	43.0%	42.2%	12.5%	2.3%
Student self-assessment prior to grading	53.5%	43.4%	3.1%	0.0%
Student options for demonstrating understandings	48.4%	48.4%	2.3%	0.8%
Forced completion of all assignments (no zeroes)	39.5%	42.6%	12.8%	5.4%
Use of grades solely for communication of achievement	44.9%	48.9%	6.2%	0.0%
Grading decisions (what, how, and weight) in teacher control	58.9%	37.2%	3.9%	0.0%

Table 1.4

*Best Practice Beliefs in Grading Systems, Likert Scale (Collective)*

<i>M</i>	<i>SEM</i>	Median	Mode	<i>SD</i>	95% <i>CI</i> [LL, UL]
2.28	.018	2.00	3.00	.798	[2.24, 2.32]

The treatment of these individual Likert-type items as a collective Likert scale allows for an additional level of certainty that the results from the survey are representative of the population. For example, when 95% confidence intervals are drawn around the mean of 2.28, we are able to say with 95% certainty that the population mean is between 2.24 and 2.32. This, combined with the evidence of the median (2.00) and the mode (3.00) allows for more certainty that the population as a whole would be in some level of agreement (*Completely* or *Mostly*) that these expert recommendations for grading practices are, indeed, best practice.

*Teacher Self-Reported Grading Practices from Survey 1*

The strategy of inquiry employed for the exploration of teachers' self-reports on grading practices was quantitative by way of a sample, cross-sectional survey in the form of an on-line questionnaire. Survey 1, the same survey used to collect data for teacher beliefs, was used for this purpose. Respondents were provided these instructions: *Think of the English class(es) you currently teach, and please answer this question with regards only to your own current and typical grading practices.* These instructions were followed by this question: *To what extent do you agree with each of the following statements?*

Results presented in Table 1.5 represent a treatment of these items as Likert-type in that they are functioning here as independent of each other. Consequently, they were analyzed using a methodology appropriate to an ordinal scale. These item results are reported

descriptively, with frequency in percentages and mode reported to indicate the central tendency. (For conciseness in reporting, the best practice topics listed in Table 1.5 are abbreviations of the complete best practice statements used on the survey.)

Table 1.5

*Self-Reported Practices in Grading Systems, Likert-type Items*

Best Practice / Topic	Agreement		Disagreement	
	Completely agree (3)	Mostly agree, but with some reservations (2)	Mostly disagree, but with some reservations (1)	Completely disagree (0)
Selective adjustment of a student's final grade	27.2%	39.0%	22.8%	11.0%
Opportunities for practice and mastery before grading	64.7%	31.6%	3.7%	0.0%
Use of exemplars for understanding expectations	46.3%	40.4%	11.8%	1.5%
Avoiding calculating formative assessments into final grades	11.8%	35.3%	43.4%	9.6%
Replacing old evidence with new evidence in grade book	25.9%	42.2%	20.7%	11.1%
Grading only on what was taught in class	39.7%	42.6%	10.3%	7.4%
Separation between behavior grades and academic grades	48.1%	28.1%	17.8%	5.9%
Emphasis on feedback rather than grades	77.2%	20.6%	1.5%	0.7%
Weighting end of unit grades more than early evidence	49.3%	37.5%	9.6%	3.7%
Student self-assessment prior to grading	46.3%	43.4%	9.6%	0.7%
Student options for demonstrating understandings	54.4%	36.0%	8.1%	1.5%
Forced completion of all assignments (no zeroes)	26.5%	41.9%	22.8%	8.8%
Use of grades solely for communication of achievement	51.9%	45.9%	2.2%	0.0%
Grading decisions (what, how, and weight) in teacher control	50.4%	43.7%	4.4%	1.5%

In nine of the 14 items for this question, the highest frequency response was found in the category of more extreme agreement – *Completely agree*. Additionally, five other modes fell into a category of agreement – *Mostly agree, but with some reservations*. Combined, 13 of the 14 responses represented some level of agreement. The mode for the one remaining item – *Avoiding calculating formative assessments into final grades* – was in the disagreement category *Mostly disagree, but with some reservations*. This same topic had the highest frequency of response in the more extreme disagreement category option – *Completely disagree* – with 11.1%.

The respondents to this survey self-reported at some level that they engaged in 13 of the 14 grading system practices listed.

Again, in addition to these results determined by the Likert-type treatment of these items, the data were analyzed with regards to how these 14 items worked together as a single Likert scale. Once all responses provided for the 14 items related to the self-reporting of grading practices were summed, they were analyzed using a methodology appropriate to an interval scale, including both descriptive statistics (measures of the mean, median, mode, and standard deviations) and inferential statistics (the standard error of the mean and upper and lower limits of confidence intervals). These results are reported in Table 1.6.

Table 1.6

*Self-Reported Practices in Grading Systems, Likert Scale (Collective)*

<i>M</i>	<i>SEM</i>	Median	Mode	<i>SD</i>	95% CI [LL, UL]
2.22	.019	2.00	3.00	.838	[2.19, 2.25]

The treatment of these individual Likert-type items as a collective Likert scale allows for an additional level of certainty that the results from the survey are representative of the population. For example, when 95% confidence intervals are drawn around the mean of 2.22, we are able to say with 95% certainty that the population mean occurs between 2.19 and 2.25. This, combined with the evidence of the median (2.00) and the mode (3.00) allow for more certainty in concluding that the population as a whole would report that they practice, to some degree, expert grading recommended practices.

#### *Comparisons of Best Practice and Self-Reported Practice Results from Survey 1*

An analysis of the relationship between best practice results and self-reported results was conducted to determine if there were discrepancies between teachers' beliefs about grading practices and self-reports about their own practices. This analysis was essential at this point in time because the design of the Interview/Portfolio Analysis and the subsequent Survey 2 would be affected by the outcomes.

A comparison of the results from the best practice belief responses and their corresponding self-reported practice responses was completed in two ways – (1) by completing a comparison study between the best practice Likert-type items and their corresponding reported practice Likert-type items, and (2) by completing a comparison study between the whole of the best practice responses as a Likert scale with its corresponding whole of the self-reported practice responses also as a Likert scale.

Respondents reported that they believed that it is best practice to avoid calculating formative assessments into final grades; however, they reported that they do not avoid calculating formative assessment into final grades. Because of the results showing a

discrepancy between this belief and this practice, this item was not further examined in the comparison analysis and was not analyzed in the interview/portfolio assessment for actual practice. It was, however, part of Survey 2 which asked teachers to consider causes for discrepancies.

### *Likert-type Item Comparisons from Survey 1*

A comparison of the remaining Likert-type items was completed using studies of the mode, median, mean rank, and the Mann-Whitney U Test to study a comparison of the medians (as is appropriate for ordinal scale data). The Mann-Whitney U Test was used to determine if a statistically significant difference existed between the recorded results for teachers' beliefs about best practice (in both grading systems and recorded assignments graded) and their self-reported practices (in both grading systems and recorded assignments graded). The measure of mean rank was employed to report which grouping variable – the best practice responses or own practice responses – indicated a stronger response toward agreement with expert recommendations. Data were analyzed using SPSS (Analyze > Nonparametric Tests > Legacy Dialogues > 2 Independent Samples). Results are reported in Table 1.7 and include measures of median, mode, mean rank (*M Rank*), Levels of statistical significance (Sig.), and a determination of statistically significant difference between best and own practices (*Not equal or equal*).

The null hypothesis for this analysis was that the median difference between the pairs was zero, or  $H_0: Mdn_1 = Mdn_2$ . In cases where the p-value was smaller than alpha (p-value < .05), the null hypothesis was rejected, and it was concluded that the median scores of the best practice responses were not equal to the median scores of the self-report practice responses.

Table 1.7

*Comparisons of Best Practice and Self-Reported Practice Results– Likert-type Items*

Item	Median	Mode	M Rank	Sig.	Statistically equal or not to each other
Selective adjustment of a student's final grade					
Best Practice	2.00	2.00	130.88	.724	Equal
Own Practice	2.00	2.00	134.02		
Opportunities for practice and mastery before grading					
Best Practice	3.00	3.00	137.27	.276	Equal
Own Practice	3.00	3.00	128.95		
Use of exemplars for understanding expectations					
Best Practice	3.00	3.00	141.69	.045	Not Equal
Own Practice	2.00	3.00	124.76		
Replacing old evidence with new evidence in grade book					
Best Practice	2.00	2.00	134.39	.602	Equal
Own Practice	2.00	2.00	129.74		
Grading only on what was taught in class					
Best Practice	2.00	2.00	130.12	.522	Equal
Own Practice	2.00	2.00	135.73		
Separation of behavior grades from academic grades					
Best Practice	2.00	3.00	135.82	.397	Equal
Own Practice	2.00	3.00	128.38		
Emphasis on feedback rather than grades					
Best Practice	3.00	3.00	134.35	.698	Equal
Own Practice	3.00	3.00	131.72		
Weighting end of unit grades more than early evidence					
Best Practice	2.00	3.00	138.39	.354	Equal
Own Practice	2.00	3.00	136.37		
Student self-assessment prior to grading					
Best Practice	3.00	3.00	141.27	.057	Equal
Own Practice	2.00	3.00	125.27		
Student options for demonstrating understandings					
Best Practice	2.00	2.00	130.80	.693	Equal
Own Practice	3.00	3.00	134.10		

*(table continues)*

*Comparisons of Best Practice and Self-Reported Practice Results– Likert-type Items (continued).*

Forced completion of all assignments (no zeroes)						
Best Practice	2.00	2.00	144.99	.006	Not Equal	
Own Practice	2.00	2.00	120.57			
Use of grades solely for communication of achievement						
Best Practice	2.00	2.00	129.03	.418	Equal	
Own Practice	3.00	3.00	135.81			
Grading decisions (what, how, and weight) in teacher control						
Best Practice	3.00	3.00	138.65	.145	Equal	
Own Practice	3.00	3.00	126.62			

Two items in the grading systems section of the survey showed a statistically significant difference in the median between the best practice responses and the self-reported responses: *use of exemplars for understanding expectations* and *forced completion of all assignments (no zeroes)*. Additionally, one item was very close to a statistically significant difference – *Student self-assessment prior to grading*.

The summed frequencies for these three items for the levels of agreement (i.e. *Completely agree* and *Mostly agree*) in the best practice study and the self-reported practice study were 94.6%/86.7%, 82.1%/68.4%, and 95.5%/91.1%, respectively. Thus, the difference of the medians is not reporting the difference between agreement and disagreement so much as it is reporting that the degree to which teachers report use of this practice in their own classrooms is less than the degree to which they report that they believe it is a best practice.

*Collective Likert Scale Item Comparisons from Survey 1*

The comparison of the Likert scale items was completed using studies of the means, standard deviations, standard errors of the mean, and an independent samples T-test (as is appropriate for interval scale data). A comparison was made between teachers' overall beliefs

about grading systems and their overall self-reported practices. The items in each of these queries were summed for the purpose of this comparison study. Table 1.8 and include measures of mean (*M*), standard deviation (*SD*), standard error of the mean (*SEM*), levels of statistical significance (*Sig.*), and the mean difference (*M* Difference).

Table 1.8

*Comparisons of Best Practice and Self-Reported Practice Results- Likert Scale (Collective)*

	<i>M</i>	<i>SD</i>	<i>SEM</i>	<i>Sig.</i> (2-tailed)	<i>M</i> Difference
Best practice beliefs in grading systems	2.28	.798	.018		
Self-reported best practices in grading systems	2.22	.838	.019	.015	.064

The null hypothesis for this analysis was that the mean difference between the pairs was zero, or  $H_0: M_1 = M_2$ . Because the *p*-value was smaller than alpha (*p*-value < .05), the null hypothesis was rejected, and it was concluded that the mean scores of the best practice belief responses were not equal to the mean scores of the self-report practice responses. Or, in other words, the degree to which teachers self-report using certain grading practices is less, to a statistically significant degree, than the degree to which teachers identify those practices as best practices. The results from Levine’s Test for Equality of Variances (*p* = .147) that this test meets the assumption of homogeneity of variances.

*Teachers’ Actual Grading Practices from Interview/Portfolio Analysis*

To collect data with regards to teachers’ actual practices, an interview/portfolio analysis process to analyze teachers’ grading records was utilized; this process employed a mixed-

methods design in the form of a standardized, open-ended interview regarding teachers' actual grading practices as recorded in their grade books. The wording of each interview question and the sequence in which all interview questions were asked was pre-determined based on the results from Survey 1.

The instrument for this interview/grade book analysis was comprised of 18 questions, and the sample was comprised of 39 volunteer members of the population who also participated in Survey. Respondents were asked to select one class period of the day that was a typical representation of their English classes as a whole and to reference that class period during the most recently completed grading period of the school year.

Data collected from the interview indicated that 25.6% of the respondents worked within a school system that used 6-weeks grading periods, and 74.4% of the respondents worked within a school system that used 9-weeks grading periods. No respondents indicated that they used grading periods of any other lengths. Figure 1 reports responses from Question 5 indicating the grade levels and courses represented by the respondents.

The open-response items regarding the grading programs of the respondents were reported on an interval scale and were analyzed both descriptively using mean and standard deviation, and inferentially using standard error of the mean and the calculation of 95% confidence intervals and their upper and lower limits. When appropriate and possible, the approximate percentage of grades affected by the topic (as calculated by considering the average number of grades recorded during a single grading period) was reported, as well. (For conciseness in reporting, the topics listed in Table 1.9 are abbreviations of the complete questions used during the interview/portfolio assessment.)

Table 1.9

*Interview/Portfolio Assessment Results*

Topic	<i>M</i>	<i>SD</i>	<i>SEM</i>	95% CI [LL, UL]	~ % of assignments
Number of assignments recorded during the grading period	12.23	3.07	.492	[11.27, 13.19]	--
Number of times teacher selectively adjusted students' final score	0.15	0.59	.094	[-.03, .33]	--
Number of assignments that represent few or no opportunities for practice	4.03	2.47	.395	[3.26, 4.80]	26%-39%
Number of times old evidence replaced by new evidence	2.87	2.54	.407	[2.07, 3.67]	--
Number of assignments measuring a skill or knowledge taught or learned in another class	3.00	2.84	.454	[2.11, 3.89]	17%-30%
Percentage of students earning a score of "0" on an assignment.	12.39	14.52	2.32	[7.84, 16.94]	--
Percentage of students earning a grade reduction for procedural reasons	8.15	9.69	1.55	[5.11, 11.19]	--
Number of assignments collected and graded following initial feedback	2.08	2.47	.396	[1.30, 2.86]	16%-23%

From these data, the following conclusions regarding actual grading systems can be drawn:

- Teachers selectively adjusted students' final scores at a rate of less than one-third of one student per grading period.

- On average, four or more assignments in a grading period (or 26%-39% of assignments) were collected and graded immediately following instruction, prior to a student having an opportunity to practice or master the skill or necessary knowledge.
- Old evidence of ability recorded in the grade book was replaced with new evidence of ability somewhere between 2 - 4 times during a grading period. Class sizes were not recorded in this study; however, based on an estimation of typical class sizes in high school English classes, the percentage of grades replaced by new evidence was less than 1%.
- Between 17% and 30% of the assignments graded and recorded relied on a skill or knowledge that was taught or learned in a previous class.
- Approximately 8% to 17% of students in a class earned a score of a zero on an assignment during a given grading period.
- Approximately 5% to 11% of students in a class earned a score reduction based on procedural (non-academic) reasons.
- On average, 1 - 3 assignments (or 16% to 23% of the assignments) were collected and graded following an opportunity for students to receive feedback.

Additionally, the interview/portfolio analysis sought evidence of the causes for a student earning a recorded zero grade for an assignment. The responses to this open-response item were analyzed for patterns, organized into categories by theme, and then summed for the purpose of reporting. Each response was forced into a single category, with an "Other" category included for the management of responses that did not appear to be part of a pattern. The resulting data are reported out in Figure 1 using frequencies reported as percentages.

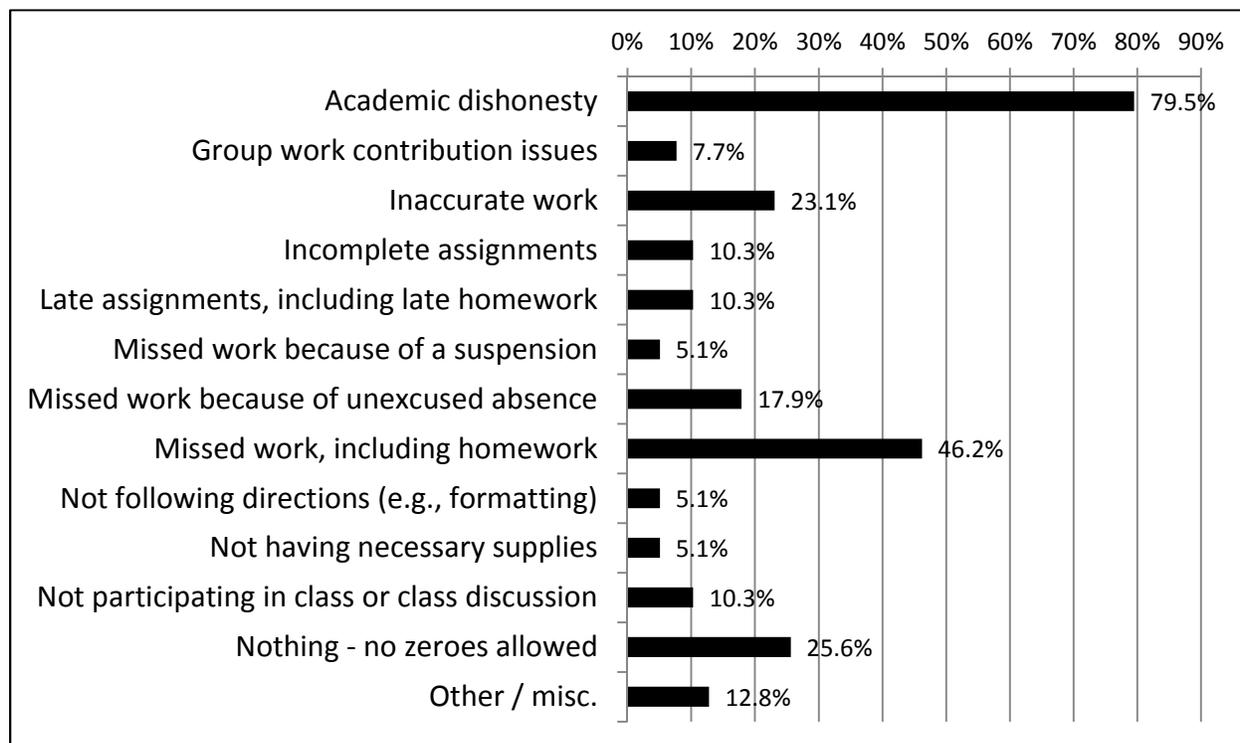


Figure 1.1. Percentage of respondents assigning a “0” grade for reasons indicated.

These data reveal that 23.1% of the respondents used zeroes for academic reasons (i.e., *inaccurate work*), but zeroes were also used for non-academic reasons such as not having supplies or not following directions. Behavioral issues accounted for a portion of the zeroes used, including penalizing students with grades for days missed because of suspensions or unexcused absences and zeroes received for not following directions such as formatting headings on papers.

The interview/portfolio analysis asked respondents to indicate their reasons for recording and reporting grades. In this question, respondents selected all possible responses that applied and were offered an opportunity to add or qualify a response. Figure 2 presents the data collected in response to this question.

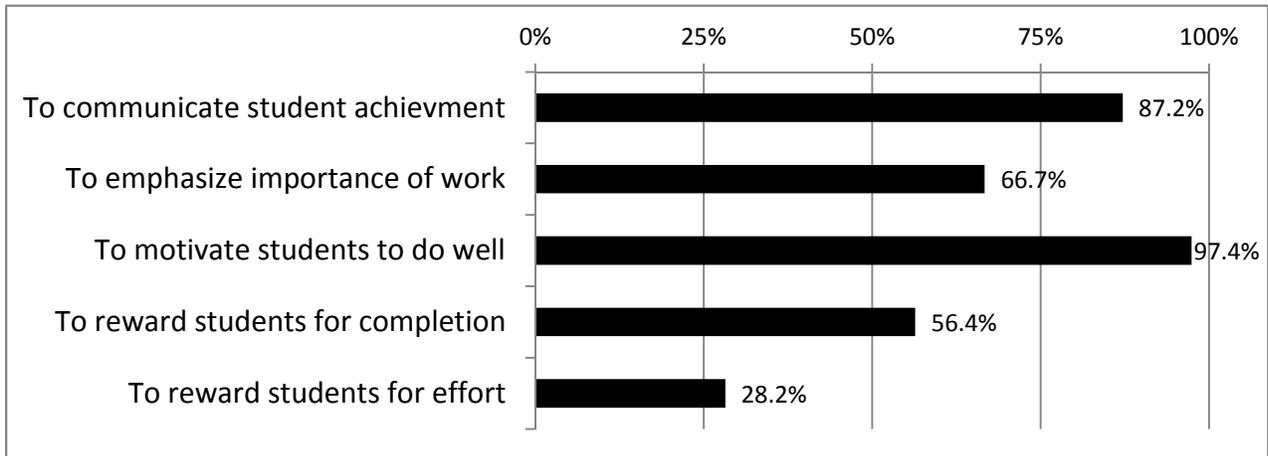


Figure 1.2. (Question 15) Percentage of respondents using grades for reasons indicated.

From these data, a conclusion can be drawn that the primary aim for the use of grades was student motivation. By order of frequency, the next purpose for use was communication of student achievement. Because these two response rates were seemingly analogous, a one sample T-test between percentages was conducted to determine if there was a significant difference between the percent of respondents who reported the use grades for student achievement reporting and the percent of respondents who reported the use of grading for motivating students. Data were analyzed using SPSS (Analyze > Compare Means > One Sample T T-Test).

The null hypothesis for this analysis was that the difference of proportions between the two results was zero, or  $H_0: \text{Percent}_1 = \text{Percent}_2$ . Because the  $p$ -value of .000 (Sig., two-tailed) was less than alpha ( $p$ -value < .05), the null hypothesis was rejected, and it was concluded that the percentage of the respondents using grades to report student achievement was not equal to the percentage of the respondents using grades to motivate students. Because the percentage was higher for teachers using grades to motivate students, it could be concluded

that teachers use grades to motivate students more often than they do to report student achievement.

The interview/portfolio analysis sought evidence of control of grading decisions. In one question, respondents were asked to report who controlled the number of grades collected in a grading period; in the following question, they were asked to report who controlled decisions regarding which assignments would be collected and graded. These results are reported in comparison by answer choice in Figure 3.

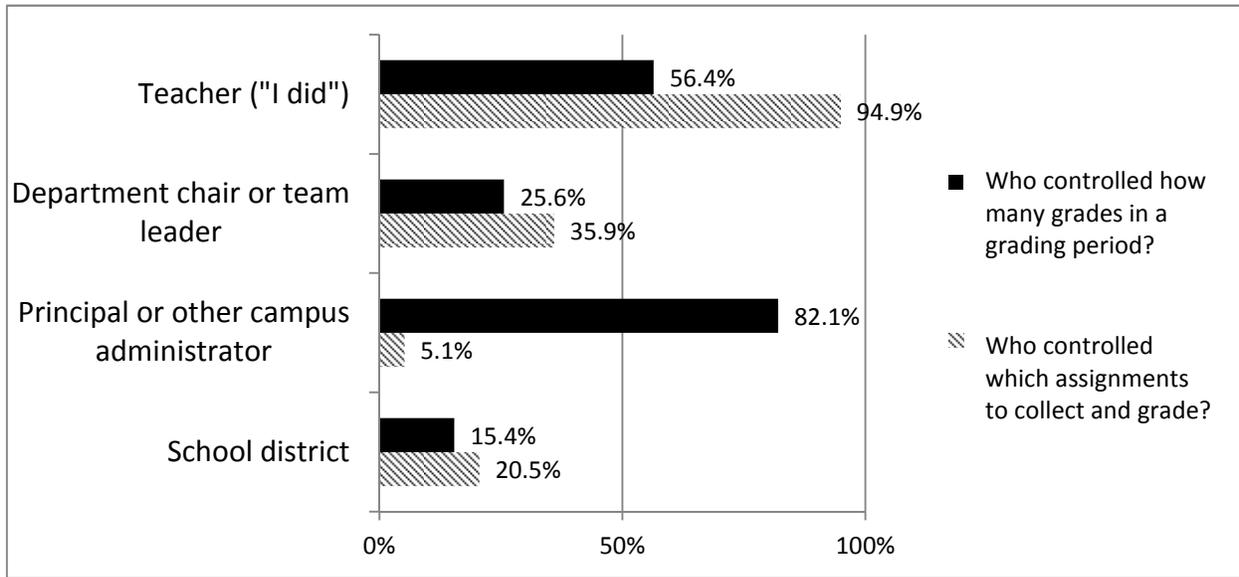


Figure 1.3. Control of Grades

Results here indicate that with regards to determining the number of grades collected, teachers have a lesser degree of control than principals or other administrators but more control than the department chair/team leader or the school district. Additionally, with regards to which assignments were collected and graded, the teacher has more control, notably, than administrators.

## Interview/Portfolio Analysis Results Summary

In summation, the interview/portfolio analysis revealed that high school English-language arts teachers' grading practices in this study included:

- Control over some grading decisions, i.e, *which* assignments were collected and graded
- Emphasis on feedback rather than grades (to a limited degree)

Their grading practices did not include:

- Selective adjustment of a student's final grade
- Replacing old evidence with new evidence in grade book
- Opportunities for practice and mastery before grading
- Control over some grading decisions, i.e, *how many* assignments were collected and graded
- Grading only on what was taught in class
- Separation of behavior grades from academic grades
- Forced completion of all assignments (no zeroes)
- Use of grades solely for communication of achievement

## Comparisons of Results

An analysis of the relationship among results in research questions 1, 2, and 3 was conducted to determine if there were discrepancies between teachers' beliefs about grading practices, their self-reports about their own practices, and their actual practices. This analysis was essential at this point in time because the design of Survey 2 would be affected by the outcomes.

The results of this comparison are presented in Table 1.10. In addition to the columns representing the results from the conducted studies (on beliefs, self-reported practices, and actual practices), an additional column labeled *Discrepancy* is employed for the purpose of indicating whether or not a discrepancy was found and, consequently, if the item in that row would be included in Survey 2. A discrepancy would be considered to be in existence in the following scenarios:

- A reported belief was not self-reported in practice.
- A reported belief was found in self-practice but not in actual practice.
- A non-belief was self-reported in practice.
- A non-belief was not self-reported in practice but was found in actual practice.

Based on this comparative analysis of results, nine items were found to have a discrepancy between or among what teachers believe, what they self-report to be their practices, and their actual practices.

Table 1.10

*Identifying Discrepancies*

Practice	Best Practice Belief	Self-Reported Practice	Actual Practice	Discrepancy
Selective adjustment of a student's final grade	Yes	Yes	No	Yes
Opportunities for practice and mastery before grading	Yes	Yes	No	Yes
Use of exemplars for understanding expectations	Yes	Yes	Other <sup>1</sup>	n/a
Avoiding calculating formative assessments into final grades	Yes	No	n/a	Yes

*(table continues)*

*Identifying Discrepancies (continued).*

Replacing old evidence with new evidence in grade book	Yes	Yes	No	Yes
Grading only on what was taught in class	Yes	Yes	No	Yes
Separation of behavior grades from academic grades	Yes	Yes	No	Yes
Emphasis on feedback rather than grades	Yes	Yes	Yes	No
Weighting end of unit grades more than early evidence	Yes	Yes	Other <sup>2</sup>	n/a
Student self-assessment prior to grading	Yes	Yes	Other <sup>1</sup>	n/a
Student options for demonstrating understandings	Yes	Yes	Other <sup>1</sup>	n/a
Forced completion of all assignments (no zeroes)	Yes	Yes	No	Yes
Use of grades solely for communication of achievement	Yes	Yes	No	Yes
Grading decisions in teacher control				
(a) Which grades to record	Yes	Yes	Yes	No
(b) How many grades to record	Yes	Yes	No	Yes

Other<sup>1</sup> indicates an item that was inadvertently left off of the interview/portfolio analysis (due to errors in pre-calculations) and therefore cannot be studied further.

Other<sup>2</sup> indicates an item that was omitted mid-way through the interview/portfolio analysis when it was determined that the end of a grading period and the end of a unit of study might not be concurrent.

## Survey 2 Description and Results

Survey 2, called “Survey on Causes for Discrepancies between High School English Teacher Beliefs and Practices,” was designed to capture teachers’ perceptions of the causes for the nine discrepancies found to exist between teachers’ beliefs about best practices, their self-reported practices, and evidence of actual practice.

The process of reaching conclusions from the results from Survey 2 was broader and less restrictive in design than the process used to draw conclusions from the more quantitative-like data from Survey 1 and from the qualitative-quantitative design found in the interview/portfolio analysis study. In one sense, the responses here were treated more as insights into bigger concepts and causes than as explicit and definitive reasons for why certain things happen in school and school systems. In other words, the results determined from Survey 2 were considered to be ways in which to see what, if any, political/power systems were influencing the implementation of best practices and/or practices teachers believed to be essential.

There was no hypothesis being tested here except to say that it is believed that teachers are able to discuss and identify a variety of causes for the discrepancies between what they believe and what they say they do and what actual practice reveals.

Each item in Survey 2 was designed to accommodate both selected- and open-responses. There were no limitations on how many causes a respondent might select, and all respondents were provided the opportunity to qualify or add their own or additional responses. Responses were first analyzed qualitatively through a process that identified and characterized trends in themes, and then these patterns were summed and are presented here using descriptive statistics including frequencies reported as percentages and mode as the measure of central tendency. This mixed method approach allowed for conclusions to be based on the observed trends, while quantified evidence is presented as the degree to which the responses support the conclusions.

In all, the presented causes for the observed discrepancies fell into one of six categories: (1) interference by someone with authority greater than the teacher, (2) interference by limitations imposed by the electronic grade book, (3) limited time for effective implementation, (4) limited student capacity (suggesting that students would have a difficult time engaging in or understanding the practice), (5) limited parent capacity (suggesting that parents would not be able to comprehend or support a practice without extensive explanations), and (6) limited teacher capacity (suggesting that teachers may believe in the practice but may not actually know how to implement it).

Table 1.11 presents the perceived causes for each of the nine discrepancies. An *X* indicates that a particular cause was identified as being a barrier to implementation by at least 1% of the respondents. An asterisk (\*) marks the perceived cause with the highest frequency of responses.

Table 1.11

*Causes for Discrepancies, Responses Selected (X) and Mode ( \* )as Measure of Central Tendency*

	Interference by authority	Interference by electronic grade book	Limited time available	Limited student capacity	Limited parent capacity	Limited teacher capacity
Causes for no selective adjustment of a student's final grade	*	X	X		X	X
Causes for no opportunities for practice or mastery before grading	X		X			*
Causes for the inclusion of formative assessment grades in final grades	*	X	X	X	X	
Causes for not replacing old evidence with new evidence in grade book	*	X	X	X	X	
Causes for not grading only what was taught in class	X					*

*(table continues)*

*Causes for Discrepancies... (continued).*

Causes for not separating behavior grades from academic grades	X			*		
Causes for not using forced completion of assignments (no zeroes)	X	X		*	X	X
Causes for not using grades solely for communicating student achievement	X					X *
Causes for not having control over deciding how many grades to take	*					X X
Causes for not using writing samples that measure grammar/usage				*		X X
Causes for not including classroom discussions in final grades				*		X
Causes for not using measures of the writer's notebook in final grades				*		X X
Causes for not using student presentation grades in final grades				*	*	X

### Discussion

Despite the fact that they may not be well-informed about grading practices, as evidenced by their limited knowledge of the authors in the field, the respondents in this study agree fully that the 14 expert-recommended grading practices make sense and should be used in our school systems. Even with recommended practices as disassociated from current practice as adjusting students' final scores based on non-quantified evaluations of student learning, or replacing old understandings of achievement in the grade book with new evidence of learning, teachers resoundingly report support for these types of grading systems.

When asked if they used these practices in their own classroom grading systems, the respondents reported that they did, with some hesitations. The Likert scale comparison of the sets of items revealed a statistically significant difference overall, and most item by item

comparisons revealed a lower measure of mean rank for teachers' reported use than for beliefs. One item – the use of formative assessments to inform instruction but not to affect final grades – was reported as not being in use though it was reported as being supported as a best practice.

The disconnect between teacher beliefs and teacher practices begins to become clear, and the discrepancies grow as the study included actual teacher practices as compared to beliefs or reported practices. Results here show that teachers' beliefs are sometimes applied to as few as one-third of one student in a single grading period. Grades are being used for reasons that teachers report to not agree with, and the profile of the uses for a zero is much different than reported in the study of teacher beliefs. In fact, although teachers report at a rate of 97.9% some level of agreement that a grade should solely be used for communication of achievement, a study of teachers' grading records shows that grades are as often used in an attempt to motivate students.

In all, discrepancies were found more often than not. Nine of the 11 grading practices that survived the whole of the study represented some level of discrepancy between beliefs, self-reported practices, and actual practices.

In a comparison of the modes and the identification of causes for these discrepancies, a pattern emerged indicating that participants are likely to hold accountable most often on one of these three sources: campus or district authorities, the limited time available, and the interferences caused by parents. Among these three, the most frequently identified cause was the limited time available. In great summation, it could be concluded that respondents argue

that there just simply is not enough time in the high school English classroom to engage the practices we know to be best for student grading.

Compound this cause with the second and third causes ranked by frequency – the campus and district authorities and parents – and teachers are essentially expressing that they do not have enough time and enough support to do their work well. Both campus and district administrators, it would seem, would be working within professional understandings of best practice much like these teachers are and would be in positions to positively influence the functionality of a school's grading system, and yet respondents find administrators as much of or more of a barrier than parents, students, and teacher capacity.

Recommendations on grading practices that are effective, fair, and valid -- in other words, "best practice grading practices" -- have been provided to teachers in various forms since the turn of the 20<sup>th</sup> century (Cross & Frary, 1999), and yet evidence shows that despite these clear cut recommendations, grading practices remain stagnant, similar to those that have been used for years (Stiggins & Conklin, 1992). If teachers are reporting an understanding of the value of the practice, but there is still no evidence of its use, then it could be argued that the system as a whole and as it currently exists is not able to accommodate effective practices.

Certainly there are other explanations. One way to consider these data is to presume that teachers know best practices, wish to report that they are mostly using them, but know that truly they are not. The discrepancies could be explained if it was known that teachers find the recommended grading practices meaningful only in theory but impractical in application.

Frary (1992) reports that teachers often directly pursue practices that work in direct opposition to those recommended by educational measurement specialists. Teachers may

specifically choose to not allow grades to completely represent achievement, and they may do so because of an agenda built around student nurturing and student advocacy (Brookhart, 1991). Teachers may intentionally “use non-achievement factors in determining grades to mitigate negative social consequences associated with what they consider to be inappropriate use of grades, such as determining eligibility for nonacademic privileges either at home or school” (Cross & Frary, 1999). Teachers may also be concerned that if grades are only a measure of achievement, then low-achieving students will be discriminated against based on their grades. They may believe that there are social consequences for grades as well as efficacy issues that could have long-term effects on students as learners (Brookhart, 1991). In other words, teachers are concerned that true and actual grades could damage a student’s sense of control over learning and result in his failing to attempt to accomplish a task that feels daunting, out of reach, or a task where he feels he has already tried his best and has failed.

Bishop (1992) argues that “teachers can’t be coach and judge.” The conflict for teachers as graders emerges through the incompatibility of the two identities teachers otherwise have to have – the advocate for the student *and* the judge of his work – and results in grading that is based on other factors.

Along these same lines, teachers have also reported the need to avoid scrutiny from parents, students, and administrators (Cross & Frary, 1999). Troug and Friedman (1996) conducted a study in which they concluded that many teachers’ grading practices are not the result of lack of understanding or conflicting professional roles; they are, instead, simply reflective of the practices that teachers believe are expected of them by parents and students.

In their *Respite for Teachers: Reflection and Renewal in the Teaching Life*, Pearson-Casanave and Sosa (2007) write: “How strange that our practices encourage students to believe that a number or a letter grade...is the primary evidence by which they and others know or don’t know something” (p. 14). Others in the educational arena concur that herein lies the problem with grading as it is used in current practice – that is has taken over our school systems and has become the primary focus of student attention in academia. The process of scoring work, of recording grades, and of reporting out standings and rankings provides students, parents, and other community members with the impression that teachers and students alike are functioning as intended. Our ability in the field to seemingly quantify the learning processes has provided must assurance to the community that we have ample evidence of student growth or lack thereof (Pearson-Casanave & Sosa, 2007).

## References

- Black, P., C. Harrison, C. Lee, B. Marshall, and D. Williams. (2003). *Assessment for learning: Putting it into practice*. Maidenhead: Open University Press.
- Boone, H. & Boone, D. (2012, April). Analyzing Likert data. *Journal of Extension*, 50(2). Retrieved from <http://www.joe.org/joe/2012april/tt2.php>.
- Britzman, D. (1991). *Practice makes practice: A critical study of learning to teach*. Albany, New York: State University of New York Press.
- Brookhart, S. (1991). Grading practices and validity. *Educational Measurement: Issues and Practices*, 10(1), 35-36.
- Clason, D., and Dormody, T. (1994). Analyzing data measured by individual Likert-type items. *Journal of Agricultural Education*, (34)4, 31-35.
- Cross, L. H. & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike. *Applied Measurement in Education*, 12(1), 53-72.
- Durm, M. (1993). An A is not an A is not an A: A history of grading. *Educational Forum*, (57)3, 294-297.
- Finkelstein, I. E. (1913) *The marking system in theory and practice*. Baltimore: Warwick & York, Inc.
- Frary, R. B., Cross, L. H., & Weber, L. J. (1992, April 21-23). *Testing and grading practice and opinions in the nineties: 1890s or 1990s?* Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.
- Guskey, T.R. and Bailey, J. M. (2001) *Developing grading and reporting systems for student learning*. Thousand Oaks, CA. Corwin Press.

- Guskey, T. (2010). Lessons of mastery learning. *Educational Leadership*, (68)2, 53-57.
- Jamison, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38, 1217-1218.
- Krosnick, J. A., et al. (2002). The impact of 'no opinion' response options on data quality: Nonattitude reduction or invitation to satisfice? *Public Opinion Quarterly*, 66, 371-403.
- Kuzon, W.M., Jr., Urbanchek, M.G., & McCabe, S. (1996). The seven deadly sins of statistical analyses. *Annals of Plastic Surgery*, 37, 265-272.
- Lamb, T., Allen, G., Green, K. (2010, May). Is 'neutral' on a Likert scale the same as 'don't know' for informed and uninformed respondents? Effects of serial position and labeling on selection of response options. Paper presented at the annual meeting of the *National Council on Measurement in Education*. Denver, CO.
- Likert, R. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 1-55.
- Marzano, R. J. (2001). *Classroom instruction that works*. Alexandria, VA: Association for Supervision and Curriculum Development.
- McTighe, J. & O'Connor, K. (2005). Seven practices for effective learning. *Educational Leadership*, 63(3), 10-17.
- O'Connor, K. (2006). *A repair kit for grading: 15 fixes for broken grades*. Portland: Educational Testing Institute.
- Pearson-Casanave, C. & Sosa, M. (2007). *Respite for teachers: Reflection and renewal in the teaching life*. Ann Arbor, MI: University of Michigan Press.
- Reeves, D. (2008, Feb.) Effective grading practices. *Educational Leadership*, 65(5), 85-87.

- Rich, R. H. (2001). *Hidden factors in teachers' secondary grading practices*. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, NY: State University of New York Press.
- Tomlinson, C. A. (2001). Grading for success. *Educational Leadership*, 58(6), 12-15.
- Truog, A. L. & Friedman, S. J. (1996). *Evaluating high school teachers' written grading policies from a measurement perspective*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New York.

APPENDIX A  
EXTENDED LITERATURE REVIEW

## Growing Interest in Grading Practices

In recent years there has been a notable resurgence of interest on the subject of grades and grading, possibly brought about by far-reaching conversations and reform efforts tied to standards-based education. Historically, conversations about standards – *what* should students learn – have resulted in myriad reforms. From the Committee of Ten report published in 1894 and the National Education Association’s “Cardinal Principles” in 1918 to the Nation at Risk report in 1985, we have evidence of a consistent interest in the field in determining what is worth learning. And while discussions of grading practices have a long history as well, the connections between standards and grading reached a more public level when conversations about what students should be learning led to federal government interventions in the form of legislation.

The United States Congress’ “Goals 2000: Educate America Act” (1994) included a blue print for content and skills standards and a delineation of specific performance standards that would require assessment, measurement, and reporting. Likewise, the “No Child Left Behind” Act of 2001 referenced measurement needs by way of accountability requirements and growth models and reminded the public by way of its “Report Card Guidance” document that “...the more parents and community members know about the academic achievement of their children and their schools, the more likely they are to be involved in their local schools and the public school system” (p. 10).

Likewise, In June of 2010, the National Governors Association Center for Best Practices (NGA Center) and the Council of Chief State School Officers (CCSSO) led a national effort to create the “Common Core State Standards” – a catalog of student expectations designed to be

implemented nationally to provide parents and other stakeholders the promise of consistency in content and skills goals for all United States students, “regardless of where they live” (Common Core State Standards Initiative, 2011). One of the tenets of the Common Core Standards is the promise of evidence-based goals. With equality in standards and the promise of evidence of learning comes the promise of equitable and efficient measurement tools and systems.

In the wake of the standards movement, educational researcher Thomas Guskey (2009) argues that school leaders must be responsible for translating standards into specific classroom experiences that are built to guard against the historically ineffective assessments that did not measure authentic or meaningful learning. Rick Stiggins (2007), another researcher and proponent of assessment reform, agrees that with the adoption of learning standards schools have become obligated to change the ways in which they report student achievement.

In addition to the surge in interest in grading practices being linked to *what* students are learning, some evidence suggests that it may also be linked to *how* students learn. McMillan & Workman (1998) argue that advances in research to support different learning theories invite conversations around the ways in which we measure and assess student learning.

Neuroscience, for example, has created an “unprecedented revolution of knowledge about the human brain, including how it processes, interprets, and stores information” (Sousa, 1998, p. 52). Knowing now about the relationships between the initial acquisition of content and skills and the extensive shaping phases and subsequent practices that are necessary for complete understanding (Marzano, 2001), educators are encouraged to shift their measurement efforts – where assessment was once fixated on highlighting the quality of outcomes, attention is now

given to the manner and development of learning. And where isolated skills and facts were often regarded as the terminal desirable outcome, the integration of skills and the application of knowledge are now recognized as the more meaningful goals of learning (McMillan, 1997). In another example, Grant Wiggins and Jay McTighe (2005) argue that a sound curriculum starts with long-term goals based on “meaning-making” experiences that result in the student’s ability to seamlessly transfer actual and true learning from one situation to another. This concept is in contrast with the more traditional goal of using grades to measure how effectively a student performs a singular, de-contextualized skill. With this focus on the transfer of learning and with the use of new measures of learning come a new focus on how teachers come to view student achievement and, ultimately, on how that level of achievement and understanding is being communicated.

### The Purpose of Grades

There is very little debate among experts in the field of educational measurement as to what a grade is; most researchers and educational leaders maintain that a grade is intended to be a communication tool used to provide information to stakeholders with regards to student achievement, and some argue additionally that it reflects the qualities of “meaningfulness, explicitness, and fairness” (Anderson, 2003, p. 149) Though “stakeholders” is not fully defined by each researcher, collectively the term can refer to students, parents, teachers (current and future), administrators, guidance counselors, college admissions clerks, community members, and employers (current and future).

It seems well-known that at an administrative level, the purpose of grades often includes employing them in making decisions regarding student promotion or retention. However, in addition to this common use, in some school systems grades also inform course decisions, such as whether a student will advance to an Honors course or be scheduled into an at- or below-grade level course. A study by Dornbusch (Stanford University, 1994) found that in schools where tracking systems were being utilized, grades from a previous class or course were the factor most likely to determine a student's tracking placement. Additionally they found flaws in tracking systems that revealed students of equal abilities being placed into both high and low tracks, with the equal ability students in the low tracks learning less and subsequently earning lower grades than their equal counterparts in the higher tracks. These findings expose the likely "high stakes" nature of the use of grades in administrative decision-making in that they potentially result in a long-term impact on students' educational opportunities.

Colleges rely on grade point averages (GPA) as one of the primary factors in admissions; according to The College Board (2011), a student's high school grades are "the most frequently used predictors in college and university admission decision." Likewise, for students preparing to graduate, scholarships and grade rank are directly linked to the average of grades assigned in some or all courses, and guidance counselors in working with large numbers of students often rely on the GPA when counseling students on post-high school plans. It is not difficult to find examples of school boards promoting the use of grades for the purpose of student rankings; Marzano (2009) reports that a 1992 Austin and McCann study found 44 percent of school board documents emphasizing this purpose despite the fact that it is difficult to find leaders or

prominent organizations in the educational field who support the use of grades to this end. The National Association of Secondary School Principals (2010) found that while grade point averages can be illustrative of a students' performance in comparison to his same-school peers, grades do not reliably inform a university about an individual student's actual ability, achievement, or even course mastery. In the same breath, the NASSP warns schools against creating a culture where grades are used to set students against one another in competition, and instead encourages schools to recognize "academic excellence in a spirit of cooperation."

Educational activist Alfie Kohn (1999) takes a different approach to responding to these administrative purposes for grades. He argues first that "college admissions is not as rigid and reactionary as many people may think" (p. 41), offering that there are myriad non-grade-related factors that colleges are willing to consider in the application and acceptance process. And then he issues the challenge to schools to not "be dragged down to [the] level" (p. 42) of the colleges that resolve to rely on an admissions process that focuses too heavily on high school grades. Others agree with Kohn's "multiple measures" stance. In 1995 the National Council on Measurement in Education refined and published their "Code of Professional Responsibilities in Educational Measurement" where they warn that educators have a "professional responsibility to use multiple sources... [when] making educational decisions."

Generally speaking, teachers agree with the idea that a grade should primarily be a representation of student achievement and nothing more (Allen & Lambating, 2001). Robert Marzano reported in *Transforming Classroom Grading* (2009) that studies have consistently shown support for this particular purpose for grades. Called the "information function," using grades to provide feedback can result in significant increases in student achievement. In this

same text, Marzano reports that researcher John Hattie (1992) found that “providing students with specific information about their standing in terms of particular objectives increased their achievement by 37 percentile points” (p. 9). (While this argument does support the “information function” of grading, it is not an argument exclusively in favor of the use of grades, per se, as Hattie reports that the use of effective written or oral feedback can serve this same purpose.)

The practice of using grades as a tool to communicate student achievement against pre-established criterion, is often referred to as mastery learning, or criterion-referenced grading. In theory, this system offers the highest available grade to all students since no processes such as curving or comparison ranking could exist in a system where the emphasis is on a measure related solely to what a student knows and is able to do. Educational researcher James McMillan (1998) suggests that a desired symptom of grading for the purpose of communicating a level of mastery in a non-competitive environment is a learning culture based on positive relationships among students and teachers.

In addition to administrative purposes and information functions, some may pose that the purpose of grading is to motivate students. McMillan (2008) agrees – to an extent. He contends that two distinct methods for considering grades and motivation can be considered. In the first construct, grades as extrinsic rewards motivate students, but only to gain the reward or avoid the punishment. In this case, the motivation is short-term and fallacious in that the student’s interest and dedication to learning are both disregarded and also diminished. However, when the grade reinforces what the student believes to be true about his increasing knowledge and understandings, he tends to view this feedback experience as a motivation and

incentive to continue to engage in challenging tasks and to view learning with positive dispositions.

Jerome Bruner (1966) suggested that “the will to learn is an intrinsic motive, one that finds both its source and its rewards in its own exercises” (p. 127). Educators who agree that student motivation comes from both the act of learning itself and the satisfaction with having learned will also agree that school systems that de-emphasize these personal attributes will similarly de-emphasize a student’s motivation to learn. Author and middle school teacher Paul Barnwell (2010) contends that systems that rely on motivation to come from the inauthentic practices of grading student learning ultimately ignore what psychological research has discovered – that people in general are not motivated in the long-term by arbitrary rewards and punishments. Though incentives and penalties may work short-term, they ultimately actually suppress the natural desire to learn and create a climate where students disengage with future learning interactions where rewards are not offered. Ultimately, the learning becomes driven by external motivations, and the potential for a life-long commitment to and desire for learning are lost.

Behavioral scientist and author Daniel Pink reports in his book, *Drive: The Surprising Truth about What Motivates Us* (2011) that when grades are used primarily as motivators, they run the risk of becoming the goal itself. Evidence in social science, he tells us, identifies the difference between the “performance goals” –getting a good grade – and “learning goals” – increasing our abilities and understandings; when the emphasis rests on the performance and the goal of rewards, the natural and intrinsic motivation eventually collapses. The question Pink offers to educators is this: Do we want classrooms where students learn to respond to our

tasks because they can secure some sort of external reward for doing so? Or do we want students who are “working hard and persisting through difficulties because of their internal desire to control their lives, learn about their world, and accomplish something that endures?” (p. 79).

### Student Confusion

Despite laws requiring school districts to adopt and publish their grading policies each year, students often report being confused as to how they get the grades that they are assigned. In 2009 the Southeast Comprehensive Center at SEDL found that even though just nine states in the United States maintained state-wide uniform grading practices (Arkansas, District of Columbia, Florida, Georgia, New Mexico, North Carolina, South Carolina, Tennessee, and West Virginia), many of those that did not still attempted to define the purpose of grades. The Texas Education Code, for example, requires districts to develop policies that require classroom teachers to assign grades that “reflect the student’s relative mastery of an assignment” (District Grading Policy, 2009).

And yet Aaronson et al. (1994) report that students are more likely to believe that grades are part of the “game of school” than a result of effort, achievement, or mastery. According to their study, students spend a great deal of their time “teacher-pleasing,” which includes behaviors such as trying to determine the answers that the teachers want to hear (rather than focusing on answer that makes the most sense) and determining the behaviors that the teachers will consider acceptable. These behaviors ultimately interfere with “genuine student intellectual, social, and moral growth” (abstract) since students who are focused on satisfying a teacher’s personal sense of self often become competitive with each other,

distracted from the process of learning, and unable to take risks. As early as 1971, researchers were discovering that students believed grades to be something “given” to them by teachers as opposed to reflections of their learning (Kirschenbaum, Simon, & Napier).

Students also sometimes report being more interested in the actual score or grade than in how much they actually learned. Because the act of assigning grades essentially introduces a conflict between what is real and what is perceived achievement (Bonesronning, 2004), students can easily come to care more about how their teachers perceive them to have achieved than how they actually achieved. Correa and Gruver (1987) call “perceived achievement” the result of actual achievement *plus* the teacher’s grading parameters. This could have either a positive effect or a negative effect based on whether or not the teacher’s grading structures are too difficult (grade deflation) or too easy (grade inflation). This shows that while the purpose for grades may be to communicate, the actual existence of grades creates the potential for this teacher “power” to actually affect student achievement (Bonesronning, 2004). In essence, these power struggles – between the perceived use of grades and their actual use – defines the idea of the “politics” of grading. If teachers use them to control behaviors and advance personal beliefs, then the system of grading takes on many of the characteristics of other political systems.

### Token Economy

The “grades as pay” mentality represents another key use of grades -- as a measure of student behavior (Brookhart, 1991). Teachers may see grades as something that students earn for behaving in such a way as to reflect the social order that the teacher deems to be

appropriate in the climate of her classroom. There is evidence that grades have become, in many schools, a sort of “academic token economy,” functioning as a primary method of maintaining acceptable behavior in the classroom (Brookhart, 1993). In these schools, the completion of work is as much of a “behavior” to be rewarded or punished as talking out of turn or being late to class would be. In these cases, teachers are recording student behaviors and lowering grades based on whether or not they see the student as behaving in a way deemed acceptable – based either on the teacher’s moral viewpoints or on classroom climate expectations set by the teacher.

Some teachers report developing a “grades as pay” classroom in order to reward those students who “work really hard” (NCTE High School Matters, 2008). When teachers feel the need to acknowledge effort and practice over measuring actual achievement, the result can be a system where a grade reports more about effort, attendance, tardies, and attitude than it does about achievement. Marzano (2001) reports on meta-analyses that confirm an authentic relationship between effort and achievement; the argument then can be made that if student efforts are directed appropriately at skills that support the desired learning then achievement will be the ultimate gain. Releasing the “grades as pay” scheme, however, relies on faith that achievement will ultimately reflect effort and therefore effort does not need to be recorded separately or rewarded outside of the context of achievement.

One sixteen year old blogger (TwentySided, 2011) writes that he “hates it” when teachers say, “Just make sure you do all the work, and you *will* pass my class.” His interpretation of this statement is that “the bulk of [his] grade will come from *doing* things, not from *knowing* things.” He goes on to say, “Some teachers even go so far as to grade the notes

we take in class. This is infuriating to me. In the past I saw school as this perfectly arbitrary trial of mysterious activities. Now I see it as a house of incompetents. Our goal is ostensibly to learn things, but the system of rewards and incentives is often completely divorced from this idea, and sometimes even runs counter to it.”

### Broken Grades

In their *Respite for Teachers: Reflection and Renewal in the Teaching Life*, Pearson-Casanave and Sosa (2007) write: “How strange that our practices encourage students to believe that a number or a letter grade...is the primary evidence by which they and others know or don’t know something” (p. 14). Others in the educational arena concur that herein lies the problem with grading as it is used in current practice – that is has taken over our school systems and has become the primary focus of student attention in academia. The process of scoring work, of recording grades, and of reporting out standings and rankings provides students, parents, and other community members with the impression that teachers and students alike are functioning as intended. Our ability in the field to seemingly quantify the learning processes has provided much assurance to the community that we have ample evidence of student growth or lack thereof (Pearson-Casanave & Sosa, 2007).

However, where grades are intended and expected to be communicators of achievement, they include, instead, a great deal of other information that gets in the way of the intended purpose. In this effect, they are “broken” (O’Connor, 2006). Clymer and William (2007) claim that “our current grading practices don’t do the one thing they are meant to do which is to provide an accurate indication of student achievement” (p. 36). Studies have shown

that teachers include too many non-achievement factors in students' grades. Cross and Frary (1999) show evidence of attitude and effort being included in grade reports. Other research shows that schools have been found to additionally include attendance, student growth, and behavior in grades (Brookhart, 1994; Feldman, Kropf, and Alibrandi, 1996; Robinson and Craver, 1989). Aaronson (1994) suggests that these wide-ranging practices have "extinguished" the potential for students to recognize the power of learning as the primary and more natural motivating force, more significant and more consequential than grades.

Furthermore, study of our current grading practices reveal that a teacher's grading habits may change when student demographics change. Students in low SES (socio-economic status) schools or in schools where there is a distinct teacher control atmosphere receive significantly different grades than students with comparable abilities in contrasting schools. The National Education Longitudinal Study published in 1988 revealed that students in schools situated in poorer communities were likely to experience grade inflation. Where a student would earn a B or C in a low SES school, a similar student performing similarly (as measured through standardized achievement tests) would receive a D or F in a school situated in a more affluent community. Conversely, in schools with more compliant students -- typically more affluent schools -- teachers are less likely to rely on non-achievement factors than in schools that are considered to have a more custodial-oriented control. In the more controlling environment, teachers were found to be using significantly more non-achievement facets in determining grades (Brantlinger, 1993). A similar study found that students with the highest capacity for learning are often graded solely on achievement whereas less-abled students are

offered the benefit of being graded also on effort and disposition (Stiggins, Frisbie, & Griswold, 1989).

Others (Kohn, 1999, O'Connor, 2009) have found that it is not just in establishing and applying the purpose for grades that we falter, but in the methods and systems we use to compute and report them. Seemingly, grades become unreliable as communicators when they are recorded and averaged using mathematical processes instead of teacher interpretations and thoughtful reflections. For example, when grades are averaged, the whole of the student's learning experiences throughout the grading period or unit of study become equally weighted. This causes grades recorded when the student's understandings were immature to be weighted against grades recorded toward the end of the grading period when the student had gained a more complete understanding. If grades are to reflect what students know and are able to do – as opposed to the extent to which they didn't already know the content at the beginning of the unit– then grades would not be averaged to include evidence prior to having opportunities to fully learn and understand the concepts.

Richard Francis (2006), a measurement and statistics professor at California State University, suggests that current practices of grade averaging are flawed since each grade is typically weighted equally, despite the fact that multiple factors go in to each grade. For example, a teacher may record a unit test with 30 questions in the gradebook and then later record a unit test with 45 questions in the gradebook. These two tests are then averaged, perhaps along with other grades, to determine a final grade. The teacher may have overlooked, however, additional skills or knowledge that were required in the former test, if indeed its greater length represents more skills being measured. On the other hand, a test with only 20

questions could be a summative measure of two separate skills, while a different test with 20 questions might be a summative measure of one skill. The first test should be counted twice; otherwise, the three measured skills are not equally represented in the final grade.

O'Connor (2006) reports on the often overlooked effects of the inclusion of zeroes in grade averages. Where the difference between an A and a B or a B and a C is a range of 10 points, the difference between the highest failing grade, a 69, and a zero is 70 points. Mathematically, a student's zero is equal to seven failing scores (Francis, 2006), what he argues to be an unbalanced punishment for missing or late work. Arguably, then, in a system where final grades are determined through an averaging process, a zero is only a fair grade when, in fact, the student is completely incapable of completing any portion of the assignment. In that case, the reported and recorded zero is, in fact, a communication of student achievement – it is saying, in essence, that a student has a zero percent mastery on a particular learning goal. Rick Wormelli (2006) puts it this way: "A zero has an undeserved and devastating effect on students and their grades – so much so that no matter what the student does, the grade distorts the final grade as a true indicator of mastery. Mathematically and ethically, this is unacceptable" (p. 12).

Rick Stiggins, Director of the ETS Assessment Training Institute in Oregon argues (2007) that one issue with current grading systems – what he calls "The Assessment Experience" – is that our current grading practices rely exclusively on assessment "that verify learning" as opposed to those that "support learning." When grading comes at the end of the learning experience, it merely measures what students did well and what they did not. For students who are not successful, traditional grading is merely evidence of failure, leaving the student to

believe that feedback is simply criticism, that frustration and fear are normal in learning systems, and – most importantly – that teacher feedback comes too late to do anything about it. A process of “assessment *for* learning” where feedback is primary and “just in time” to correct mis-steps, re-learn as needed, and learn at a personal pace would enhance and not just monitor student growth, and it would replace a “winners and losers” system. Kings College professors and researchers Paul Black and Dylan William (1998) agree that giving grades favors the summative assessment model where recording data does not have a viable impact on on-going learning. Evidence they found supports an approach that focuses on formative assessments, which would “restore the balance” the assessment and learning experience.

### The History of Grading Practices

Formal grading systems as we know them today are a rather recent phenomenon in United States education (Guskey, 2001). The one-room school houses that existed in and before the 18<sup>th</sup> century relied not on “grades” but on a system of feedback by teacher comments reported orally to students and parents (Rich, 2001). By the late 1700s, it is said that word had reached both Harvard and Yale Universities of a professor at Cambridge University in England – one William Farish – who had increased his class size and thusly increased his salary by creating a system of grading that would allow him to by-pass the narrative feedback by adopting the much more efficient system of “grading,” much like the one being used in factories to measure the effectiveness of the assembly line workers (Pearson-Casanave & Sosa, 2007). Instructors at Harvard and Yale adapted this new system and respectively implemented a four-point scale and a 100-point scale by the late 1800’s (Marzano, 2000). Shortly after,

Mount Holyoke College in Massachusetts converted these numerical systems into letter grades for ease of reporting (Durm, 1993).

Though built under the guise of efficiency, it wasn't long before researchers testing these new systems deemed them to be unreliable; as early as 1911, researchers Daniel Starch and Edward Charles Elliott were arguing that teachers' previously used narrative feedback responses improved students' subsequent work whereas the use of numerical and letter grades merely served the purpose of sorting students and their abilities (Guskey, 2001). In studying grading in English, they found that some teachers would focus on certain skills – such as grammar and neatness – where others would focus on different skills – such as content; in studying grading in mathematics, they found even greater disparities, with some teachers deducting for neatness, form, and spelling and others responding only to correctness (Pollio, 2000).

In 1913 Cornell doctoral student Isidor Edward Finkelstein published a short text entitled, *The Marking System in Theory and Practice* in which he wrote that “school administrators have been using with confidence an absolutely uncalibrated instrument...” (p. 1) and he argued that the “variability in the marks given for the same subject and to the same pupils by different instructors is so great as frequently to work real injustice to the students” (p. 6). Even at this early date, educators were questioning if “marks should indicate performance or ability or accomplishment” (p. 11).

Despite the criticisms of grade use and subsequent failed attempts to create systems that were as efficient as grades but as effective as feedback, by the 1940s, almost all large-scale educational institutions in the United States had begun using either a number or letter grading

system; consequently, a 1998 study by the College Board found that of over 3,000 contemporary U.S. schools surveyed, 91% used some sort of A-F or numeric grading system (Marzano, 2001).

### Expert Recommendations

For as much criticism as is published on the problems with grading, an equal number of recommendations for fixing grading problems can be found. At the elemental level, Allen and Lambating (2001) recommend a dedication to informing teachers as to how to make grading decisions that are based on fundamental measurement principles. They recommend preparing preservice and inservice teachers to develop effective methods to assess students and to communicate clearly and accurately through their grading practices. In essence, their recommendations are grounded in the belief that teacher training can change the trajectory of current practice.

Other recommendations for creating a standard for effective grading practices come from a position document created jointly in 1990 by three professional education organizations – The American Federation of Teachers, the National Council on Measurement in Education, and the National Education Association. These recommendations are formed as standards called the “Standards for Teacher Competence in Educational Assessment of Students.” These standards focus on classroom-based competencies and on competencies connected to broader-scaled assessment decisions like those made at the school, district, state, or national levels. These standards direct teachers to be skilled in assessment methods – choosing, developing, administering, scoring, and interpreting appropriate assessment methods – and in grading

methods – developing and communicating grading procedures. By creating these standards these organizations acknowledge the importance of grading in the teaching and learning process, and in the introduction of the standards, the organizations suggest that training “to develop the competencies covered in the standards should be an integral part of preservice [teacher] preparation.”

A grading practice advocated by Virginia Polytechnic Institute and State University Professor Lawrence H. Cross (1995) includes using professional judgment to make sense of letter grades instead of relying on quantitative measures. Cross believes that teachers must make informed choices about final grades based on more information than converted raw scores. In test construction practices, he recommends varying the level of cognitive difficulty in test questions so as to discriminate among all skill levels (Cross, 1995). Also supported in Cross’s research is the practice of determining minimum passing scores on tests by working collaboratively within a department to identify test items or performance characteristics that are considered absolutely essential. These items then represent the performance standard, so that answering a large majority of these items correctly is necessary to receiving a passing grade. These steps, along with using professional judgment to determine cut points between grades will result in fairer score reports and increased student efficacy (Cross, 1995).

At a more specific level, researchers recommend that administrators and teachers have repeated conversations on the topic in order to agree up on what a grade is and reflect on what goes into a grade that is reported out to students and parents. Along these same lines, teacher verbiage should reflect an awareness of the differences between these two terms -- grades and achievement (Carlson, 2004). This is an important step in changing the culture of grading

practices in schools. When a teacher assigns or collects student work, she should be careful to indicate whether the work will be assessed or assessed *and* graded. Equally, teachers need to be diligent in communicating the purpose of work. Ultimately, the work is intended to lead to learning, not lead to a grade.

Another grading recommendation is to have new evidence of learning replace old evidence of learning so that the grade will more likely reflect student achievement (Wiggins & O'Connor, 2005). If a student is assessed on his knowledge of state capitals four times, his last grade would likely be the most accurate, given the fact that over time he has practiced memorizing the answers. But if his grades are averaged together -- e.g., 50, 70, 80, 99 -- his averaged score would reflect all of the grading checkpoints completed by the teacher (he would have a 75%) instead of his actual level of mastery (of 99%). Another student who may have started with more knowledge or one who learned more quickly would have a higher average, despite the fact that by the time the teacher measured the skill, they were all equal in their achievement (Wiggins & O'Connor, 2005). This practice of averaging representative knowledge and understanding over time rewards students for pre-requisite knowledge and for learning quickly; correcting these misuses requires reflection on the purposes of grades. Teachers must be able to defend their position—are we rewarding students for learning quickly or are we communicating an ultimate measure of their achievement in a given time period?

The process of trying to convince a campus administrator to engage in the practice of changing teacher behavior, principal Jeffrey A. Erickson (2010) argues, is much like trying to convince a politician to attempt restructure of the social security process – both are a “third rail” – “if you touch it, you’ll die” (p. 22). All the same, he offers two important

recommendations for moving toward reformed grading practices. The first he garners from professor and researcher and founder of the “Leadership and Learning Center,” Doug Reeves who warns against falling prey to the “buy-in” myth. According to Erickson, Reeves warns that the “wait for buy-in can be interminable” (p. 23); instead, he offers that schools and districts should adopt the “vision / action / buy-in / action” model, which relies on the effects of action to create understandings as opposed to waiting for buy-in before enacting change. Erickson also poses a solution to issues of academic integrity. Where cheating and missed assignments were once simply recorded as zeros, his plan provides for alternative interventions for students – including required ethics work and lost privileges – that can result in more accurate grade reporting *and* solutions to cheating and complacency issues.

Ken O’Connor’s *15 Fixes for Broken Grades* (2009) offers additional recommendations for grading practices, including teachers learning how to separate achievement from behavior in grade reporting through the use of better designed reporting systems, how to handle group work so that group grades are not part of individual students’ achievement grades, and how to avoid zero grading while also developing a system of accountability. McTighe and Tomlinson (2006) offer recommendations to create a classroom experience where “reporting practices can be a natural extension of a rich, differentiated curriculum and a seamless part of the instructional practice” (p. 128). They recommend grading only on “clearly specified learning goals and performance standards,” and avoiding factoring in formative assessments into final grades (p. 131). Guskey and Bailey (2001) pose these four schemes for successful grading – to keep communication as the primary goal of grading, to focus on grading as part of the process

of instruction, to rely only on solid evidence for grading, and to use a grade reporting system that allows teachers to do more than simply report a singular grade.

### The Politics of Grading

Recommendations on grading practices that are effective, fair, and valid -- in other words, “best practice grading practices” -- have been provided to teachers in various forms since the turn of the 20<sup>th</sup> century (Cross & Frary, 1999), and yet evidence shows that despite these clear cut recommendations, teachers are not adjusting their grading practices (Stiggins & Conklin, 1992). So why don’t teachers’ practices reflect research and understandings about learning, engagement, and motivation? What is impeding teachers from acting on these recommendations?

Stiggins, Frisbie, and Griswold (1989) suggest that teachers may find that the “real world” of teaching that has been established in our nation’s schools cannot accommodate these solutions which they find to be impractical. Recommendations like standards-based grading are big changes from current practices like entering the percentage correct on several assignments and averaging. Not using this comfortable and familiar method could mean having to find new ways to report learning in a school or district where report cards have a standard format and where parents expect grades to be reported in a way that is familiar to them.

Additionally, teachers may not consider using the grading advice because they lack the training or expertise to implement the recommended practices (Nitko, 1991). Where measurement specialists may fully explain how grades should come together as a symbol of achievement, teachers may not be able to connect the instructional experiences and curriculum

plans they have in place with these new ways to measure. The National Council on Measurement in Education (2010) points out that it is unlikely that teachers have received the training necessary to understand how grading and assessment works best. Those who are skilled in the field of psychometrics are often the ones asked to teach courses on measurement in teacher education programs; psychometricians may view the use of data differently from the ways that teachers might typically view it. Where the psychometrician is historically trained in looking at statistics and trends in large-scale assessments, the teacher needs information that is immediately applicable to his small classroom environment.

Frary (1992) reports that teachers often directly pursue practices that work in direct opposition to those recommended by educational measurement specialists. Teachers may specifically choose to not allow grades to completely represent achievement, and they may do so because of an agenda built around student nurturing and student advocacy (Brookhart, 1991). Teachers may intentionally “use non-achievement factors in determining grades to mitigate negative social consequences associated with what they consider to be inappropriate use of grades, such as determining eligibility for nonacademic privileges either at home or school” (Cross & Frary, 1999). Teachers may also be concerned that if grades are only a measure of achievement, then low-achieving students will be discriminated against based on their grades. They may believe that there are social consequences for grades as well as efficacy issues that could have long-term effects on students as learners (Brookhart, 1991). In other words, teachers are concerned that true and actual grades could damage a student’s sense of control over learning and result in his failing to attempt to accomplish a task that feels daunting, out of reach, or a task where he feels he has already tried his best and has failed.

Bishop (1992) argues that “teachers can’t be coach and judge” (p. 16). It is the mentor-like relationships found infrequently in schools – coaches, band conductors, debate team sponsors, yearbook advisors, etc. – he argues, that allow teachers to be most effective because they are long-term and ultimately the summative assessment is provided by someone else. The conflict for teachers as graders emerges through the incompatibility of the two identities teachers otherwise have to have – the advocate for the student *and* the judge of his work – and results in grading that is based on other factors. This evidence suggests to Bishop that external measures of success could play a role in measuring student learning and free teachers from having to take on the role of judging those whom they wish to support and nurture.

Along these same lines, teachers have also reported that the use of non-academic factors in grades typically raises the students’ grades, which helps the teacher avoid scrutiny from parents, students, and administrators who may otherwise accuse them of having excessive failures (Cross & Frary, 1999). Trog and Friedman (1996) conducted a study in which they analyzed the grading policies used by high school teachers. They concluded that many teachers’ grading practices are not the result of lack of understanding or conflicting professional roles; they are, instead, simply reflective of the practices that teachers believe are expected of them by parents and students. For example, prior to students reaching high school classrooms, they have had myriad experiences with grading practices that include averaging, the inclusion of nonacademic grades, and allowing behaviors – good and bad – to influence a final grade. The high school teacher doesn’t want to be the one to have conflict with parents and students over practices that seem new and radically different from the student’s

elementary and middle school experiences or from the experiences the parents had when they were students.

APPENDIX B  
DETAILED METHODOLOGY

## Strategies of Inquiry

As the variety of research questions comprising this study suggested a need for an equally diverse assortment of approaches to data collection and analysis, an array of strategies of inquiry were employed. As a whole, this work is best described as mixed-design in that some research questions were satisfied through quantitative only methods while others required mixed methods designs – i.e., qualitative and quantitative strategies used simultaneously.

### *Research Questions 1 and 2: Quantitative Methods*

To capture the current conditions of teachers' beliefs about grading and self-reported grading practices, a quantitative survey method was used. Specifically, a sample, cross-sectional survey in the form of an on-line questionnaire collected responses from participants with regards to both research question 1 (beliefs) and research question 2 (self-reported practices), as they both related to Part A of the research questions – the overall grading system – and Part B of the research questions – assignments graded.

### *Research Question 3: Quantitative-Qualitative (Mixed Methods)*

To collect data with regards to teachers' actual practices, an interview/portfolio analysis process to analyze teachers' grading records was utilized; this process employed a mixed-methods design for the purposes of improving the accuracy of the data and ultimately the results and conclusions of the study.

This mixed methods design was in the form of a standardized, open-ended interview regarding teachers' actual grading practices as recorded in their grade books; the wording of each interview question and the sequence in which all interview questions were asked was pre-

determined based on the results from Survey 1. These qualitative responses were then quantified through a process of coding, grouping, and summing.

*Research Question 4: Quantitative-Qualitative (Mixed Methods)*

Once it was determined that a discrepancy existed between beliefs, self-reported practices, and/or actual practices, a sample, cross-sectional survey in the form of an on-line, combination fixed-response/open-ended questionnaire was designed and implemented to address research question 4. The fixed responses were derived from pilot sample responses, and respondents in the sample had the additional opportunity to explain their responses or record further answers and ideas. These survey responses were coded, grouped, summed, and analyzed for interpretation and reporting purposes.

*Research Question 5: Quasi-Experimental Design*

In response to research question 5 – the cause-effect relationship between teachers' responses and order of survey questions – a quantitative strategy was used with a quasi-experimental, counterbalanced design. This analysis of variance between Treatment Group A and Treatment Group B was conducted to determine if there were limitations in the data created by the order of response to the questions found in Survey 1. Specifically, the aim was to determine if participants who responded to questions about their own actual grading practices were affected by the knowledge that this survey was also asking questions regarding their beliefs about best grading practices.

In this research design, instrumentation was manipulated in such that Treatment 1 (Treatment Group A) and Treatment 2 (Treatment Group B) varied with regards to which part of

the survey instrument was received first – either questions regarding research question 1 (beliefs) or questions regarding research question 2 (self-reported) practices.

### Instrumentation

Because this study is grounded in matters that are directly relevant to the participants in their daily professional lives – i.e., their professional beliefs about best practice and their self-reported and actual practices – the challenge regarding the design of instrumentation was to construct a method of soliciting responses that capture participants’ most honest and complete ideas while not causing them to feel personally judged or distressed.

An overview of the types of instrumentation used to meet the requirements of the research questions is presented in Table B.1.

#### *Survey 1*

Survey 1 (Appendix A), called “Survey on High School English Teachers’ Grading Practices” was designed and created for this research by me personally and served several purposes. First, it addressed research question 1 (What do high school ELA teachers believe a grade should measure?), both Part A (grading systems – *how* grades should be recorded and reported) and Part B (graded assignments – *what* should be graded and recorded). Additionally, it addressed research question 2 (What do high school ELA teachers self-report their students’ grades measure?), both Part A (grading systems – *how* grades are recorded and reported) and Part B (graded assignments – *what* is graded and recorded).

Table B.1

*Summary of Research Questions, Strategies of Inquiry, and Instrumentation*

Research Question	Strategy of Inquiry	Instrumentation
<p>1 – What do high school English-language arts teachers believe a grade should measure?</p> <p>Part A: ...with regards to grading systems, i.e. <i>how</i> should you grade, record, and report?</p> <p>Part B: ... with regards to assignments graded, i.e., <i>what</i> should you grade, record, and report?</p>	Quantitative Qualitative	Survey 1 – Sample, cross-sectional survey in two forms (1.2, 2.1), via on-line questionnaire, with Likert-scales and Likert-type items
<p>2 – What do high school English-language arts teachers self-report their students’ grades measure?</p> <p>Part A: ...with regards to grading systems, i.e. <i>how</i> do you grade, record, and report?</p> <p>Part B: ... with regards to assignments graded, i.e., <i>what</i> do you grade, record, and report?</p>	Quantitative Qualitative	Survey 1 – Sample, cross-sectional survey in two forms (1.2, 2.1), via on-line questionnaire, with Likert-scales and Likert-type items
<p>3 – What do high school English-language arts teachers’ grades actually measure?</p>	Quantitative- Qualitative (triangulation, mixed methods design)	Portfolio Analysis via Interview: analysis of grade book records, combined fixed-response and open-ended response
<p>4 – What do teachers perceive to be the causes for the discrepancies between what high school ELA teachers say a grade should measure, and what the grades actually measure?</p>	Quantitative- Qualitative (triangulation, mixed methods design)	Survey 2 – Sample, cross-sectional survey, via on-line questionnaire, with combined fixed-response and open-ended response
<p>5 – Do teachers’ self-reports on their actual grading practices change significantly when they are first asked to report on their personal beliefs about best grading practices?</p>	Quantitative (quasi-experimental design)	Use of Survey 1 forms 1.2 and 2.1 results

Furthermore, because Survey 1 was designed in two forms – 1.2 and 2.1 – it provided data for the analysis of variance of the means conducted in response to research question 5 (Do teachers’ self-reports on their actual grading practices change significantly when they are first asked to report on their personal beliefs about best grading practices?).

Survey 1 was also designed to collect demographic information on the sample so that it could be described thoroughly.

Finally, this survey offered participants the opportunity to volunteer to participate further in the interview/portfolio assessment designed in response to research question 3 (What do high school English-language arts teachers’ grades actually measure?).

This survey was comprised of twelve questions, outlined in Table B.2 and delineated by question number, topic and content description, and the research question(s) addressed.

The bases for these best practice statements used in Survey 1, Question 3 are recommendations on grading practices identified in the following five expert sources:

*Assessment for Learning* by Black, Harrison, Lee, Marshall, and Wiliam (2003); “Seven Practices for Effective Learning” by McTighe and O’Connor (2005); “Effective Grading” by Reeves (2008); “Grading for Success” by Tomlinson (2001), and *A Repair Kit for Grading: 15 Fixes for Broken Grades* by O’Connor (2007). Support for the use of these authors as experts in the field is provided in Table B.3.

Table B.2

*Design of Survey 1 by Topic/Description and Research Question Addressed*

Question	Topic/description	Research question(s) addressed
Question 1	Agreement to participate	None
Question 2	Definitions used in survey	None
Question 3	Personal beliefs about grading: 15 general statements about grading systems	Research question 1, Part A
Question 4	Personal beliefs about grading: 16 assignment types	Research question 1, Part B
Question 5	Self-reported grading practices: 15 general statements about grading systems	Research question 2, Part A
Question 6	Self-reported graded practices: 16 assignment types	Research question 2, Part B
Question 7	Familiarity with professional organizations	Participant characteristics
Question 8	Familiarity with educational authors, instructors, and researchers	Participant characteristics
Question 9	Survey respondent information	Demographic information
Question 10	Current teaching assignments	Demographic information
Question 11	Previous teaching assignments	Demographic information
Question 12	Request for extended participation	Research question 3

Table B.3

*Evidence of Expert Status*

Author/Researcher	On-line biographies
William, Dylan	<p>Author of seven books on assessment and learning            Dean of the School of Education, King’s College London            Degrees in mathematics, math education, and PhD University of London            Deputy Director, Institute of Education, University of London            Emeritus Professor of Educational Assessment, IOE            Senior Research Director, ETS, Princeton NJ</p>
McTighe, Jay	<p>Author of 13 books on assessment and curriculum development            Degrees in education from the College of William and Mary, University of Maryland, and Johns Hopkins University            Director of the Maryland Assessment Consortium            Director of the Instructional Framework, database on teaching            Member of the National Assessment Forum            Selected participant in the Educational Policy Fellowship Program through the Institute for Educational Leadership</p>
O’Connor, Ken	<p>Author of 10 books on assessment and grading            Consultant on Secondary Assessment for the Ontario Ministry of Education            Curriculum Coordinator for Student Assessment and Evaluation for the Scarborough Board of Education            Degrees in education from the University of Melbourne and the University of Toronto            Staff development presenter on assessment and grading in 43 states and 16 other countries</p>
Reeves, Douglas	<p>Author of 30 books on educational leadership and assessment            Twice named to the Harvard University Distinguished Authors Series            Distinguished Service Award from the National Association of Secondary School Principals            Recipient of the National Staff Development Council’s Contribution to the Field Award            Founder of the Leadership and Learning Center</p>
Tomlinson, Carol Ann	<p>Author of over 200 articles, books, book chapters, and professional development materials            Co-Founder &amp; Co-Director of the University of Virginia Best Practices Institute            Educational consultant to the Corporation for Public Broadcasting            Named Outstanding Professor at Curry School of Education            Professor and Chair of Educational Leadership, Foundations, and Policy – University of Virginia</p>

From these recommendations, fifteen general statements on grading systems were developed for this study – for example: “A teacher should adjust a student’s report card grade if the final calculation does not accurately reflect the teacher’s personal knowledge of the student’s success in the class.” Survey respondents were asked to indicate the degree to which they believed these statements to be best practice by selecting a level of agreement or disagreement on a Likert scale with four points – “3” representing complete agreement, “2” representing general agreement but with some reservations, “1” representing general disagreement but with some reservations, and “0” representing complete disagreement.

For this and in all questions on all survey instruments in this study, these particular ranges of options were intended to capture the intensity of the respondents’ agreement or disagreement but without the option for neutrality; the intentional omission of neutrality is fitting to this study in such that the responses to these items are not hypothetical to the respondents – i.e., they are considering statements that have immediate connections to their current professional lives. Some researchers (Clason and Dormody, 1994) have concluded this deletion of the neutral response is appropriate and does not interfere with the validity of respondents’ selections. Researchers from the University of Toronto and University of Denver (Lamb, Allen, & Green, 2010) found that “informed respondents” (p. 2) – those with professional experience in the field related to the statements of study – had a lesser need for neutral responses, and Krosnick, et al. reported (2002) that the neutral option can actually “discourage respondents from doing the cognitive work necessary to report” (p. 371) truthfully or to the best of their abilities their attitudes toward the topic at hand.

For Question 4 on the survey, also designed to reflect research question 1, respondents were presented with sixteen different assignment types commonly found in a high school English class, such as first drafts of essays, class notes, class presentations, etc. Survey respondents were asked to react to each of these assignments by identifying, based on their beliefs about best grading practices, to what extent each assignment should be considered important when determining what work should be counted toward a student's final grade. Rating options for this part of the survey had respondents select a "3" for "very important; this work should always count toward a student's final grade," a "2" for "somewhat important; this work should usually count toward a student's final grade," a "1" for "somewhat unimportant; this should usually not count toward a student's final grade," or a "0" for "completely unimportant; this work should never count toward a student's final grade."

Question 5 and Question 6 repeated the processes and concepts addressed in Question 3 and Question 4, excepting that Question 5 and Question 6 asked respondents to self-report their own actual current grading practices (research question 2) for Parts A and B, respectively, rather than their beliefs.

The language of the statements from Question 3 to its corresponding Question 5 was adjusted to reflect the first-person point of view needed for respondents to self-report their own practices. For example, whereas in Question 3 a respondent was asked to respond to the third-person statement, "A teacher *should* adjust a student's report card grade if the final computation does not accurately reflect the teacher's personal knowledge of the student's success in the class," in Question 5, that same concept was worded in the first person, as such:

“I sometimes adjust a student’s report card grade if the final computation does not accurately reflect my personal knowledge of the student’s success in my class.”

There was no change in the way that the statements in Question 4 and its corresponding Question 6 were stated for the self-report; however, the language of the rating options was altered. For example, whereas in Question 4 the respondents were asked to consider only their beliefs about grading practices, the level “3” rating was worded as, “very important; this work should always count toward a student’s final grade,” in Question 6 where respondents were asked to self-report on their own actual current practices, the level “3” rating was personalized as such: “very important; I always count this work toward a student’s final grade.”

In Survey 1, Question 7 was designed to describe the participants with regards to their levels of involvement in the following professional organizations: the National Council of Teachers of English (NCTE), the International Reading Association (IRA), and the Association for the Supervision of Curriculum Development (ASCD). Likewise, Question 8 was designed to describe the participants with regards to their awareness of expert educators, authors, and researchers who most greatly influence the field’s understandings and perceptions on grading practices, including the five experts referenced previously and an additional five. Question 10 and Question 11 were also designed for the purpose of capturing participant demographics by way of their current and past teaching assignments, by grade level, and by courses taught.

Question 9, which asked respondents to share their current age, age when they began teaching, years teaching, state in which their school is located, and the school name, was also intended for the purpose of demographic data collection. The school name allowed for the determination of the classification of the school using the classification structure outlined by

the United States Census; a note to the respondent indicating this limited purpose was linked to that portion of the question.

And finally, Question 12, aforementioned, asked participants to consider volunteering for additional participation in this research study. A space for participants to indicate their contact information – specifically, name and email address – was included in this section of the survey.

For the purpose of confidentiality, all responses collected in Questions 9 and 12 were disassociated and are stored separately from each other and separately from responses to Questions 1-8, 10, and 11.

In order to conduct the analysis of variance between the means of the two versions of the survey documents, Survey 1 was designed and released in two formats. Form 1.2 (Appendix A) was designed exactly as described here, with the questions regarding the beliefs on best grading practices preceding those questions regarding the self-report on grading practices. Form 2.1 (Appendix B) follows the design described here as well, excepting that the questions regarding the self-report on actual grading practices precede those on the beliefs on best grading practices.

This survey was designed with two specific logic features – (1) it used a “skip logic” feature that closed the survey if it was determined that the participant did not meet the requirements to be considered part of the population, and (2) it did not allow participants to return to review or change previous answers. This second feature was essential to maintaining data that would be valid for the purpose of studying research question 5 regarding effects of participants’ knowledge of survey questions.

### *Survey 1 – Pilot Study*

Reliability of Survey 1 was tested through a pilot study using a test/re-test model intended to measure the consistency of results over a time period of one week. Six members of the population not otherwise participating in the sample were asked to volunteer to participate in this pilot study. It should be noted here that these pilot study participants were excluded from the sample because they work in the same school district where I am employed as a district-level administrator. I have an indirect supervisory relationship with the participants, and so to reduce any bias in their responses, they were assured that the responses they provided would not be made part of the study and would not be analyzed for content. The language of the text used to invite participation in this pilot study is reported in Appendix C.

Initially, the pilot study participants were asked to complete the entire survey. During this administration, they were also asked to measure the length of time needed to complete the survey. The average length of time reported was 15.3 minutes with a range of 12 minutes to 18 minutes and a standard deviation of 2.50. Given these results, respondents in the sample were informed that the survey would take between 15 and 20 minutes to complete.

One week after the first administration, the pilot study participants were asked to re-take one portion of the instrument – either Question 3 or Question 4, randomly assigned. Data from both administrations were entered into SPSS, and a bivariate correlation was performed (Analyze > Correlate > Bivariate). For Question 3, the coefficient of stability measures for the test and re-test items were: .849, .825, .713. For Question 4, the coefficient of stability measures for test and re-test items were: .876, .841, .800. Though there seem to be a wide variety of acceptable lower limits of reliability measures in survey research, Nunnally and

Bernstein (1994) prescribe values of 0.70 or greater. To this end, the survey was considered reliably stable, but with the understanding that in using a test/re-test method there is a possibility that memories of the initial administration may influence the participants' responses on the second administration especially when a limited amount of time passes between the two.

For content validity, these same six pilot study participants were asked to return to the survey and consider it with regards to structure, item construction and content, instructions, and definitions. They were asked to respond to these four questions: (a) Did you understand what each question was trying to ask you, (b) When you encountered each question, were you comfortable with the instructions, (c) When you began answering each question, were you able to answer the way in which you wanted to, and (d) What other feedback would you like to provide about the survey or your experience taking it?

Several recommendations were offered, and minor changes were made to several items. No significant changes were recommended or made.

#### *Interview/Portfolio Assessment*

An instrument designed and developed for the purpose of directing the interview and portfolio analysis process, called "Interview/Portfolio Analysis: Teachers' Grading Practices in High School English" (Appendix D), was developed in response to research question 3 (What do grades in the high school ELA class actually measure?).

Overall, this interview and portfolio assessment instrument was comprised of 18 questions; the topics and descriptions of the questions along with the research question they addressed are outlined in Table B.4.

Table B.4

*Components of Interview and Portfolio Assessment Survey*

Question	Topic/Description	Research Question(s) Addressed
Question 1	Agreement to participate	None
Question 2	Definitions used in survey	None
Question 3	Grade Report - directions for selecting, a single class period's recent grade report	None
Question 4	Length of grading period	General characteristics
Question 5	Grade level and course represented in the grade report	Demographic information
Question 6	Number of assignments recorded during grading period	Research question 3
Question 7	Number of times final scores adjusted	Research question 3
Question 8	Number of grades representing early learning	Research question 3
Question 9	Number of times new evidence replaced old evidence	Research question 3
Question 10	Number of assignments measuring previously taught content/skills	Research question 3
Question 11	Causes of a zero grade	Research question 3
Question 12	Number of students earning at least one zero grade	Research question 3
Question 13	Number of students earning point deductions for behavior	Research question 3
Question 14	Number of assignments preceded by teacher feedback	Research question 3

*(table continues)*

*Components of Interview and Portfolio Assessment Survey (continued).*

Question 15	Reasons for using grades	Research question 3
Question 16	Control over collected and graded assignments	Research question 3
Questions 17	Control over number of grades	Research question 3
Questions 18	Types of assignments represented in grade book	Research question 3

Questions 1 through 3 represent the launching of the interview as each of these informs the participant of necessary information including the “Consent to Participate” information, the definitions for terms that will be used in the interview, and directions on how to select an appropriate class and grading period for the whole of the portfolio assessment. Questions 4 and 5 were designed to capture information regarding the general design of the interviewees’ classes – i.e., the length of the grading period and the courses represented.

Results from Survey 1 directed the development of the subsequent thirteen interview/portfolio analysis questions to the extent that when there was a statistically significant difference between teachers’ beliefs and their self-reported practices, there was no attempt made to identify the extent to which that belief was detected in actual practice. In other words, practices that respondents did not self-report as used in their grading programs were not included in this grade book analysis because they were unlikely to be found.

*Interview/Portfolio Assessment – Pilot Study*

Content validity of the interview/portfolio assessment was tested through a quasi-pilot study involving the first four interviewees to complete the interview.

At the conclusion of their interviews, these first four participants were informed that they were the first four to complete the interview/portfolio analysis process and, consequently, their responses to four additional questions would assist me in confirming the validity of the content of the interview. They were each asked if they would be willing to answer four additional questions, and all four agreed to do so.

I asked them to consider the interview experience with regards to the content and instructions and then asked them these four questions: (a) Did you understand what I was trying to ask you with each question, (b) When you encountered each question, were you comfortable with the instructions, (c) When you began answering each question, were you able to answer the way in which you wanted to, and (d) What other feedback would you like to provide about the survey or your experience taking it?

Several recommendations were offered, and one minor change was made – the language of “report card grade” was changed to “grade report.” No other changes were recommended or made.

### *Survey 2*

Data needed to consider research question 4 were collected using an instrument developed for this purpose called, “Survey on Causes for Discrepancies between High School English Teacher Beliefs and Practices” (Appendix E). The purpose of this survey was to capture teachers’ perceptions of the reasons for discrepancies found to exist between teachers’ beliefs about best practices, their self-reported practices, and evidence of actual practice. To that end, the development of this survey was directed by results gathered from both Survey 1 and the interview/portfolio analysis instrument.

When results from Survey 1 indicated a statistically significant difference between teachers' beliefs and their self-reported practices, a related question was included in Survey 2 and the directions to that item explained that the goal was to understand why teachers may report a particular belief but then also report that they do not engage in the related practice. Likewise, when results from Survey 1 indicated no statistically significant difference between teachers' beliefs and their self-reported practice but a statistically significant difference was found between reported practices and actual practice as measured by the interview/portfolio analysis, the related question in Survey 2 attempted to explain that the goal was to understand why a discrepancy may exist between beliefs and actual practice. Ultimately, four types of possible discrepancy results could have generated the items in this survey: (a) a reported belief not self-reported in practice, (b) a reported belief found in self-practice but not in actual practice, (c) a reported a non-belief self-reported in practice, or (d) a non-belief not self-reported in practice but found in actual practice.

This survey was comprised of fifteen questions, outlined in Table B.5 and delineated by question number, topic and content description, and the research question addressed.

Table B.5

*Design of Survey 2 by Topic/Description and Research Question Addressed*

Question	Topic/Description	Research Question(s) Addressed
Question 1	Agreement to participate	None
Question 2	Definitions used in survey	None

*(survey continues)*

*Design of Survey 2 by Topic/Description and Research Question Addressed (continued).*

Question 3	Reasons for not using selective adjust of a student's final grade	Research question 4
Question 4	Reasons for limited opportunities for practice and mastery before grading	Research question 4
Question 5	Reasons for including formative measures in final grades	Research question 4
Question 6	Reasons for not replacing old evidence with new evidence	Research question 4
Question 7	Reasons for not grading only what was taught in class	
Question 8	Reasons for including conduct/behavior information in final grades	Research question 4
Question 9	Reasons for not requiring students to complete "zero" work	Research question 4
Question 10	Reasons for using grades to motivate or punish	Research question 4
Question 11	Reasons for not having control over how many grades to record	Research question 4
Question 12	Reasons for not measuring grammar and usage through writing samples	Research question 4
Question 13	Reasons for not including classroom discussion in final grades	Research question 4
Question 14	Reasons for not including a writer's notebook in final grades	Research question 4
Question 15	Reasons for not including student presentations in final grades	Research question 4

---

### *Survey 2 - Pilot Study*

Content validity of Survey 2 was tested through a pilot study involving six members of the population who were not otherwise part of the sample. As in the pilot study for Survey 1, these participants were excluded from the sample because they worked in the same school district where I am employed as a district-level administrator. I have an indirect supervisory relationship with the participants, and so to reduce any bias in their responses, they were assured that the responses they provided would not be made part of the study and would not be analyzed for content. The language of the text used to invite participation in this pilot study is reported in Appendix F.

The pilot study participants were asked to complete the entire survey while responding to each of the open response items with as many reasonable answers as possible. Additionally, they were asked to measure the length of time needed to complete the survey. The average length of time reported was 13.8 minutes, with a range of time from 10 minutes to 18 minutes and a standard deviation of 2.79. Given these results, respondents in the sample were told the survey would likely take between 15 and 20 minutes to complete.

After this administration, these participants were asked to consider the survey with regards to structure, item construction and content, instructions, and definitions. They were asked to respond to these four questions: (a) Did you understand what each question was trying to ask you, (b) When you encountered each question, were you comfortable with the instructions, (c) When you began answering each question, were you able to answer the way in which you wanted to, and (d) What other feedback would you like to provide about the survey or your experience taking it?

Based on the responses from this pilot study, a significant change was made in the structure of the items. Three of the pilot study participants indicated that the wording felt evaluative and caused them to want to defend their own practices rather than fully consider the reasons for the discrepancies noted. Consequently, the structure of the wording was changed so that survey participants were clear that the purpose of the survey was not to challenge their practices but for them to explore possible causes as to why others may not engage in the practices indicated. Initially, the questions were worded using this structure: “Some high school English teachers report that they believe...; however, findings show that they do not always...” After revisions, the items were structured in this manner: “Some high school English teachers report that they believe... / Many teachers employ this practice. This question is trying to better understand why some teachers might not. What do you think might be the reasons that...”

Additionally, pilot survey participants provided a collection of possible answers for each open-ended response. The answers that were replicated at least twice (that is, by more than one pilot study participant) were then included in the survey questions as possible answer choices (along with options for free/open response).

It should be noted here that two versions of this survey were constructed. The first version, which was distributed to participants and received some responses over the course of several days, was soon found to be flawed in content in that three items found to have a discrepancy which should have been included had been accidentally omitted from the survey. In resolve, a corrected survey was constructed, and data collected on the original flawed survey were deleted.

## Participants

### *Population*

This study was conducted with regards to the population of U.S. high school teachers who currently teach at least one English-language arts class required for graduation, in any grade 9-12, and who are duly certified to teach that class according to the requirements established in their state. Excluded from the targeted population, for the purposes of this work, are non-certified teachers or teachers of English elective courses only such as creative writing and journalism. Additionally, because of problems with generalizability, teachers at campuses with atypical populations were not considered to be the population for this study. For the purposes of this study, schools with atypical populations include: schools not deemed “Regular” through United States Census categorization, alternative campuses, vocational schools, special education centers, charter schools, magnet schools, online/virtual schools, correspondence schools, hospital schools, juvenile/detention schools, adult only campuses, and night only campuses.

Using 2010-2011 data from the U.S. Department of Education’s National Center for Education Statistics, it was determined that there are 13,836 campuses that included students in grades 9-12. By definition, this includes campuses with other grade levels as well, such as junior high schools with 8<sup>th</sup>/9<sup>th</sup> grades or secondary schools which house grades 6-12. It also includes 9<sup>th</sup> grade centers, 9<sup>th</sup>/10<sup>th</sup> grade centers, and 11<sup>th</sup>/12<sup>th</sup> grade centers. Excluded from this list and count are schools with the atypical teacher and student populations described previously.

Using this same source for data collection, it was determined that at these 13,836 schools, there are employed 822,755 full-time teachers. The Common Core of Data program of the U.S. Department of Education reports that approximately 15.9% of full-time secondary teachers are employed as high school English language arts teachers. Using this figure, it can be estimated that there are approximately 130,818 high school English teachers working at typical high schools in the United States.

### *Sample and Sampling Procedures*

This study was conducted using non-probability sampling – specifically, a convenience sample was used, with the understanding that a sampling bias likely exists against those in the population who are not represented by the sample that was easily accessible to me personally and against those who are unlikely to volunteer to participate in the study. The parameters of the convenience sample included all those high school English teachers considered part of the population who are within my own ease of access through (a) direct contact initiated because of personal knowledge, (b) indirect contact via association with others personally known to me, or (c) indirect contact via connections on the social networking sites Facebook and Twitter. The language of the text used to invite participation in each of these scenarios is reported in Appendix G.

### *Sample for Survey 1*

Initially, direct contact was made via individually composed emails to approximately 60 former and current colleagues who would be considered part of the population for this study. To avoid a significant bias or issues with sincerity of responses, excluded from this and all contact were possible respondents who were employed in the same school district where I

work as a district-level administrator. This direct contact email provided an explanation of the reason for the contact, stated the purpose of the study, invited the recipients to volunteer to participate, and included the link to the survey. The survey was not monitored following these contacts because there was no intent to identify separately respondents who participated as a result of direct contact versus respondents participating as a result of indirect contact.

In the following weeks, approximately 40 emails addressed to personally known campus administrators, district administrators, and teachers not part of the target population were sent. This correspondence was worded similarly to that of the direct contact email except that rather than asking the recipients to participate, it requested that the recipients share the invitation with others who would be known personally to them who may wish to participate. Because there was no control over who would receive this request for participation, it was essential that the surveys had been designed with the “skip logic” feature that closed the survey if it determined the participant did not meet the requirements to be considered part of the population.

For both these direct and indirect contacts made using email, individually composed emails, as opposed to bulk/group emails, were utilized for two reasons – (1) to lend a sense of personal connection to those contacted in the hopes that the response rate would be higher than otherwise, and (2) to avoid district spam blockers that might be activated with the use of bulk emails.

During this same period of time, indirect requests for participation were made using the social networking sites Facebook and Twitter; one Facebook request was made, and three

Twitter requests were sent. These invitations were especially brief and so merely stated only the request for participation and provided the link to the survey instrument.

The survey remained opened for six weeks. One hundred and ninety-one respondents began the survey, and 136 of those completed it. Thirty-eight states were represented in the total sample, and 18 respondents did not indicate their state of residence. Eighty-five schools were represented in the sample, and 27 respondents not indicating their school of employment. It should be noted that though an attempt was made to directly or indirectly contact potential participants across the country, a large percentage of participants – 37% – represented Texas schools.

### *Demographics*

Demographic information on the participants was collected and is presented and compared to the population demographic information in Table B.6.

There are some notable differences in the age distributions between the sample and the population. Where the percentages of older teachers (50+) in the sample and population are relatively the same, the population has twice as many younger teachers ( $\leq 29$ ) as the sample. This is significant to this study because younger teachers are often more likely to embrace reforms in education (Feistritzer, 2011), so the sample may indicate a lesser approval of expert recommended grading systems if the participants considered them to be reform-like in nature.

Likewise, there are notable differences in the types of communities the sample represents as compared to the population. Participants in Survey 1 represent teachers working at schools located in the suburbs at a rate twice as great as the population, and they represent rural community schools at a markedly lesser degree than the population.

Table B.6

*Demographic Profiles of Participants vs. U.S. English Teachers*

Characteristic	Population %*	Sample %	Sample Difference
<b>Age</b>			
≤29	21	10	-52%
30-39	27	30	+11%
40-49	22	30	+36%
50+	31	30	-03%
<b>Gender</b>			
Male	16	11	-31%
Female	84	89	+05%
<b>Years of Teaching Experience</b>			
1-5	26	15	-42%
6-9	16	16	--
10-14	16	21	+31%
15-24	23	33	+43%
25+	17	15	-12%
<b>Types of School Community</b>			
Rural	42	25	-40%
Town	18	12	-33%
City	16	14	-12%
Suburb	24	49	+104%

\*Feistritzer (2011) and U.S. Department of Education's National Center for Education Statistics (2012).

In addition to the demographic information collected, information regarding the professional characteristics of the sample respondents was also collected and analyzed for the purpose of understanding the sample more thoroughly.

*Familiarity with Professional Organizations*

The degree to which respondents had knowledge of educational organizations was collected in Question 7. The three organizations targeted for this inquiry were NCTE (the National Council of Teachers of English), IRA (the International Reading Association), and ASCD (the Association for the Supervision of Curriculum Design). Levels of degree of recognition of these organizations offered for responses ranged from self-identification as a member of the

organization to indicating no familiarity whatsoever. A summary of the results of these data are found in Table B.7.

Table B.7

*Participant Familiarity with Professional Organizations*

Professional Organization	A member of the organization	Not a member but very familiar with the work of the organization	Not a member, somewhat familiar with the organization	Not familiar with the organization whatsoever
NCTE	61%	24%	15%	0%
IRA	09%	23%	34%	34%
ASCD	11%	25%	25%	39%

Of these three professional organizations, it could be argued that NCTE is the one which is most closely aligned to the everyday work of the high school English educator. The NCTE itself argues that members of their organization “demonstrate their desire to be better teachers by staying abreast of the latest teaching techniques...” (NCTE, n.d.). If it can be assumed that participation in a national organization, particularly NCTE, is evidence of a desire to be committed to trends and developing understandings in the field, then this report showing that at least 61% of respondents as members could somewhat affect the perspective by which one might consider the survey results.

*Knowledge of Grading Experts*

Information regarding the level at which participants had familiarity with educators and researchers who have widely published on the topic of grading was also collected. This list includes all five of the experts whose grading recommendations were used to construct the items in Survey 1 as well as an additional five. A summary of this information is provided in

Table B.8

*Participant Knowledge of Grading Experts*

Grading Expert	I have an extensive understanding of his/her work.	I have a limited understanding of his/her work.	I know the name but am not otherwise familiar with his/her work.	I do not recognize the name.
Alfie Kohn	12.6%	21.0%	13.4%	52.9%
Carol Ann Tomlinson	22.7%	16.0%	20.2%	41.2%
Douglas Reeves	7.6%	7.6%	11.9%	72.9%
Grant Wiggins	22.0%	16.9%	11.0%	50.0%
Jay Mctighe	18.5%	16.0%	13.4%	52.1%
Ken O'Connor	5.1%	11.9%	11.9%	71.2%
Ric Stiggins	10.1%	12.6%	11.8%	65.5%
Robert Marzano	48.3%	25.0%	8.3%	18.3%
Susan Brookhart	3.4%	5.0%	12.6%	79.0%
Thomas Guskey	4.3%	7.7%	9.4%	78.6%

The only expert on this list who garnered a majority above the “I do not recognize the name” level was Robert Marzano. In addition to writing on grading in his *Classroom Assessment and Grading That Work*, Marzano has written extensively in other publications in the educational field including *The Art and Science of Teaching*, *Coaching Classroom Instruction*, *Classroom Instruction That Works*, and *Classroom Management That Works*. Because of his work in so many areas it is difficult to determine if the sample identifies with him because of his work with grading research. Aside from Marzano, the next two highest levels of recognition were Carol Ann Tomlinson, who is widely known for her work with differentiation, and Grant Wiggins, who is well-known for his expertise in curriculum design. The two researchers whose

work is most tightly aligned to grading – Susan Bookhart and Ken O’Connor – had the two lowest percentages of responses in the “I have an extensive understanding of his/her work.”

### *Grades and Courses Taught*

Data regarding the level of courses currently taught and the past histories of courses and grade levels taught were also collected and a summation is provided in Table B.9. A total of 46 “Other” courses were specified; of these courses, approximately 70% of them are electives such as The History of Rock and Roll, Computer Literacy, and Debate, and approximately 30% of them are non-electives such as Gifted and Talented English, ESL English, and AP US History. With 31.9% of respondents reporting that they currently also teach another class other than a credit bearing high school English class, it suggests that the majority of teachers’ classes – i.e., 67.1% -- are teaching only credit bearing and high stakes courses. Additionally, 30% of the time that teachers are teaching an “Other” course, it is also a required, credit-bearing class. This information serves to inform as to the grading loads and the level of significance placed on the grades students receive from these respondents.

### *Sample for Interview/Portfolio Analysis*

Forty-nine respondents to Survey 1 agreed to participate in the interview/portfolio analysis, as indicated by a favorable response to Question 12 in Survey 1. Each volunteer was contacted using the email address they provided. The language of the text used to contact these participants and further explain the interview/portfolio analysis is reported in Appendix H.

Table B.9

*Grade Levels and Courses Taught by Respondents*

Level and Course Taught	Current	Previous
6th grade English language arts	N/A	7.9%
6th grade English language arts - Honors or Pre-AP	N/A	2.6%
7th grade English language arts	N/A	28.9%
7th grade English language arts - Honors or Pre-AP	N/A	8.8%
8th grade English language arts	N/A	28.1%
8th grade English language arts - Honors or Pre-AP	N/A	13.2%
9th grade English	26.7%	62.3%
9th grade English - Honors or Pre-AP	12.1%	22.8%
10th grade English	37.1%	63.2%
10th grade English - Honors or Pre-AP	15.1%	28.9%
11th grade English	31.0%	58.8%
11th grade English - Honors or Pre-AP	10.3%	26.3%
12th grade English	31.9%	50.9%
12th grade English - Honors or Pre-AP	5.2%	9.6%
AP Language and Composition	14.7%	20.2%
AP Literature and Composition	8.6%	8.8%
International Baccalaureate, middle grades	N/A	9.0%
International Baccalaureate, high school	10.3%	4.4%
Other (please specify)	31.9%	21.9%

Of the 49, 42 replied to this email contact, and 41 scheduled times over the next three weeks for the interview/portfolio analysis. Most interviews were scheduled in the late afternoons, based on the most convenient times for the interviewees or on Saturdays and Sundays. A small number occurred during the work day (presumably during the teachers' conference periods). In all, 39 respondents participated in the full interview/portfolio analysis process. Of these 39, all were part of the original sample from Survey 1. The three of the 42 volunteers who did not participate after scheduling the interview did not do so because they were not available at the planned time. A voice mail was left for them indicating that I had called. Two never returned that call, and one called back two weeks after the process had closed and results were already being tabulated and interpreted.

#### *Sample for Survey 2*

The convenience sample for Survey 2 was derived using the same methods as those used for Survey 1, except for the omission of direct contact made to my colleagues who may be considered part of the population. Because those known personally to me had been previously asked to participate in Survey 1, they were not approached again for Survey 2. To that end, indirect contact via association with other educators personally known to me and indirect contact via connections on social networking sites Twitter and NCTE's Teaching and Learning site were made. The language of the text used to invite participation in each of these two scenarios is reported in Appendix I.

Over the course of 11 days, approximately 30 emails addressed to campus administrators, district administrators, and teachers not part of the target population were distributed. It would be fair to say that the "net was cast wider" for this second survey, given

that many of those more closely known to me had already been contacted. Again, individually composed emails, as opposed to bulk/group emails, were utilized. A total of three requests were made on Twitter and two requests were sent through the NCTE's Teaching and Learning site.

An initial version of this survey remained open for several days and several responses were collected before a content error was found. The revised version of the survey was constructed and replaced the original, and the original responses to the first version of the survey were deleted. The corrected version of the survey remained open for weeks. Ninety-nine respondents began the survey, and 68 completed it. One hundred percent of the participants in this survey were members of the intended population; additional demographic information was not collected.

#### Data Analysis

Data collection was initiated in the early fall of 2012 and was completed by January of 2013. All data collection was completed in full compliance with the Institutional Review Board guidelines and received IRB approval in April of 2013.

#### *Research Questions 1 and 2*

For the purpose of data collection for research question 1 (What do high school English teachers believe a grade should measure?) and research question 2 (What do high school teachers report their grades measure?), survey response items were designed and responses collected from the sample via Survey 1. These survey response items were considered both Likert scales and Likert-type items – the primary difference being whether or not the individual items work separately from each other or work together as a collective, summed whole (Clason

and Dormody, 1994). The original intent of the Likert scale was to combine a series of responses from related items into a single composite score in order to create a quantified measure of an attitude or personality trait (Likert, 1932). The evolution of this process has produced Likert-type items which have been designed to work individually to inform researchers of attitudes or opinions on singular, focused items. The items in Survey 1 were designed to do both – to work separately with regards to the particular grading behaviors and assignments graded within the teacher’s grading program and to work collectively to inform me of teachers’ overall beliefs and self-reported practices regarding grading behaviors and graded assignments.

To this end, four sets of items (grouped by question) used in Survey 1 act independently of each other in that each item in the set represents a directed, more focused sub topic of the grading program. Two of those four sets of items work together in such a way that they can be considered true Likert scales.

Table B.10 identifies the Likert-type items (62 in total) and the Likert-type items (2 in total) and shows the relationship between these two types of survey items used in Survey 1.

Table B.10

*Relationship Between Likert Scale Items and Likert-type Items in Survey 1*

Research Question Addressed	Likert-type Items	Likert-Scale Items
Research question 1, Part A	15 Items - Selective adjustment of a student’s final grade - Opportunities for practice and mastery before grading - Use of exemplars for understanding expectations - Avoiding calculating formative assessments into final grades - Replacing old evidence with new evidence in grade book - Grading only on what was taught in class - Separation of behavior grades from academic grades - Emphasis on feedback rather than grades - Weighting end of unit grades more than early evidence - Student self-assessment prior to grading - Student options for demonstrating understandings - Forced completion of all assignments (no zeroes) - Use of grades solely for communication of achievement - Grading decisions (what, how, and weight) in teacher control - Equal weighting of reading and writing in final grades	1 Item - Teachers’ beliefs about grading systems
Research question 1, Part B	16 Items - First drafts of an essay/composition - A fully revised and edited essay/composition - Short answer responses to reading/literature questions - Reading checks/quizzes - Grammar worksheets - Writing samples that measure grammar/usage - Spelling tests - Handwriting samples - Multiple choice tests for reading/literature - Multiple choice tests for grammar/usage - True/false tests - Classroom discussions - Writer’s notebook - Class notes - Annotations made on texts or in books - Presentations (in front of an audience)	None

*(table continues)*

*Relationship Between Likert Scale Items and Likert-type Items in Survey 1 (continued).*

Research question 2, Part A	<p>15 Items</p> <ul style="list-style-type: none"> <li>- Selective adjustment of a student’s final grade</li> <li>- Opportunities for practice and mastery before grading</li> <li>- Use of exemplars for understanding expectations</li> <li>- Avoiding calculating formative assessments into final grades</li> <li>- Replacing old evidence with new evidence in grade book</li> <li>- Grading only on what was taught in class</li> <li>- Separation of behavior grades from academic grades</li> <li>- Emphasis on feedback rather than grades</li> <li>- Weighting end of unit grades more than early evidence</li> <li>- Student self-assessment prior to grading</li> <li>- Student options for demonstrating understandings</li> <li>- Forced completion of all assignments (no zeroes)</li> <li>- Use of grades solely for communication of achievement</li> <li>- Grading decisions (what, how, and weight) in teacher control</li> <li>- Equal weighting of reading and writing in final grades</li> </ul>	<p>1 Item</p> <ul style="list-style-type: none"> <li>- Teachers’ self-reported grading systems</li> </ul>
Research question 2, Part B	<p>16 Items</p> <ul style="list-style-type: none"> <li>- First drafts of an essay/composition</li> <li>- A fully revised and edited essay/composition</li> <li>- Short answer responses to reading/literature questions</li> <li>- Reading checks/quizzes</li> <li>- Grammar worksheets</li> <li>- Writing samples that measure grammar/usage</li> <li>- Spelling tests</li> <li>- Handwriting samples</li> <li>- Multiple choice tests for reading/literature</li> <li>- Multiple choice tests for grammar/usage</li> <li>- True/false tests</li> <li>- Classroom discussions</li> <li>- Writer’s notebook</li> <li>- Class notes</li> <li>- Annotations made on texts or in books</li> <li>- Presentations (in front of an audience)</li> </ul>	<p>None</p>
Totals:	62 Likert-type Items	2 Likert Items

The analysis processes appropriate for Likert scales and Likert-type items differ from each other. The 62 Likert-type items are not treated as a collective whole and are consequently analyzed using a methodology appropriate to an ordinal scale (Boone & Boone, 2012; Clason & Dormody, 1994; Jamison, 2004; Kuzon, Urbanchek, & McCabe, 1996.). For analysis of the two Likert scale item sets, data were summed and therefore analyzed using a methodology

appropriate to an interval scale (Boone & Boone, 2012; Likert, 1932). A summary expression of the data analysis strategies for Survey 1 data is reported in Table B.11.

Table B.11

*Analysis Design for Survey 1: Research Questions 1 and 2*

Research Question	Item type	Scale type	Analysis strategies
RQ1: What do high school English-language arts teachers believe a grade should measure?			
Part A: Grading Systems ( <i>How</i> do you grade, record, and report)			
	1 - Likert scale item	Interval Scale	M, SD, SEM
	16 - Likert-type Items	Ordinal Scale	Frequency, Mode, Median, Range
Part B: Graded Assignments ( <i>What</i> do you grade, record, and report)			
	16 - Likert-type Items	Ordinal Scale	Frequency, Mode, Median, Range
RQ2: What do high school English-language arts teachers self-report their students' grades measure?			
Part A: Grading Systems ( <i>How</i> do you grade, record, and report)			
	1 - Likert scale item	Interval Scale	M, SD, SEM
	16 - Likert-type Items	Ordinal Scale	Frequency, Mode, Median, Range
Part B: Graded Assignments ( <i>What</i> do you grade, record, and report)			
	16 - Likert-type Items	Ordinal Scale	Frequency, Mode, Median, Range

Before moving forward to research question 3, a consideration of the results of these research question 1 and research question 2 analyses had to be completed to determine if there were discrepancies between teachers' beliefs about grading practices and their self-reports, as these findings would inform the upcoming research. If there were found to be

expert grading recommendations that teachers did not believe to be best practices, those would be removed from the remainder of the study since there was no intent to determine the extent to which teachers engage practices they report to not find valuable.

Then a comparison of the results from the remaining best practice belief responses and their corresponding self-reported practices responses was needed. To complete this process, the data from the Survey 1 questions related to beliefs were compared to the data from the Survey 1 questions related to self-reported practices. This analysis occurred in two ways – first in terms of the corresponding pairs of Likert-type items which were analyzed using studies of the mode, median, mean rank, and the Mann-Whitney U Test to study a comparison of the medians (as is appropriate for ordinal scale data). The Mann-Whitney U Test was used to determine whether a statistically significant difference existed between the recorded results for teachers' beliefs about best practice (in both grading systems and recorded assignments graded) and their self-reported practices (in both grading systems and recorded assignments graded). The measure of mean rank was utilized to report which grouping variable – the best practice responses or own practice responses – indicated a stronger response toward agreement with the expert recommendations. These data were analyzed using SPSS (Analyze > Nonparametric Tests > Legacy Dialogues > 2 Independent Samples).

And then, secondly, the comparison of the Likert scale item results for best practice beliefs for grading systems and self-report practices for grading systems was completed using studies of the means, standard deviations, standard errors of the mean, and an independent samples T-test (as is appropriate for interval scale data). These data were analyzed using SPSS (Analyze > Compare Means > Independent Samples T-Test).

It is worth noting here that the choice to use of the Mann-Whitney U Test was a difficult one. Because data were not collected as paired samples – that is, respondents’ answers to questions regarding their beliefs were not directly paired with their answers regarding their self-reported practices, the one test that would have met all of the needs of this study, including the use of ordinal data – The Wilcoxon Sub Test – could not be used. All assumptions for the Mann-Whitney U Test were met except that the two collections of data do not represent samples from two populations completely independent of each other. That being said, however, this assumption was explored further in research question 5 as I attempted to determine what influence, if any, the order of the questions on the survey had on responses.

### *Research Question 3*

For the purpose of data analysis for research question 3 (What do high school English teachers’ grades actually measure?) the responses to the structured items from the Interview/Portfolio Analysis work were treated in a variety of ways based on their intent and design.

Questions 4 and 5, meant to describe the characteristics of the sample – such as the length of grading period used in the respondents’ schools and their grade levels and courses taught – were analyzed descriptively using frequencies reported as percentages with a consideration of mode as the measure of central tendency.

Open-response items that reflected the grading programs of the respondents and were reported on an interval scale – such as the number of assignments recorded during a given period of time – were analyzed both descriptively using mean and standard deviation, and inferentially using standard error of the mean and the calculation of 95% confidence intervals.

Additionally, because these same items included opportunities for additional respondent comments, they were studied qualitatively through a process that sought to identify and categorize trends in language and over-arching themes in responses. Questions 6-10 and 12-16 from the interview/portfolio analysis represent these types of items that were analyzed both quantitatively and qualitatively.

And lastly, the responses for Question 11, which was an open-response item that elicited qualitative data, was analyzed for patterns, organized into meaningful categories by theme, and then summed for the purpose of reporting. Each response was forced into a single category, with an “Other” category included for the management of responses that did not appear to be part of a pattern. The resulting data were reported out using frequencies reported as percentages with a consideration of mode as the measure of central tendency.

#### *Research Question 4*

Central to the work of analyzing data for research question 4 was the goal of identifying themes and patterns that emerged from responses to Survey 2. The research question concerned the perceived causes for discrepancies between what high school ELA teachers believe grades should measure and what grades in the high school ELA classroom actually measure. By definition, this research question is primarily interested in inviting participants to identify their own perceptions as to the causes for the identified discrepancies; to some extent, it is the teacher’s own perceptions about causes and controls that defines the concept of “politics” in this study. There were no limitations on how many causes a respondent might select for each discrepancy, and they were provided the opportunity to qualify or add their own or additional responses.

Responses regarding the causes for the discrepancies were first analyzed qualitatively through a process that sought to identify and characterize trends in respondents' comments. Once patterns emerged and a series of particular causes were identified, the frequency with which each of these was mentioned in the responses was calculated. This quantification allowed for some reporting using descriptive statistics including frequencies calculated as percentages and mode as the measure of central tendency. This mixed method approach allowed me to base the conclusions on perceived trends in the responses while also presenting measurable evidence as to the degrees to which the responses supported the conclusions. Because no measure of statistical significance could be made given the ordinal and qualitative nature of the data, analysis included the consideration of any identified cause reported by at least 1% of the sample. Additionally, the identified causes with the greatest frequencies for each of the discrepancies and the overall causes as a whole were identified as such.

#### *Research Question 5*

In response to the final research question, a quantitative strategy was used with a quasi-experimental, counterbalanced design to determine if there were limitations in the data collected in Survey 1 created by the order of the questions. Specifically, the aim was to determine if participants who responded to questions about their beliefs about best grading practices were affected by the knowledge that this survey was also asking questions regarding their own actual grading practices.

Consequently, data were collected from two independent samples – the first sample (Treatment Group A) was comprised of seventy-nine members of the population, and the second sample (Treatment Group B) was comprised of fifty members of the population. In this

research design, all participants received the same “treatment” – i.e. completion of Survey 1 – but in a different order. The instrumentation was manipulated in such that Treatment Group A and Treatment Group B varied with regards to which part of the survey instrument was received and answered first – either questions regarding research question 1 (beliefs) or questions regarding research question 2 (self-reported) practices. The survey was designed with a logic feature that prevented respondents from returning to earlier questions to alter responses. Distribution of the two survey versions was random through a process of alternating the survey addresses (URLs) provided to potential participants.

A Mann-Whitney U Test was used to determine if there was a statistically significant difference of the means (as is appropriate for an ordinal scale) on a dependent variable (measure of best practice) between two groups of an independent variable (version 1.2 and version 2.1). The measure of mean rank was used to report which grouping variable – version 1.2 or version 2.1 – indicated a stronger response toward agreement with experts on best practices. This data were analyzed using SPSS (Analyze > Nonparametric Tests > Legacy Dialogues > 2 Independent Samples).

APPENDIX C  
COMPLETE/UNABRIDGED RESULTS

Results are presented in order of research question with the segmented research questions 1-5 examined first followed by a consideration of how these results work together to inform the primary research question – What is the status among high school English teachers’ beliefs about grading practices, their self-reported grading practices, and their actual grading practices?

### Research Question 1

Research question 1 invited an exploration of what high school English-language arts teachers believe a grade should measure. This question was considered in two ways – Part A represented beliefs regarding the whole of the teachers’ grading systems (i.e. *how* teachers should grade, record, and report), and Part B represented beliefs regarding graded assignments (i.e. *what* teachers should grade, record, and report).

The strategy of inquiry employed for this exploration of teacher beliefs was quantitative by way of a sample, cross-sectional survey in the form of an on-line questionnaire. Two forms of the same survey were designed and used for the purpose of considering research question 5; these forms were named *1.2* and *2.1* and differed in their sequencing of questions. However, for the purpose of research question 1, responses from identical questions on the two forms were combined for reporting. In total, 136 members of the population of high school English teachers served as the sample for research question 1.

#### *Research Question 1, Part A – Survey 1*

The following instructions were provided to respondents on Question 3 on version 1.2 and its corresponding 5 on version 2.1 of Survey 1, which addressed only Part A of research question 1: *Please answer this question with regards to your personal beliefs about best*

*grading practices that should be implemented by high school English teachers in general.* These instructions were followed by this question: *To what extent do you agree with each of the following statements?*

Results presented in Table C.12 represent a treatment of these items as Likert-type in that they are functioning here as independent of each other. Consequently, they were analyzed using a methodology appropriate to an ordinal scale (Boone & Boone, 2012; Clason & Dormody, 1994; Jamison, 2004; Kuzon, Urbanchek, & McCabe, 1996). These item results are reported descriptively, with frequency reported in percentages and mode to report the central tendency. For conciseness in reporting, the best practice topics listed in Table C.12 are abbreviations of the complete best practice statements used on the survey; the complete wording used on the survey statements can be found in Appendices A and B.

In 10 of the 15 items for this question, the response with the highest frequency was found in the category of more extreme agreement – *Completely agree*. In the remaining five of the 15 items, the mode was found in the category representing the next level of agreement – *Mostly agree, but with some reservations*. Additionally, in 13 of these 15 items, greater than a two-thirds majority of participants responded with one of these two levels of agreement. The two items with the lowest frequencies of agreement were *Selective adjustment of a student's final grade* and *Avoiding calculating formative assessments into final grades*. The former of these also represents the item with the highest frequency in the category of more extreme disagreement – *Completely disagree* – at 9.4%.

Table C.12

*Survey 1, Part A Results – Best Practice Beliefs in Grading Systems, Likert-type Items*

Best Practice / Topic	Agreement		Disagreement	
	Completely agree (3)	Mostly agree, but with some reservations (2)	Mostly disagree, but with some reservations (1)	Completely disagree (0)
Selective adjustment of a student's final grade	21.1%	46.9%	22.7%	9.4%
Opportunities for practice and mastery before grading	72.8%	23.3%	3.9%	0.0%
Use of exemplars for understanding expectations	57.4%	37.2%	5.4%	0.0%
Avoiding calculating formative assessments into final grades	13.2%	43.4%	38.8%	4.6%
Replacing old evidence with new evidence in grade book	30.4%	37.5%	22.7%	9.4%
Grading only on what was taught in class	36.4%	42.6%	14.7%	6.2%
Separation between behavior grades and academic grades*	8.6%	14.8%	35.2%	41.4%
Emphasis on feedback rather than grades	79.1%	19.4%	1.5%	0.0%
Weighting end of unit grades more than early evidence	43.0%	42.2%	12.5%	2.3%
Student self-assessment prior to grading	53.5%	43.4%	3.1%	0.0%
Student options for demonstrating understandings	48.4%	48.4%	2.3%	0.8%
Forced completion of all assignments (no zeroes)	39.5%	42.6%	12.8%	5.4%
Use of grades solely for communication of achievement	44.9%	48.9%	6.2%	0.0%
Grading decisions (what, how, and weight) in teacher control	58.9%	37.2%	3.9%	0.0%
Equal weighting of reading and writing in final grades	59.7%	35.7%	3.1%	1.5%

\*Note that in the actual survey documents (Appendices A and B), the item regarding the separation of academic and behavior grades was inadvertently asked in the inverse of best practice order. For the purposes of calculating and reporting, it has been corrected here and the response frequencies have been reversed, respective to the levels of agreement or disagreement. While this type of correction might otherwise create a validity issue with the

survey item, in this instance, the responses represent such a noteworthy inclination toward a particular opinion, that I defend the revision and believe it to be accurately reporting respondents' beliefs.

Though there are limitations in the ability to interpret ordinal data expressed only in descriptive statistics, I feel confident in concluding that the respondents to this survey agree that each of the 16 grading system practices listed here are indeed best practices.

In addition to these results determined by the Likert-type treatment of these items, I analyzed the data again but this time with regards to how these 15 items worked together as a single Likert scale item. Once all responses provided for the 15 items related to beliefs about best grading practices were summed, they were analyzed using a methodology appropriate to authentic Likert items (Boone & Boone, 2012; Likert, 1932), including both descriptive statistics (measures of the mean, median, and mode, and standard deviations) and inferential statistics (the standard error of the mean and confidence intervals). These results are presented in Table C.13.

Table C.13

*Survey 1, Part A Results – Best Practice Beliefs in Grading Systems, Likert Scale (Collective)*

<i>M</i>	<i>SEM</i>	Median	Mode	<i>SD</i>	95% CI [LL, UL]
2.28	.018	2.00	3.00	.798	[2.24, 2.32]

The treatment of these individual Likert-type items as a collective Likert scale allows for an additional level of certainty that the results from the survey are representative of the population. For example, when 95% confidence intervals are drawn around the mean of 2.28, we are able to say, with 95% certainty, that our population mean is between 2.24 and 2.32.

This, combined with the evidence of the median (2.00) and the mode (3.00) allows us to say with more certainty that the population as a whole would be in some level of agreement (*Completely or Mostly*) that these expert recommendations for grading practices are, indeed, best practice.

#### *Research Question 1, Part B – Survey 1*

Respondents were provided this question in Questions 4 on version 1.2 and corresponding Question 6 on version 2.1 of Survey 1, which addressed only Part B of research question 1: *Please answer this question with regards to your personal beliefs about best grading practices that should be implemented by high school English teachers in general.* These instructions were followed by this question: *To what extent is each of these assignments important when considering what should count toward a student's final grade in English?*

Results presented in Table C.14 represent a treatment of these items as Likert-type in that they are functioning here as independent of each other. Consequently, they were analyzed using an appropriate methodology for Likert-type items, and item results are reported descriptively, with frequency in percentages and mode to report the central tendency.

In two of the 16 items in this question the mode fall in the category of the more extreme level of importance – *Very important; this work should always count toward a student's final grade.* The assignment type garnering the highest percentage response for the greater degree of importance was *A fully revised and edited essay/composition* at 91.9%; the other assignment type whose mode was found in the same category was *Presentations (in front of an audience)* at 58.1%.

Table C.14

*Survey 1, Part B Results – Best Practice Beliefs in Assignments Graded, Likert-type Items*

Best Practice / Topic	Agreement		Disagreement	
	Very important; this work should always count toward a student's final grade (3)	Mostly important; this work should usually count toward a student's final grade (2)	Mostly unimportant; this work should not usually count toward a student's final grade (1)	Completely unimportant; this work should never count toward a student's final grade (0)
First drafts of an essay/composition	21.8%	33.9%	37.9%	6.4%
A fully revised and edited essay/composition	91.9%	8.1%	0.0%	0.0%
Short answer responses to reading/literature questions	36.3%	55.6%	7.3%	0.8%
Reading checks/quizzes	25.2%	48.8%	21.9%	4.1%
Grammar worksheets	7.3%	23.4%	45.9%	23.4%
Writing samples that measure grammar/usage	28.7%	47.5%	17.2%	6.6%
Spelling tests	10.4%	22.6%	30.0%	37.1%
Handwriting samples	1.6%	1.6%	22.6%	74.2%
Multiple choice tests for reading/literature	25.8%	45.9%	23.4%	4.8%
Multiple choice tests for grammar/usage	20.2%	41.9%	28.2%	9.7%
True/False tests	10.5%	25.0%	34.7%	29.8%
Classroom discussions	31.5%	50.8%	15.3%	2.4%
Writer's notebooks	34.1%	48.8%	13.8%	3.3
Class notes	8.9%	26.8%	44.7%	19.5%
Annotations made on texts or in books	18.5%	42.7%	29.8%	8.9%
Presentations (in front of an audience)	58.1%	39.5%	2.4%	0.0%

Conversely, for two items the mode was found in the category of the more extreme level of unimportance – *Completely unimportant; this work should never count toward a student's final grade*. Those two items were *Handwriting samples* at 74.2% and *Spelling tests* at 37.1%.

Two-thirds or more of the respondents indicated they found eight of the 16 items from this question to be important to some degree (*Completely or Mostly*). Of the remaining eight items, five items were identified by a majority of respondents ( $\geq 63\%$ ) to be unimportant to some degree. The data representing the remaining three types of graded assignments were challenging to analyze because despite having a single mode, they did not suggest an overwhelmingly obvious majority response. Where a measure of levels of statistical significance or the assigning of confidence levels would have made this interpretation easier, such strategies would have relied on inferential measures which would require an assumption that the numerical values assigned to each response truly represented equal degrees of differences between the response options. Given the inability to consider this further using a quantitative strategy, I elected to conclude that the variance of responses in the three remaining items makes it problematic for me to confidently conclude that the sample revealed a notable belief about these three items as being best practice. At this point in the study, I rejected the plan to continue the analysis of the importance of grading these three assignment types: first drafts of an essay/composition, multiple choice tests for grammar/usage, and annotations made on texts or in books.

For these Part B items related to graded assignments, no attempt was made to treat them collectively as a single Likert scale since they were not designed to work together in such

a fashion. Whereas the Part A items were all identified through expert texts as being recommended practices and therefore it was relevant to consider them wholly in terms of the degree to which respondents considered them all best practices, these graded assignments were not based on expert recommendations and would not, therefore, be treated collectively as representing best practice.

#### *Research Question 1 Results - Summary*

In summation, high school English-language arts teachers believe that a grading system should include:

- Selective adjustment of a student's final grade
- Opportunities for practice and mastery before grading
- Use of exemplars for understanding expectations
- Avoiding calculating formative assessments into final grades
- Replacing old evidence with new evidence in grade book
- Grading only on what was taught in class
- Separation of behavior grades from academic grades
- Emphasis on feedback rather than grades
- Weighting end of unit grades more than early evidence
- Student self-assessment prior to grading
- Student options for demonstrating understandings
- Forced completion of all assignments (no zeroes)
- Use of grades solely for communication of achievement
- Grading decisions (what, how, and weight) in teacher control

- Equal weighting of reading and writing in final grades

They believe that it is important to grade and record these types of assignments:

- A fully revised and edited essay/composition
- Short answer responses to reading/literature questions
- Reading checks/quizzes
- Writing samples that measure grammar/usage
- Multiple choice tests for reading/literature
- Classroom discussions
- Writer's notebook
- Presentations (in front of an audience)

They do not believe that it is important to grade and record these types of assignments:

- Class notes
- True/false tests
- Handwriting samples
- Spelling tests
- Grammar worksheets

The following items are being removed from the study heretofore because of a variance in responses that makes it difficult to know with certainty whether or not English teachers believe it is important or unimportant to grade and record them:

- First drafts of an essay/composition
- Multiple choice tests for grammar/usage
- Annotations made on texts or in books

## Research Question 2

Research question 2 asked what high school English-language arts teachers self-report their grades to measure. Like research question 1, this question was considered in two ways – Part A represented beliefs regarding the whole of the teachers’ grading systems (i.e. *how* teachers should grade, record, and report), and Part B represented beliefs regarding graded assignments (i.e. *what* teachers should grade, record, and report).

The strategy of inquiry employed for this exploration of teachers’ self-reports on grading practices was quantitative by way of a sample, cross-sectional survey in the form of an on-line questionnaire. Survey 1, the same survey used to collect data for research question 1, was used for this purpose. The response differences in relation to the distinct forms of the survey (i.e. 1.2 and 2.1) were not relevant, and so responses from identical questions on the two forms were combined for reporting here. In total, 136 of the population of high school English teachers served as the sample for this survey.

### *Research Question 2, Part A – Survey 1*

Respondents were provided these instruction for Question 5 on version 1.2 and its corresponding Question 3 on version 2.1 of Survey 1, which addressed only Part A of research question 2: *Think of the English class(es) you currently teach, and please answer this question with regards only to your own current and typical grading practices.* These instructions were followed by this question: *To what extent do you agree with each of the following statements?*

Results presented in Table C.15 represent a treatment of these items as Likert-type in that they are functioning here as independent of each other. Consequently, they were analyzed using a methodology appropriate to an ordinal scale. These item results are reported

descriptively, with frequency in percentages and mode to report the central tendency. For conciseness in reporting, the best practice topics listed in Table C.15 are abbreviations of the complete best practice statements used on the survey. The actual wording of the survey statements can be found in Appendices A and B.

In nine of the 15 items for this question, the highest frequency response was found in the category of more extreme agreement – *Completely agree*. Additionally, five other modes fell into a category of agreement – *Mostly agree, but with some reservations*. Combined, 14 of the 15 responses represented some level of agreement. The mode for the one remaining item – *Avoiding calculating formative assessments into final grades* – was in the disagreement category *Mostly disagree, but with some reservations*. This same topic had the highest frequency of response in the more extreme disagreement category option – *Completely disagree* – with 11.1%.

Though there are limitations in the ability to interpret ordinal data expressed only in descriptive statistics, I feel confident in concluding that the respondents to this survey self-reported at some level that they engaged in 14 of the 15 grading system practices listed.

Again, in addition to these results determined by the Likert-type treatment of these items, I analyzed the data with regards to how these 15 items worked together as a single Likert scale. Once all responses provided for the 15 items related to the self-reporting of grading practices were summed, they were analyzed using a methodology appropriate to an interval scale, including both descriptive statistics (measures of the mean, median, mode, and standard deviations) and inferential statistics (the standard error of the mean and confidence intervals). These results are reported in Table C.16.

Table C.15

*Survey 1, Part A Results – Self-Reported Practices in Grading Systems, Likert-type Items*

Best Practice / Topic	Agreement		Disagreement	
	Completely agree (3)	Mostly agree, but with some reservations (2)	Mostly disagree, but with some reservations (1)	Completely disagree (0)
Selective adjustment of a student's final grade	27.2%	39.0%	22.8%	11.0%
Opportunities for practice and mastery before grading	64.7%	31.6%	3.7%	0.0%
Use of exemplars for understanding expectations	46.3%	40.4%	11.8%	1.5%
Avoiding calculating formative assessments into final grades	11.8%	35.3%	43.4%	9.6%
Replacing old evidence with new evidence in grade book	25.9%	42.2%	20.7%	11.1%
Grading only on what was taught in class	39.7%	42.6%	10.3%	7.4%
Separation between behavior grades and academic grades*	48.1%	28.1%	17.8%	5.9%
Emphasis on feedback rather than grades	77.2%	20.6%	1.5%	0.7%
Weighting end of unit grades more than early evidence	49.3%	37.5%	9.6%	3.7%
Student self-assessment prior to grading	46.3%	43.4%	9.6%	0.7%
Student options for demonstrating understandings	54.4%	36.0%	8.1%	1.5%
Forced completion of all assignments (no zeroes)	26.5%	41.9%	22.8%	8.8%
Use of grades solely for communication of achievement	51.9%	45.9%	2.2%	0.0%
Grading decisions (what, how, and weight) in teacher control	50.4%	43.7%	4.4%	1.5%
Equal weighting of reading and writing in final grades	44.4%	46.7%	7.4%	1.5%

Table C.16

*Survey 1, Part A Results – Self-Reported Practices in Grading Systems, Likert Scale (Collective)*

<i>M</i>	<i>SEM</i>	Median	Mode	<i>SD</i>	95% CI [LL, UL]
2.22	.019	2.00	3.00	.838	[2.19, 2.25]

The treatment of these individual Likert-type items as a collective Likert scale allows for an additional level of certainty that the results from the survey are representative of the population. For example, when 95% confidence intervals are drawn around the mean of 2.22, we are able to say, with 95% certainty, that our population mean is between 2.19 and 2.25. This, combined with the evidence of the median (2.00) and the mode (3.00) allow us to say with more certainty that the population as a whole would report that they practice, to some degree, at least 14 of these 15 expert grading recommended practices.

*Research Question 2, Part B – Survey 1*

Respondents were provided these instruction for Questions 6 on version 1.2 and corresponding Question 4 on version 2.1 of Survey 1, which addressed only Part B of research question 2: *Think of the English class(es) you currently teach, and please answer this question with regards only to your own current and typical grading practices.* These instructions were followed by this question: *To what extent do you find each of these assignments important when considering what should count toward your students’ final grades in English?*

Results presented in Table C.17 represent a treatment of these items as Likert-type in that they are functioning here as independent of each other. Consequently, they were

analyzed using an appropriate methodology including descriptive statistics, with frequency in percentages and mode to report the central tendency.

In two of the 16 items in this question did the mode fall in the category of the more extreme level of importance – *Very important; I always count this work toward a student’s final grade*. The assignment type reported in this category with the highest percentage of responses was *A fully revised and edited essay/composition* at 93.4%; the other assignment type whose mode fell in the same category was *Presentations (in front of an audience)* at 54.9%. Eight additional items were indicated to be *Mostly important* by measure of mode as central tendency and/or frequency counts at or above 67%.

For three other items the mode was in the category of the more extreme level of unimportance – *Completely unimportant; I never count this work toward my students’ final grades*. Those three items were *Spelling tests* at 62.8%, *Handwriting samples* at 62.8% and *True/False tests* at 41.3%. The modes for the remaining three types of assignments were found in the category *Mostly unimportant*, though all three (as in the corresponding Part A study) had results with such a variance that it seemed difficult for me to reach a reasonable conclusion as to whether not teachers in general felt that it was important to include these types of assignments in their students’ final grades.

As was found with research question 1, for these Part B items related to graded assignments, no attempt was made to treat them collectively as a single Likert scale item since they were not designed to work together in such a way.

Table C.17

*Survey 1, Part B Results – Self-Reported Practices in Assignments Graded, Likert-type Items*

Best Practice / Topic	Agreement		Disagreement	
	Very important; I always count this work toward a student's final grade (3)	Mostly important; I usually count this work toward a student's final grade (2)	Mostly unimportant; I usually do not count this work toward a student's final grade (1)	Completely unimportant; I never count this work toward a student's final grade (0)
First drafts of an essay/composition	22.1%	34.4%	35.2%	8.2%
A fully revised and edited essay/composition	93.4%	5.7%	0.8%	0.0%
Short answer responses to reading/literature questions	40.2%	50.8%	8.2%	0.8%
Reading checks/quizzes	28.7%	49.3%	18.9%	3.0%
Grammar worksheets	7.4%	24.0%	41.3%	27.1%
Writing samples that measure grammar/usage	25.4%	40.2%	28.7%	5.7%
Spelling tests	8.3%	10.7%	18.2%	62.8%
Handwriting samples	0.8%	3.3%	13.1%	82.8%
Multiple choice tests for reading/literature	30.6%	40.5%	23.1%	5.8%
Multiple choice tests for grammar/usage	22.1%	36.1%	27.0%	14.8%
True/False tests	9.9%	19.8%	28.9%	41.3%
Classroom discussions	27.9%	41.8%	26.2%	4.1%
Writer's notebooks	30.3%	43.4%	14.8%	11.5%
Class notes	8.4%	23.5%	36.1%	31.9%
Annotations made on texts or in books	14.9%	43.0%	27.3%	14.9%
Presentations (in front of an audience)	54.9%	37.7%	6.6%	0.0%

### *Research Question 2 Results - Summary*

In summation, high school English-language arts teachers report that their own grading systems include:

- Selective adjustment of a student's final grade
- Opportunities for practice and mastery before grading
- Use of exemplars for understanding expectations
- Replacing old evidence with new evidence in grade book
- Grading only on what was taught in class
- Separation of behavior grades from academic grades
- Emphasis on feedback rather than grades
- Weighting end of unit grades more than early evidence
- Student self-assessment prior to grading
- Student options for demonstrating understandings
- Forced completion of all assignments (no zeroes)
- Use of grades solely for communication of achievement
- Grading decisions (what, how, and weight) in teacher control
- Equal weighting of reading and writing in final grades

They report that their own grading systems do not include:

- Avoiding calculating formative assessments into final grades

They report that they included these types of assignments in their students' final grades:

- A fully revised and edited essay/composition
- Short answer responses to reading/literature questions

- Reading checks/quizzes
- Writing samples that measure grammar/usage
- Multiple choice tests for reading/literature
- Classroom discussions
- Writer's notebook
- Presentations (in front of an audience)

They report that they do not include these types of assignments in their students' final grades:

- Grammar worksheets
- Spelling tests
- Handwriting samples
- Class notes
- True/false tests

The responses to the following items were so varied that it was difficult to know with certainty whether or not English teachers find it important to include these types of assignment in their students' final grades and so were heretofore omitted from the study:

- First drafts of an essay/composition
- Multiple choice tests for grammar/usage
- Annotations made on texts or in books

#### *Comparisons of Results – Research Questions 1 and 2*

An analysis of the relationship between results in research question 1 and research question 2 was conducted to determine if there were discrepancies between teachers' beliefs about grading practices and their self-reports about their own practices. This analysis was

essential at this point in time because the design of the Interview/Portfolio Analysis and the subsequent Survey 2 would be affected by the outcomes.

A comparison of the results from the best practice belief responses and their corresponding self-reported practice responses was completed in two ways – (1) by completing a comparison study between the best practice Likert-type items and their corresponding reported practice Likert-type items, and (2) by completing a comparison study between the whole of the best practice responses as a Likert scale with its corresponding whole of the self-reported practice responses also as a Likert scale.

In response to research question 1, teachers reported that they believed that it is best practice to avoid calculating formative assessments into final grades; however, in response to research question 2, they reported that they do not avoid calculating formative assessment into final grades. Because of the results showing a discrepancy between this belief and this practice, this item was not further examined in the comparison analysis and was not analyzed in the interview/portfolio assessment for actual practice. It was, however, part of Survey 2 which asked teachers to consider causes for discrepancies.

Additionally, the following graded assignment types were not analyzed further because it was already determined that the variations in responses made a confident interpretation of teachers' perceived levels of importance difficult: *First drafts of an essay/composition*, *Multiple choice tests for grammar/usage*, and *Annotations made on texts or in books*.

#### *Likert-type Item Comparisons*

A comparison of the remaining Likert-type items was completed using studies of the mode, median, mean rank, and the Mann-Whitney U Test to study a comparison of the

medians (as is appropriate for ordinal scale data). The Mann-Whitney U Test was used to determine if a statistically significant difference existed between the recorded results for teachers' beliefs about best practice (in both grading systems and recorded assignments graded) and their self-reported practices (in both grading systems and recorded assignments graded). The mean was employed to report which grouping variable – the best practice responses or own practice responses – indicated a stronger response toward agreement. Data were analyzed using SPSS (Analyze > Nonparametric Tests > Legacy Dialogues > 2 Independent Samples). Results for grading systems are reported in Table C.18.

Table C.18

*Comparisons of Research Questions 1 and 2 Results – Likert-type Items – Best Practice*

Item	Median	Mode	M Rank	Sig.	Difference
Selective adjustment of a student's final grade					
Best Practice	2.00	2.00	130.88	.724	no
Own Practice	2.00	2.00	134.02		
Opportunities for practice and mastery before grading					
Best Practice	3.00	3.00	137.27	.276	no
Own Practice	3.00	3.00	128.95		
Use of exemplars for understanding expectations					
Best Practice	3.00	3.00	141.69	.045	yes
Own Practice	2.00	3.00	124.76		
Replacing old evidence with new evidence in grade book					
Best Practice	2.00	2.00	134.39	.602	no
Own Practice	2.00	2.00	129.74		
Grading only on what was taught in class					
Best Practice	2.00	2.00	130.12	.522	no
Own Practice	2.00	2.00	135.73		
Separation of behavior grades from academic grades					
Best Practice	2.00	3.00	135.82	.397	no
Own Practice	2.00	3.00	128.38		

*(table continues)*

*Comparisons of Research Questions 1 and 2 Results...Best Practice (continued).*

Emphasis on feedback rather than grades						
Best Practice	3.00	3.00	134.35	.698	no	
Own Practice	3.00	3.00	131.72			
Weighting end of unit grades more than early evidence						
Best Practice	2.00	3.00	138.39	.354	no	
Own Practice	2.00	3.00	136.37			
Student self-assessment prior to grading						
Best Practice	3.00	3.00	141.27	.057	no	
Own Practice	2.00	3.00	125.27			
Student options for demonstrating understandings						
Best Practice	2.00	2.00	130.80	.693	no	
Own Practice	3.00	3.00	134.10			
Forced completion of all assignments (no zeroes)						
Best Practice	2.00	2.00	144.99	.006	yes	
Own Practice	2.00	2.00	120.57			
Use of grades solely for communication of achievement						
Best Practice	2.00	2.00	129.03	.418	no	
Own Practice	3.00	3.00	135.81			
Grading decisions (what, how, and weight) in teacher control						
Best Practice	3.00	3.00	138.65	.145	no	
Own Practice	3.00	3.00	126.62			
Equal weighting of reading and writing in final grades						
Best Practice	3.00	3.00	143.42	.011	yes	
Own Practice	2.00	2.00	122.07			

The null hypothesis for this analysis was that the median difference between the pairs was zero, or  $H_0: Mdn_1 = Mdn_2$ . In cases where the p-value was smaller than alpha (p-value < .05), the null hypothesis was rejected, and I concluded that the median scores of the best practice responses were not equal to the median scores of the self-report practice responses.

Three items in the grading systems section of the survey showed a statistically significantly different measure of the median between the best practice responses and the self-reported responses. These three items were: *Use of exemplars for understanding*

*expectations, Forced completion of all assignments (no zeroes), and Equal weighting of reading and writing in final grades.*

The summed frequencies for these three items for the levels of agreement (i.e. *Completely agree* and *Mostly agree*) in the best practice study and the self-reported practice study were 94.6%/86.7%, 82.1%/68.4%, and 95.5%/91.1%, respectively. Given these data, it could be concluded that the difference of the medians is not reporting the difference between agreement and disagreement so much as it is reporting that the degree to which teachers report use of this practice in their own classrooms is statistically significantly less than the degree to which they report that they believe it is a best practice.

The results of these same analyses applied to the graded assignments (Part B) comparisons are reported in Table C.19.

Table C.19

*Comparisons of Research Questions 1 and 2 Results – Likert-type Items – Graded Assignments*

Item	Median	Mode	Mean Rank	Sig.	Difference
<i>A fully revised and edited essay/composition</i>					
Best Practice	3.00	3.00	122.62	.665	no
Own Practice	3.00	3.00	124.39		
<i>Short answer responses to reading/literature questions</i>					
Best Practice	2.00	2.00	121.69	.652	no
Own Practice	2.00	2.00	125.34		
<i>Reading checks/quizzes</i>					
Best Practice	2.00	2.00	119.63	.418	no
Own Practice	2.00	2.00	126.40		
<i>Grammar worksheets</i>					
Best Practice	1.00	1.00	124.28	.762	no
Own Practice	1.00	1.00	121.69		

*(table continues)*

*Comparisons of Research Questions 1 and 2 Results...Graded Assignments (continued).*

Writing samples that measure grammar/usage						
Best Practice	2.00	2.00	128.11			
Own Practice	2.00	2.00	116.89	.186		no
Spelling tests						
Best Practice	1.00	1.00	138.31			
Own Practice	0.00	0.00	107.31	.000		yes
Handwriting samples						
Best Practice	0.00	0.00	128.46			
Own Practice	0.00	2.00	118.46	.123		no
Multiple choice tests for reading/literature						
Best Practice	2.00	2.00	121.38			
Own Practice	2.00	2.00	124.66	.701		no
True/false tests						
Best Practice	1.00	1.00	129.67			
Own Practice	1.00	0.00	116.16	.118		no
Classroom discussions						
Best Practice	2.00	2.00	130.56			
Own Practice	2.00	2.00	116.33	.092		no
Writer's notebooks						
Best Practice	2.00	2.00	129.28			
Own Practice	2.00	2.00	116.66	.134		no
Class notes						
Best Practice	1.00	2.00	128.00			
Own Practice	1.00	1.00	114.79	.122		no
Presentations (in front of an audience)						
Best Practice	3.00	3.00	126.65			
Own Practice	3.00	3.00	120.30	.422		no

Again, the null hypothesis for this analysis was that the median difference between the pairs was zero, or  $H_0: Mdn_1 = Mdn_2$ . In cases where the p-value was smaller than alpha (p-value < .05), the null hypothesis was rejected, and I concluded that the median scores of the best practice responses were not equal to the median scores of the self-report practice responses.

Or, in other words, there was a statistically significant difference in the median scores where the p-value is less than or equal to .05.

For just one item in the graded assignments section of the survey a statistically significant difference was found between the median scores of reported best practice and self-reported practices. For *Spelling tests*, the best practice mode was a 1.0, and the actual practice mode was a 0.0. The summed frequencies for this for the levels of unimportance (in brief, *mostly unimportant* and *completely unimportant*) were 61.1%/81.0%, so the difference of the medians is not reporting the difference between importance or unimportance so much as it is reporting that the degree which teachers report to use this practice in their own classrooms is statistically significantly less than the degree to which they report that they believe it is a best practice.

#### *Likert Scale Item Comparisons*

The comparison of the Likert scale items was completed using studies of the means, standard deviations, standard errors of the mean, and an independent samples T-test (as is appropriate for interval scale data). A comparison was made between teachers' overall beliefs about grading systems and their overall self-reported practices. The 15 items in each of these queries were summed for the purpose of this comparison study. Table C.20 reports these results.

The null hypothesis for this analysis was that the mean difference between the pairs was zero, or  $H_0: M_1 = M_2$ . Because the p-value was smaller than alpha (p-value < .05), the null hypothesis was rejected, and I concluded that the mean scores of the best practice belief responses were not equal to the mean scores of the self-report practice responses. Or, in other

words, the degree to which teachers self-report using certain grading practices is statistically less than the degree to which teachers identify those practices as best practices. The results from Levine’s Test for Equality of Variances indicated ( $p = .147$ ) that this test meets the assumption of homogeneity of variances.

Table C.20

*Survey 1, Part A Results – Self-Reported Practices in Grading Systems, Likert Scale (Collective)*

	<i>M</i>	<i>SD</i>	<i>SEM</i>	Sig. (2-tailed)	<i>M</i> Difference
Best practice beliefs in grading systems	2.28	.798	.018		
Self-reported best practices in grading systems	2.22	.838	.019	.015	.064

### Research Question 3

In research question 3, I initiated an exploration of what high school English-language arts teachers’ grades actually measure (as opposed to what they say they believe about grades and what they self-report their own grades to measure). To collect data with regards to teachers’ actual practices, an interview/portfolio analysis process to analyze teachers’ grading records was utilized; this process employed a mixed-methods designs in the form of a standardized, open-ended interview regarding teachers’ actual grading practices as recorded in their grade books. The wording of each interview question and the sequence in which all interview questions were asked was pre-determined based on the results from research questions 1 and 2.

At this point in the process, the distinction between Part A and Part B of the questions was abandoned, and all elements of the grading practices – the items making up the study of the system (Part A) and the items making up the study of the graded assignments (Part B) – were amassed.

The survey instrument for research question 3 was comprised of 18 questions, and the sample was comprised of 39 members of the population who also participated in Survey 1. Question 1 was the informed consent agreement, and Question 2 made up the definitions for the survey. In Question 3, respondents were asked to select one class period of the day that was a typical representation of their English classes as a whole and to reference that class period during the most recently completed grading period of the school year.

Questions 4 and 5 of the interview were used to capture some characteristics of the respondents' work environments that affect grading programs. Data collected from Question 4 indicated that 25.6% of the respondents worked within a school system that used 6-weeks grading periods, and 74.4% of the respondents worked within a school system that used 9-weeks grading periods. No respondents indicated that they used grading periods of any other lengths. Figure 1 reports responses from Question 5 indicating the grade levels and courses represented by the respondents.

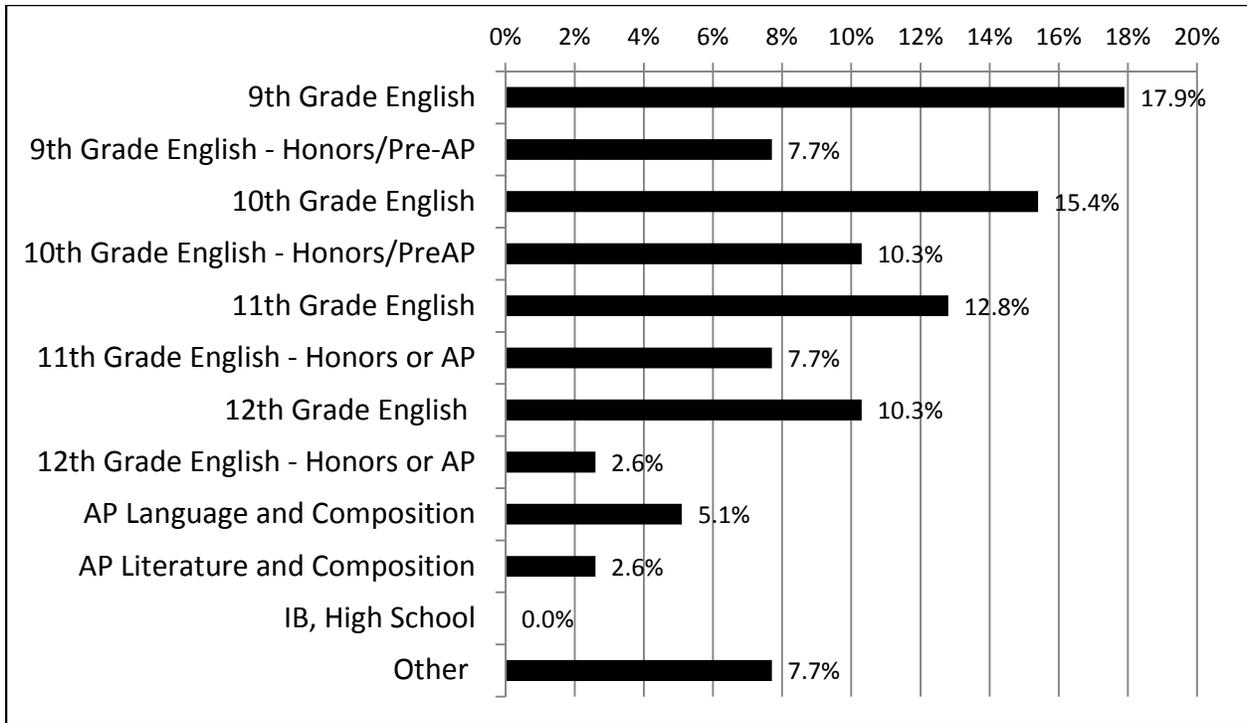


Figure C.1. Percentage of grade levels/courses represented in sample.

The “Other” courses indicated in Figure 1 making up the 7.7% of respondents in this sample included credit-bearing English language arts courses that had an additional level of specificity such as Dual Credit or Gifted/Talented.

The open-response items regarding the grading programs of the respondents were reported on an interval scale and were analyzed both descriptively using mean and standard deviation, and inferentially using standard error of the mean and the calculation of 95% confidence intervals. When appropriate and possible, the approximate percentage of grades affected by the topic (as calculated by considering the average number of grades recorded during a single grading period) was reported, as well. For conciseness in reporting, the topics listed in Table C.21 are abbreviations of the complete questions used during the interview/portfolio assessment; the actual and complete wording can be found in Appendix J.

Table C.21

*Interview/Portfolio Assessment Results – Questions 6-11 and 12-14*

Topic	<i>M</i>	<i>SD</i>	<i>SEM</i>	95% CI [LL, UL]	~ % of assignments
6. Number of assignments recorded during the grading period	12.23	3.07	.492	[11.27, 13.19]	--
7. Number of times teacher selectively adjusted students' final score	0.15	0.59	.094	[-.03, .33]	--
8. Number of assignments that represent few or no opportunities for practice	4.03	2.47	.395	[3.26, 4.80]	26%-39%
9. Number of times old evidence replaced by new evidence	2.87	2.54	.407	[2.07, 3.67]	--
10. Number of assignments measuring a skill or knowledge taught or learned in another class	3.00	2.84	.454	[2.11, 3.89]	17%-30%
12. Percentage of students earning a score of "0" on an assignment.	12.39	14.52	2.32	[7.84, 16.94]	--
13. Percentage of students earning a grade reduction for procedural reasons	8.15	9.69	1.55	[5.11, 11.19]	--
14. Number of assignments collected and graded following initial feedback	2.08	2.47	.396	[1.30, 2.86]	16%-23%

From these data, seven conclusions regarding actual grading systems can be drawn:

- Teachers selectively adjusted students' final scores at a rate of less than one-third of one student per grading period.

- On average, four or more assignments in a grading period (or 26%-39% of assignments) were collected and graded immediately following instruction, prior to a student having an opportunity to practice or master the skill or necessary knowledge.
- Old evidence of ability recorded in the grade book was replaced with new evidence of ability somewhere between 2 - 4 times during a grading period. Class sizes were not recorded in this study; however, based on an estimation of typical class sizes in high school English classes, the percentage of grades replaced by new evidence was less than 1%.
- Between 17% and 30% of the assignments graded and recorded relied on a skill or knowledge that was taught or learned in a previous class.
- Approximately 8% to 17% of students in a class earned a score of a zero on an assignment during a given grading period.
- Approximately 5% to 11% of students in a class earned a score reduction based on procedural (non-academic) reasons.
- On average, 1 - 3 assignments (or 16% to 23% of the assignments) were collected and graded following an opportunity for students to receive feedback.

Question 11 in the interview/portfolio analysis sought evidence of the causes for a student earning a recorded zero grade for an assignment. The responses to this open-response item were analyzed for patterns, organized into categories by theme, and then summed for the purpose of reporting. Each response was forced into a single category, with an “Other” category included for the management of responses that did not appear to be part of a pattern. The resulting data are reported out in Figure 2 using frequencies reported as percentages.

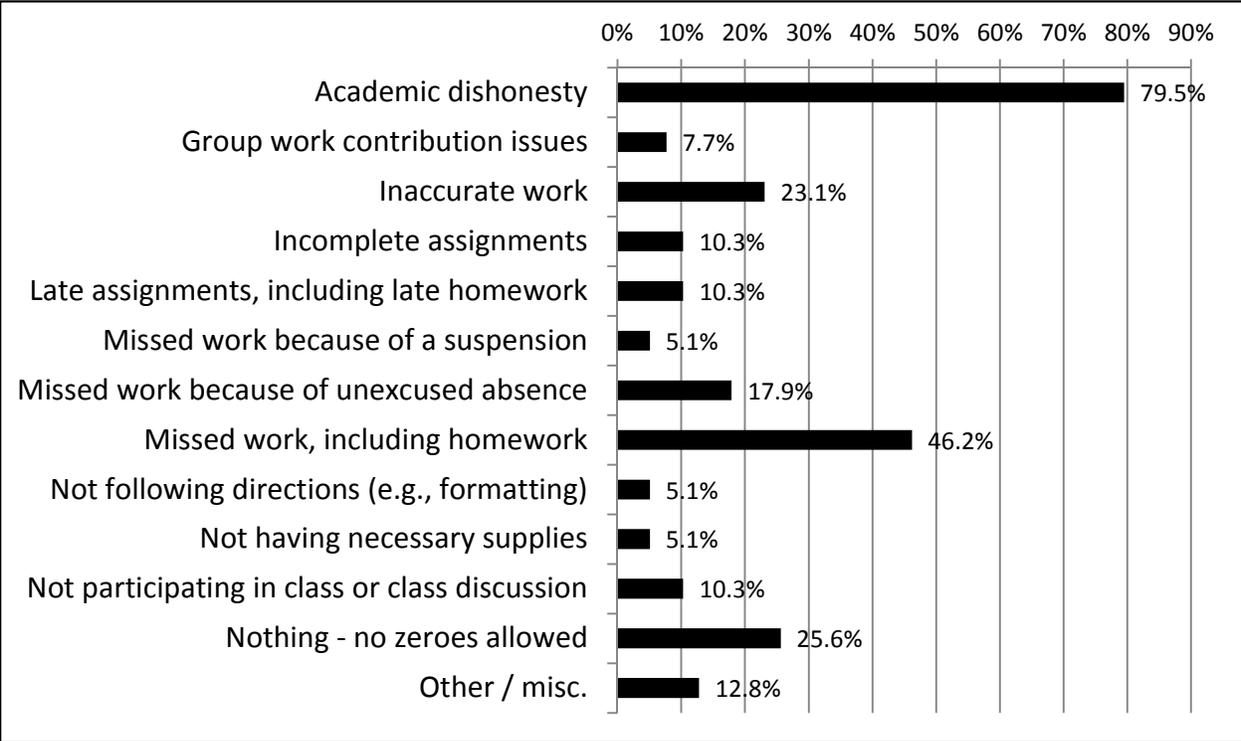


Figure C.2. (Question 11) Percentage of respondents assigning a “0” grade for reasons indicated.

The “Other” responses were either explanations of how zero grades were assigned (“notebook quiz grades are only 0, 70, 100”), responses using unfamiliar terminology (“not completing ZAPP work”), or responses that were only reported once during the interviews (e.g., “not attending required tutoring” or “not having work signed by parent when required”).

Though 25.6% of respondents indicated that no zeroes were allowed in their grade books, 79.5% also reported the use of zeroes for academic dishonesty. The sum of these two responses, though they are in conflict with one another, is 105.1%. This indicates that at least some of the respondents who indicate not using zeroes or not being allowed to use zeroes actually do use them at least to report or punish academic dishonesty.

These data reveal that 23.1% of the respondents used zeroes for academic reasons (i.e., *inaccurate work*), but zeroes were also used for non-academic reasons such as not having supplies or not following directions. Behavioral issues accounted for a portion of the zeroes used, including penalizing students with grades for days missed because of suspensions or unexcused absences and zeroes received for not following directions such as formatting headings on papers.

Question 15 in the interview/portfolio analysis asked respondents to indicate their reasons for recording and reporting grades. In this question, respondents selected all possible responses that applied and were offered an opportunity to add or qualify a response. Figure 3 presents the data collected in response to this question.

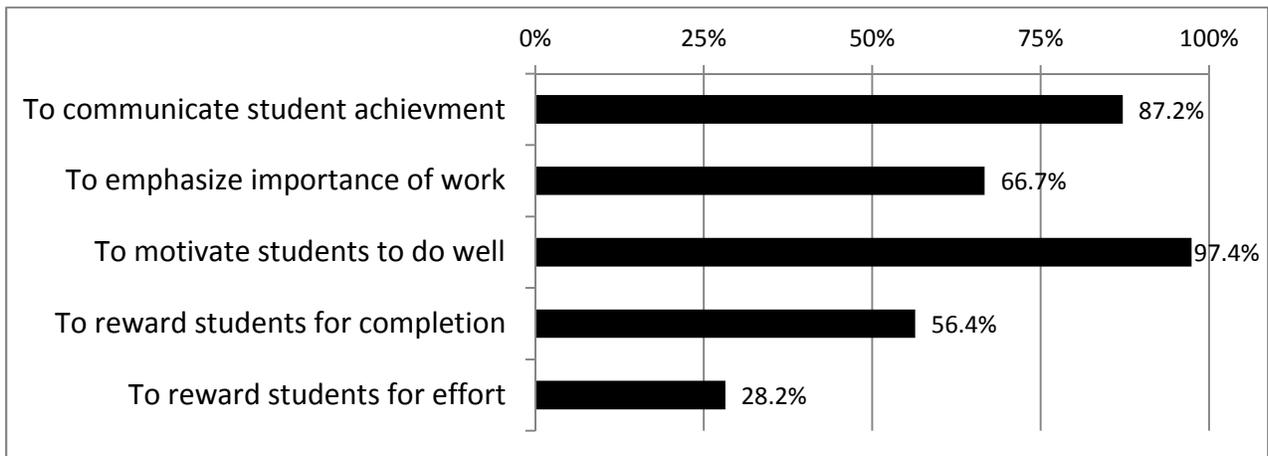


Figure C.3. (Question 15) Percentage of respondents using grades for reasons indicated.

From these data, a conclusion can be drawn that the primary aim for the use of grades was student motivation. By order of frequency, the next purpose for use was communication of student achievement. Because these two response rates were seemingly analogous, a one sample T-test between percentages was conducted to determine if there was a significant

difference between the percent of respondents who reported the use grades for student achievement reporting and the percent of respondents who reported the use of grading for motivating students. Data were analyzed using SPSS (Analyze > Compare Means > One Sample T-Test).

The null hypothesis for this analysis was that the difference of proportions between the two results was zero, or  $H_0: \text{Percent}_1 = \text{Percent}_2$ . Because the  $p$ -value of .000 (Sig., two-tailed) was less than alpha ( $p$ -value < .05), the null hypothesis was rejected, and I concluded that the percentage of the respondents using grades to report student achievement was not equal to the percentage of the respondents using grades to motivate students. Because the percentage was higher for teachers using grades to motivate students, I concluded that teachers use grades to motivate students more often than they do to report student achievement.

Other purposes indicated by teachers for use of grades were to emphasize the importance of work, to reward students for completion, and to reward students for doing well. One respondent advised that student achievement was reported as a purpose for grading because it provided a means of “determining which students have the highest achievement,” which I considered an indication of the use of grades for ranking purposes.

Questions 16 and 17 of the interview/portfolio analysis sought evidence of control of grading decisions. In Question 16, respondents were asked to report who controlled the number of grades collected in a grading period; in Question 17 they were asked to report who controlled decisions regarding which assignments would be collected and graded. These results are reported in comparison by answer choice in Figure 4.

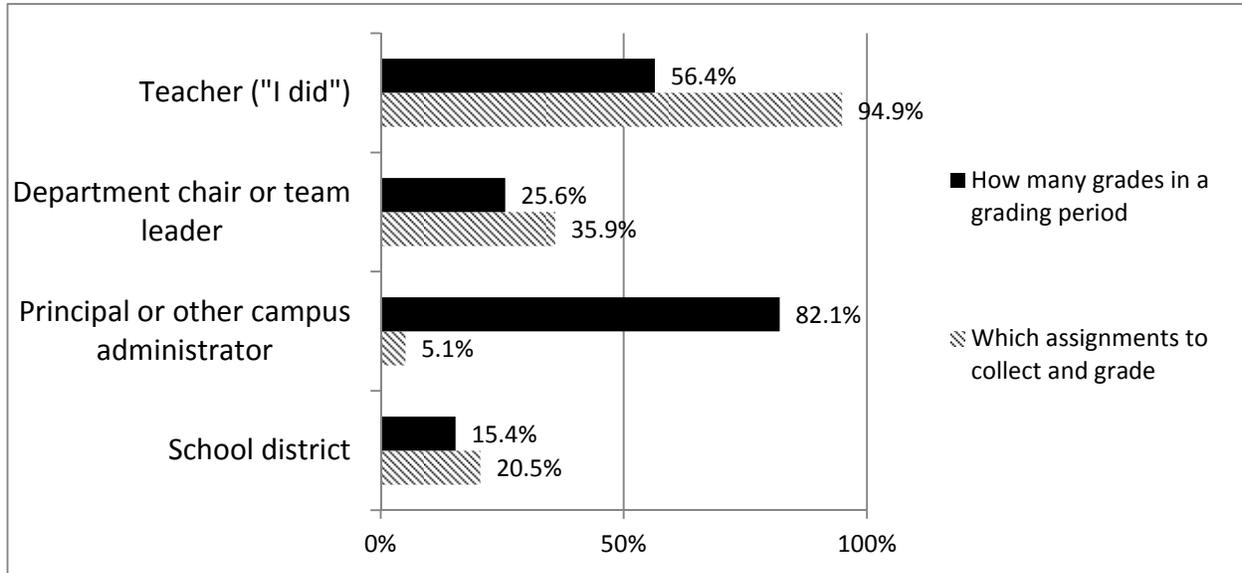


Figure C. 4. (Questions 16 and 17) Control of Grades

Results here indicate that with regards to determining the number of grades collected, teachers have a lesser degree of control than principals or other administrators but more control than the department chair/team leader or the school district. Additionally, with regards to which assignments were collected and graded, the teacher has more control, notably, than administrators.

And finally, Question 18 in the interview/portfolio assessment collected responses relevant to what types of assignments are graded and recorded. Based on analysis of entries in the grade book for the class and grading period selected, respondents indicated which types of assignments were recorded and represented in the students' final grades. Figure 5 relates these results in terms of frequency listed as percentages.

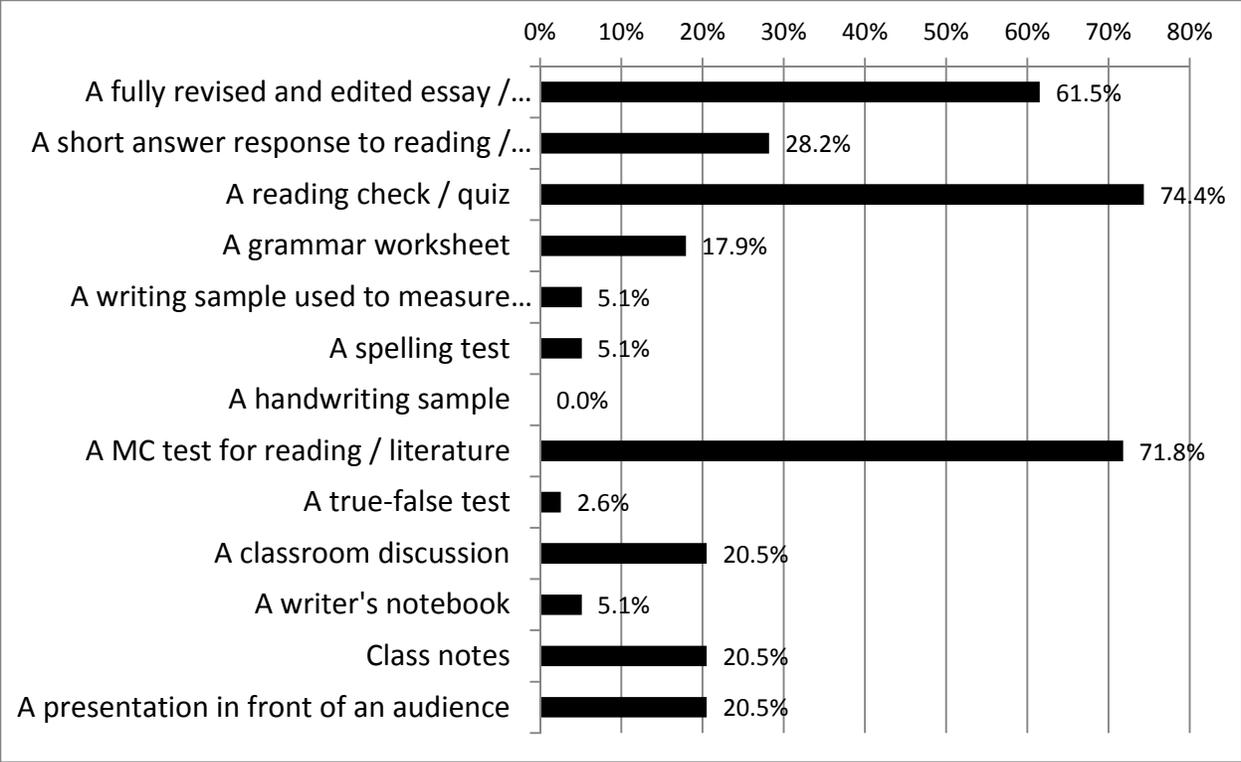


Figure C.5. (Question 18) Percentage of respondents grading and recording assignment types indicated.

Analyzing the significance of the results for Question 18 required a determination to be made regarding the frequency at which the assignment type would be considered “used” as a type of graded assignment by the sample.

This calculation was made in reference to the number of grading periods in the year which informed me as to the likelihood that an assignment used would be evidenced in the grading period referenced during the interview/portfolio analysis. While 74.4% of respondents reported working within a 9-weeks grading period, 26.6% reported using 6-weeks grading periods. A 9-weeks system is typically equal to four grading periods per year, and a 6-weeks system is typically equal to six grading periods per year. Consequently, a teacher within a 9-weeks period who does, in fact, use a particular type of assignment, would be expected to

record that assignment in a grade book at least once out of the four grading periods per year (or at 25%). Likewise, a teacher within a 6-weeks period who does, in fact, use a particular type of assignment, would be expected to record that assignment in a grade book at least once out of the six grading periods per year (or at a rate of 16.7%). An analysis of the mean percentage calculated in a ratio relative to the number of respondents reporting each type of grading period resulted in a determination that an assignment type used by a teacher in the sample would be reported at a rate of at least 23.1%.

Four of the assignment types were used at this rate, and the remaining ten assignment types were not.

#### Research Question 3 Results - Summary

In summation, the interview/portfolio analysis revealed that high school English-language arts teachers' grading practices include:

- Control over some grading decisions, i.e, *which* assignments were collected and graded
- Emphasis on feedback rather than grades (to a limited degree)

Their grading practices did not include:

- Selective adjustment of a student's final grade
- Replacing old evidence with new evidence in grade book
- Opportunities for practice and mastery before grading
- Control over some grading decisions, i.e, *how many* assignments are collected and graded
- Grading only on what was taught in class
- Separation of behavior grades from academic grades
- Forced completion of all assignments (no zeroes)

- Use of grades solely for communication of achievement

Teachers include these types of assignments in their students' final grades:

- A fully revised and edited essay/composition
- Short answer responses to reading/literature questions
- Reading checks/quizzes
- Multiple choice tests for reading/literature

They do not include these types of assignments in their students' final grades:

- Grammar worksheets
- A writing sample to measure grammar/usage
- A spelling test
- A handwriting sample
- A true/false test
- A classroom discussion
- A writer's notebook
- Class notes
- A presentation (in front of an audience)

#### Comparisons of Results – Research Questions 1, 2, and 3

An analysis of the relationship among results in research questions 1, 2, and 3 was conducted to determine if there were discrepancies between teachers' beliefs about grading practices, their self-reports about their own practices, and their actual practices. This analysis was essential at this point in time because the design of Survey 2 would be affected by the outcomes.

The results of this comparison are presented in Table C.22. In addition to the columns representing the results from the conducted studies (on beliefs, self-reported practices, and actual practices), an additional column labeled *Discrepancy* is employed for the purpose of indicating whether or not a discrepancy was found and, consequently, if the item in that row will be included in research question 4 work. A discrepancy would be considered to be in existence in the following scenarios:

- A reported belief was not self-reported in practice
- A reported belief was found in self-practice but not in actual practice
- A non-belief was self-reported in practice
- A non-belief is not self-reported in practice but was found in actual practice

Based on this comparative analysis of results from research questions 1-3, 13 items were found to have a discrepancy between or among what teachers believe, what they self-report to be their practices, and their actual practices.

Table C.22

*Identifying discrepancies in relationships among results from research questions 1-3*

Practice	Best Practice Belief	Self-Reported Practice	Actual Practice	Discrepancy
Selective adjustment of a student's final grade	Yes	Yes	No	Yes
Opportunities for practice and mastery before grading	Yes	Yes	No	Yes
Use of exemplars for understanding expectations	Yes	Yes	Other <sup>1</sup>	n/a
Avoiding calculating formative assessments into final grades	Yes	No	n/a	Yes
Replacing old evidence with new evidence in grade book	Yes	Yes	No	Yes
Grading only on what was taught in class	Yes	Yes	No	Yes
Separation of behavior grades from academic grades	Yes	Yes	No	Yes
Emphasis on feedback rather than grades	Yes	Yes	Yes	No
Weighting end of unit grades more than early evidence	Yes	Yes	Other <sup>2</sup>	n/a
Student self-assessment prior to grading	Yes	Yes	Other <sup>1</sup>	n/a
Student options for demonstrating understandings	Yes	Yes	Other <sup>1</sup>	n/a
Forced completion of all assignments (no zeroes)	Yes	Yes	No	Yes
Use of grades solely for communication of achievement	Yes	Yes	No	Yes
Grading decisions in teacher control				
(c) Which grades to record	Yes	Yes	Yes	No
(d) How many grades to record	Yes	Yes	No	Yes
Equal weighting of reading and writing in final grades	Yes	Yes	Other <sup>1</sup>	n/a
A fully revised and edited essay/composition	Yes	Yes	Yes	No
Short answer responses to reading/lit questions	Yes	Yes	Yes	No

*(table continues)*

*Identifying discrepancies in relationships among results from research questions 1-3 (continued).*

Reading checks/quizzes	Yes	Yes	Yes	No
Writing samples that measure grammar/usage	Yes	Yes	No	Yes
Multiple choice tests for reading/literature	Yes	Yes	Yes	No
Classroom discussions	Yes	Yes	No	Yes
Writer's notebook	Yes	Yes	No	Yes
Presentations (in front of an audience)	Yes	Yes	No	Yes
Class notes	No	No	No	No
True/false tests	No	No	No	No
Handwriting samples	No	No	No	No
Spelling tests	No	No	No	No
Grammar worksheets	No	No	No	No
First drafts of an essay/composition	Inconclusive	n/a	n/a	No
Multiple choice tests for grammar/usage	Inconclusive	n/a	n/a	No
Annotations made on texts or in books	Inconclusive	n/a	n/a	No

Other<sup>1</sup> indicates an item that was inadvertently left off of the interview/portfolio analysis and therefore cannot be studied further.

Other<sup>2</sup> indicates an item that was omitted mid-way through the interview/portfolio analysis when it was determined that the end of a grading period and the end of a unit of study might not be concurrent.

#### Research Question 4

Data in response to research question 4 were collected using Survey 2, called “Survey on Causes for Discrepancies between High School English Teacher Beliefs and Practices.” This survey was designed to capture teachers’ perceptions of the causes for the 13 discrepancies found to exist between teachers’ beliefs about best practices, their self-reported practices, and evidence of actual practice.

The process of reaching conclusions from the results from Survey 2 was broader and less restrictive in design than the process used to draw conclusions from the more quantitative-like data from Survey 1 and from the qualitative-quantitative design found in the interview/portfolio Analysis study. In one sense, the responses here were treated more as insights into bigger concepts and causes than as explicit and definitive reasons for why certain things happen in school and school systems. In other words, the results determined from Survey 2 were considered to be ways in which to see what, if any, political/power systems were influencing the implementation of best practices and/or and practices teachers believed to be essential.

The distinctions between the Part A and Part B elements of the study are worth resurrecting. Recall from research questions 1 and 2 that the original items considered in this study were defined as elements of a grading system (which was labeled Part A) or were classified simply as types of assignments that might be graded (which was labeled Part B). Where the Part A items were derived from expert recommendations on best practice, the Part B items were not. Here we are able to see if there are differences in the ways best practice recommendations are treated as opposed to general teacher practices.

There was no hypothesis being tested here except to say that I believed that teachers would be able to discuss and identify a variety of causes for the discrepancies between what they believe and what they say they do and what actual practice reveals.

Each item in Survey 2 was designed to accommodate both selected- and open-responses. There were no limitations on how many causes a respondent might select, and all respondents were provided the opportunity to qualify or add their own or additional responses. Responses were first analyzed qualitatively through a process that identified and characterizes trends in themes, and then these patterns were summed and are presented here using descriptive statistics including frequencies reported as percentages and mode as the measure of central tendency. This mixed method approach allows me to base conclusions on the trends I observe in the responses while also presenting evidence as to the degrees to which the responses support the conclusions.

In all, the presented causes for the observed discrepancies fell into one of six categories: (1) interference by someone with authority greater than the teacher, (2) interference by limitations imposed by the electronic grade book, (3) limited time for effective implementation, (4) limited student capacity (suggesting that students would have a difficult time engaging in or understanding the practice), (5) limited parent capacity (suggesting that parents would not be able to comprehend or support a practice without extensive explanations), and (6) limited teacher capacity (suggesting that teachers may believe in the practice but may not actually know how to implement it).

Table C.23 presents the perceived causes for each of the 13 discrepancies. Where an *X* indicates that a particular cause was identified as being a barrier to implementation by at least

1% of the respondents, an asterisk (\*) marks the perceived cause with the highest frequency of responses.

Table C.23

Causes for discrepancies, responses selected (X) and mode ( \* )as measure of central tendency

	Interference by authority	Interference by electronic grade book	Limited time available	Limited student capacity	Limited parent capacity	Limited teacher capacity
Causes for no selective adjustment of a student's final grade	*	X	X		X	X
Causes for no opportunities for practice or mastery before grading	X		X			*
Causes for the inclusion of formative assessment grades in final grades	*	X	X	X	X	
Causes for not replacing old evidence with new evidence in grade book	*	X	X	X	X	
Causes for not grading only what was taught in class	X					*
Causes for not separating behavior grades from academic grades	X			*		
Causes for not using forced completion of assignments (no zeroes)	X	X	*	X	X	
Causes for not using grades solely for communicating student achievement	X				X	*
Causes for not having control over deciding how many grades to take	*				X	X
Causes for not using writing samples that measure grammar/usage			*		X	X
Causes for not including classroom discussions in final grades			*			X
Causes for not using measures of the writer's notebook in final grades			*		X	X
Causes for not using student presentation grades in final grades			*	*	X	

The pattern emerging here indicates that the participants are likely to hold accountable most often one these three sources: campus or district authorities, the limited time available,

and the interferences caused by parents. The greatest cause among these three was identified as the limited time available. In greater summation, it could be said that respondents argue that there is not enough time to practice best and desired practices. The significant difference between the second and third indicated causes – campus and district authorities and parents – is that one of these two is arguably more culpable for having awareness of the need to implement certain practices than the other. Both campus and district administrators, it would seem, would be working within professional understandings of best practice and would be in positions to positively influence the functioning ability of a school system, and yet respondents find administrators as much of or more of a barrier than parents, students, and other teachers.

Respondents identify students as having the second to least degree of responsibility, with only blame going to the electronic grade book less often.

#### Research Question 5

In response to the final research question, a quantitative strategy was used with a quasi-experimental, counterbalanced design to determine if there were limitations in the data created by the order of questions presented in Survey 1. Specifically, the aim was to determine if participants who responded to questions about their beliefs about best grading practices were affected by the knowledge that the survey was also asking questions regarding their own actual grading practices.

In this research design, all participants received the same treatment – i.e. completion of Survey 1 – but in a different order. The instrumentation was manipulated in such that Treatment Group A and Treatment Group B varied with regards to which part of the survey

instrument was received and answered first – either questions regarding research question 1 (beliefs) or questions regarding research question 2 (self-reported) practices. The survey was designed with a logic feature that prevented respondents from returning to earlier questions to alter responses. Survey 1 was designed in two forms – 1.2 and 2.1 – and distribution was random through a process of distributing alternating survey addresses (URLs). Consequently, data were collected from two independent samples – the first sample (Treatment Group A) was comprised of 79 members of the population, and the second sample (Treatment Group B) was comprised of 50 members of the population.

A Mann-Whitney U Test was used to determine if there was a statistically significant difference of the means (as is appropriate for an ordinal scale) on a dependent variable (measure of best practice) between two groups of an independent variable (version 1.2 and version 2.1). The mean rank was used to report which grouping variable – version 1.2 or version 2.1 – indicated a stronger response toward agreement with experts on best practices. This data were analyzed using SPSS (Analyze > Nonparametric Tests > Legacy Dialogues > 2 Independent Samples). Table C.24 presents the results of this study.

Table C.24

*Research question 5 – Comparison of mean, versions 1.2 and 2.1, Likert Scale (Collective)*

	<i>M Rank</i>	<i>Sig. (2-tailed)</i>	<i>M Rank Difference</i>
Survey 1.2 – Self-reported grading systems results, following questions on best practices	1038.27	.128	37.20
Survey 2.1 – Self-reported grading system results, preceding questions on best practices	1001.07		

The null hypothesis for this analysis was that the mean difference between the treatment groups was zero, or  $H_0: M_1 = M_2$ . Because the  $p$ -value was greater than alpha ( $p$ -value .128), I failed to reject the null hypothesis, and concluded that the mean scores of the Treatment group 1 self-reported practice responses were not statistically different from the mean scores of the Treatment group 2 self-reported practice responses. Or, in other words, regardless of the order in which the questions of self-reported practices were asked – i.e., either prior to or following the questions regarding beliefs about best practices --, the response rates on the Likert scales were statistically similar.

APPENDIX D  
SURVEY AND INTERVIEW DOCUMENTS

## SURVEY ON HIGH SCHOOL ENGLISH TEACHERS' GRADING PRACTICES, SURVEY 1, FORM 1.2

### **Question 1**

#### Informed Consent Notice

Before agreeing to participate in this research study, it is important that you read and understand the following explanation of the purpose, benefits and risks of the study and how it will be conducted.

**Title of Study:** A Study of High School English Teachers' Personal Grading Beliefs and Self-Reported Grading Systems

**Student Investigator:** Lisa Thibodeaux, University of North Texas (UNT), College of Education.

**Supervising Investigator:** Dr. Carol Wickstrom

**Purpose of the Study:** You are being asked to participate in a research study which involves the study of grading practices used in high school English classrooms. The intention of this work is to invite teachers, as a whole, into a conversation around the role that grades and grading play in classrooms and schools and why these systems exist and are used thus. This study is designed to answer the following research questions: (a) What do high school English-language arts teachers believe a grade should measure? and (b) What do high school English-language teachers report their students' grades measure.

**Study Procedures:** You will be asked to complete a confidential on-line survey that will take approximately 15-20 minutes of your time.

**Foreseeable Risks:** No foreseeable risks are involved in this study.

**Benefits to the Subjects or Others:** This study is not expected to be of any direct benefit to you, but we hope to learn more about grading practices in high school English classes. Though much information is available about grading practices in general, this study will contribute to our field by identifying practices specific to the characteristics of assignments and instruction found only in high school English-language arts classes.

**Compensation for Participants:** None

**Procedures for Maintaining Confidentiality of Research Records:** This study will be kept confidential, and the records of this study will be kept private. No identifiers linking you to this study will be included in any sort of report that might be published. Research records will be stored securely and only Lisa Thibodeaux (the student investigator) and Dr. Carol Wickstrom (the supervising investigator) will have access to the confidential records.

Questions about the Study: If you have any questions about the study, you may contact me, Lisa Thibodeaux, at xxxxxxxxxxxx@gmail.com or my Supervising Investigator, Dr. Carol Wickstrom, at Carol.Wickstrom@unt.edu.

Review for the Protection of Participants: This research study has been reviewed and approved by the UNT Institutional Review Board (IRB). The UNT IRB can be contacted at (940) 565-3940 with any questions regarding the rights of research subjects.

Research Participants' Rights: Your participation in the survey confirms that you have read all of the above and that you agree to all of the following:

- Lisa Thibodeaux has explained the study to you and you have had an opportunity to contact her with any questions about the study. You have been informed of the possible benefits and the potential risks of the study.
- You understand that you do not have to take part in this study, and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits. The study personnel may choose to stop your participation at any time.
- You understand why the study is being conducted and how it will be performed.
- You understand your rights as a research participant and you voluntarily consent to participate in this study.
- You understand you may print a copy of this form for your records.

Agreement to Participate: Please indicate your agreement to participate in this study by clicking "Agree" below. If you choose not to participate, please indicate by clicking "Disagree" below, and then exit the survey. Participants must be 18 years of age or older. Thank you.

Agree  
 Disagree

## Question 2

### Definitions

For the purposes of this study, the following terms are used in the ways in which they are defined here:

Grades -- are the symbols that are assigned to student work. They may be numerical grades such as numbers on a 1-100 scale or 1-4 scale where the higher number represents the higher level of achievement, or they may be alpha grades which are most typically A, B, C, D, and F where A is the highest rating and F is the lowest. Grades will be used interchangeably to refer to grades given to a single assignment and grades that represent a marking period as recorded on a report card.

Grade book -- will be used in reference to the system a teacher uses to record assignments and students' grades on those assignments, whether it be a sophisticated electronic system or a simpler, more traditional pen-and-paper method for recording the information.

"Recorded as a grade" -- means that the work was assessed, assigned a grade, and recorded in a grade book

Grading period -- refers to the period of time determined by the campus or school district for which a single grade is recorded and reported to stakeholders. Grading periods can vary by district; most commonly grading periods are 6-weeks, 9-weeks, 12-weeks, or semester-long.

Best practice -- will be used when referring to instructional or grading practices that are considered by teachers and/or researchers to be the ones that are most likely to achieve the desired results.

English-language arts -- indicates a state-required (non-elective) course. In this study, this designation does NOT refer to reading only classes or other elective classes such as "Creative Writing" or "Literary Studies" or other standardized test preparation courses such as "SAT Prep" or state exam remediation courses.

Teachers -- will refer to those who are certified to teach the course and who are the designated record-keeper for the course. Co-teachers or support-teachers who do not bear the responsibility for grading, recording grades, and reporting grades to stakeholders should not participate in this study.

\_\_\_\_\_ I am a certified teacher of a state required English/language arts course in high school (grades 9-12).

\_\_\_\_\_ I am not a certified teacher of state required English/language arts course in high school (grades 9-12).

### Question 3

Please answer this question with regards to your personal beliefs about best grading practices that should be implemented by high school English teachers in general.

To what extent do you agree with each of the following statements?

3 = Completely agree

2 = Mostly agree, but with some reservations

1 = Mostly disagree, but with some reservations

0 = Completely disagree

\_\_\_\_\_ A teacher should adjust a student's report card grade if the final computation does not accurately reflect the teacher's personal knowledge of that student's success in the class.

\_\_\_\_\_ Before an assignment is graded, students should have an opportunity to practice and master the skill that is being measured.

\_\_\_\_\_ Before beginning an assignment, a student should be shown a well done model (e.g., a sample essay or project) in order to gain a better understanding of what is expected of him.

\_\_\_\_\_ Formative assessments – those assignments used to let the teacher know how the student is performing during the instruction and practice portion of the learning process – should be recorded but not factored into the final grade.

\_\_\_\_\_ It is more important to know how well a student mastered the designated content than when; therefore, new evidence of understandings should replace old evidence in the gradebook.

\_\_\_\_\_ Only what has been taught in class should be graded.

\_\_\_\_\_ Student tardies, mis-behaviors, absences, or other conduct issues should be considered when determining a student's academic grade.

\_\_\_\_\_ Students benefit more from meaningful feedback (written or oral) than from a single letter or number grade.

\_\_\_\_\_ Students should be evaluated more heavily on their work completed at the end of a unit of study because it reflects their eventual understandings.

\_\_\_\_\_ Students should be given an opportunity to self-assess and revise their work using a criterion or grading guide before the teacher grades it.

\_\_\_\_\_ Students who are given options for demonstrating their understandings perform better than students who are not.

\_\_\_\_\_ The best consequence for students who do not complete an assignment is to require them to complete the assignment.

\_\_\_\_\_ The purpose of a grade is to communicate a student's level of academic achievement to students, parents, and others. 

\_\_\_\_\_ The teacher should be ultimately in control of the grading decisions.

\_\_\_\_\_ Writing and reading achievement should be equally reflected in a student's grade.

#### **Question 4**

Please answer this question with regards to your personal beliefs about best grading practices that should be implemented by high school English teachers in general.

To what extent is each of these assignments important when considering what should count toward a student's grade in English?

3 = Very important; this work should always count toward a student's final grade

2 = Mostly important; this work should usually count toward a student's final grade

1 = Mostly unimportant; this work should not usually count toward a student's final grade

0 = Completely unimportant; this work should never count toward a student's final grade

\_\_\_\_\_ First drafts of an essay/composition

\_\_\_\_\_ A fully revised and edited essay/composition

\_\_\_\_\_ Short answer responses to reading/literature questions

\_\_\_\_\_ Reading checks/quizzes

\_\_\_\_\_ Grammar worksheets

\_\_\_\_\_ Writing samples that measure grammar/usage

\_\_\_\_\_ Spelling tests

\_\_\_\_\_ Handwriting samples

\_\_\_\_\_ Multiple choice tests for reading/literature

\_\_\_\_\_ Multiple choice tests for grammar/usage

\_\_\_\_\_ True/false tests

\_\_\_\_\_ Classroom discussions

\_\_\_\_\_ Writer's notebook

\_\_\_\_\_ Class notes

\_\_\_\_\_ Annotations made on texts or in books

\_\_\_\_\_ Presentations (in front of an audience)

### Question 5

Think of the English class(es) you currently teach, and please answer this question with regards only to your own current and typical grading practices.

To what extent do you agree with each of the following statements?

3 = Completely agree

2 = Mostly agree, but with some reservations

1 = Mostly disagree, but with some reservations

0 = Completely disagree

\_\_\_\_\_ I sometimes adjust a student's report card grade if the final computation does not accurately reflect my personal knowledge of the student's success in the class.

\_\_\_\_\_ Before an assignment is graded, my students should have an opportunity to practice and master the skill that is being measured.

\_\_\_\_\_ Before beginning an assignment, my students are shown a well done model (e.g., a sample essay or project) in order to gain a better understanding of what is expected of them.

\_\_\_\_\_ Formative assessments – those assignments I use to let me know how the student is performing during the instruction and practice portion of the learning process – are recorded but not factored into the final grade.

\_\_\_\_\_ It is more important for me to know how well a student mastered the designated content than when; therefore, new evidence of understandings replace old evidence in my grade book.

\_\_\_\_\_ Only what has been taught in my class is graded.

\_\_\_\_\_ Student tardies, mis-behaviors, absences, or other conduct issues are considered when I determine a student's academic grade.

\_\_\_\_\_ Students benefit more from meaningful feedback (written or oral) than when I give a single letter or number grade.

\_\_\_\_\_ Students in my class are evaluated more heavily on their work completed at the end of a unit of study because it reflects their eventual understandings.

\_\_\_\_\_ Students in my class are given an opportunity to self-assess and revise their work using a criterion or grading guide before I grade it.

\_\_\_\_\_ Students in my class are given options for demonstrating their understandings.

\_\_\_\_\_ When a student in my class does not complete an assignment, the consequence is that they have to complete the assignment.

\_\_\_\_\_ I use grades to communicate a student's level of academic achievement to the student, his parents, and others (such as counselors, college admissions officers, etc.)

\_\_\_\_\_ In my class, I am ultimately in control of the grading decisions.

\_\_\_\_\_ Writing and reading achievement are equally reflected in my students' grades.

### Question 6

Think of the English class(es) you currently teach, and please answer this question with regards only to your own current and typical grading practices.

To what extent do you find each of these assignments important when considering what should count toward your students' final grades in English?

3 = Very important; I always count this work toward a student's final grade

2 = Mostly important; I usually count this work toward a student's final grade

1 = Mostly unimportant; I usually do not count this work toward a student's final grade

0 = Completely unimportant; I never count this work toward a student's final grade

\_\_\_\_\_ First drafts of an essay/composition

\_\_\_\_\_ A fully revised and edited essay/composition

\_\_\_\_\_ Short answer responses to reading/literature questions

\_\_\_\_\_ Reading checks/quizzes

\_\_\_\_\_ Grammar worksheets

\_\_\_\_\_ Writing samples that measure grammar/usage

\_\_\_\_\_ Spelling tests

\_\_\_\_\_ Handwriting samples

\_\_\_\_\_ Multiple choice tests for reading/literature

\_\_\_\_\_ Multiple choice tests for grammar/usage

\_\_\_\_\_ True/false tests

\_\_\_\_\_ Classroom discussions

\_\_\_\_\_ Writer's notebook

\_\_\_\_\_ Class notes

\_\_\_\_\_ Annotations made on texts or in books

\_\_\_\_\_ Presentations (in front of an audience)

**Question 7**

To what extent are you familiar with each of the following organizations?

I am a member of this organization

I am not a member, but I am very familiar with the work of this organization.

I am not a member, but I am somewhat familiar with this organization.

I am not familiar with this organization whatsoever.

\_\_\_\_\_ NCTE (National Council of Teachers of English)

\_\_\_\_\_ IRA (International Reading Association)

\_\_\_\_\_ ASCD (Association for the Supervision of Curriculum Development)

**Question 8**

To what extent are you familiar with the following educational authors, instructors, and researchers?

I have an extensive understanding of his/her work.

I have a limited understanding of his/her work.

I know the name but am not otherwise familiar with his/her work.

I do not recognize the name.

\_\_\_\_\_ Jay McTighe

\_\_\_\_\_ Ric Stiggins

\_\_\_\_\_ Grant Wiggins

\_\_\_\_\_ Carol Ann Tomlinson

\_\_\_\_\_ Susan Brookhart

\_\_\_\_\_ Alfie Kohn

\_\_\_\_\_ Thomas Guskey

\_\_\_\_\_ Douglas Reeves

\_\_\_\_\_ Ken O'Connor

\_\_\_\_\_ Robert Marzano

**Question 9**

Survey Respondent Information (for data collection purposes only)

Age: \_\_\_\_\_

Age when you began teaching: \_\_\_\_\_

Years teaching: \_\_\_\_\_

State in which your school is located: \_\_\_\_\_

School name (to determine census classification only): \_\_\_\_\_

# of total students whose grades you are responsible for this grading period: \_\_\_\_\_

**Question 10**

CURRENT teaching assignments (grades and courses). Please check all that apply:

- 9<sup>th</sup> grade English
- 9<sup>th</sup> grade English – Honors or PreAP
- 10<sup>th</sup> grade English
- 10<sup>th</sup> grade English – Honors or PreAP
- 11<sup>th</sup> grade English
- 11<sup>th</sup> grade English Honors or PreAP
- 12<sup>th</sup> grade English
- 12<sup>th</sup> grade English – Honors or PreAP
- AP Language and Composition
- AP Literature and Composition
- International Baccalaureate, 9-12
- Other: \_\_\_\_\_

**Question 11**

PREVIOUS teaching assignments (grades and courses). Please check all that apply:

- 6<sup>th</sup> grade English/Language Arts
- 6<sup>th</sup> grade English/Language Arts – Honors or PreAP
- 7<sup>th</sup> grade English/Language Arts
- 7<sup>th</sup> grade English/Language Arts – Honors or PreAP
- 8<sup>th</sup> grade English/Language Arts
- 9<sup>th</sup> grade English/Language Arts – Honors or PreAP
- 9<sup>th</sup> grade English
- 9<sup>th</sup> grade English – Honors or PreAP
- 10<sup>th</sup> grade English
- 10<sup>th</sup> grade English – Honors or PreAP
- 11<sup>th</sup> grade English
- 11<sup>th</sup> grade English Honors or PreAP
- 12<sup>th</sup> grade English
- 12<sup>th</sup> grade English – Honors or PreAP
- AP Language and Composition
- AP Literature and Composition
- International Baccalaureate, 6-8
- International Baccalaureate, 9-12
- Other: \_\_\_\_\_

**Question 12**

A separate component of this research study can only be conducted through an interview (in person, phone, or email).

Won't you please consider volunteering for this additional step in this research study?

\_\_\_\_\_ Yes. I am willing to participate. Please contact me to further explain the process and arrange a time that is convenient for me. My contact information is below.

\_\_\_\_\_ No. I am unable to further participate in this study.

Name and email: \_\_\_\_\_

### **Question 1**

#### Informed Consent Notice

Before agreeing to participate in this research study, it is important that you read and understand the following explanation of the purpose, benefits and risks of the study and how it will be conducted.

Title of Study: A Study of High School English Teachers' Personal Grading Beliefs and Self-Reported Grading Systems

Student Investigator: Lisa Thibodeaux, University of North Texas (UNT), College of Education.

Supervising Investigator: Dr. Carol Wickstrom

Purpose of the Study: You are being asked to participate in a research study which involves the study of grading practices used in high school English classrooms. The intention of this work is to invite teachers, as a whole, into a conversation around the role that grades and grading play in classrooms and schools and why these systems exist and are used thus. This study is designed to answer the following research questions: (a) What do high school English-language arts teachers believe a grade should measure? and (b) What do high school English-language teachers report their students' grades measure.

Study Procedures: You will be asked to complete a confidential on-line survey that will take approximately 15-20 minutes of your time.

Foreseeable Risks: No foreseeable risks are involved in this study.

Benefits to the Subjects or Others: This study is not expected to be of any direct benefit to you, but we hope to learn more about grading practices in high school English classes. Though much information is available about grading practices in general, this study will contribute to our field by identifying practices specific to the characteristics of assignments and instruction found only in high school English-language arts classes.

Compensation for Participants: None

Procedures for Maintaining Confidentiality of Research Records: This study will be kept confidential, and the records of this study will be kept private. No identifiers linking you to this study will be included in any sort of report that might be published. Research records will be stored securely and only Lisa Thibodeaux (the student investigator) and Dr. Carol Wickstrom (the supervising investigator) will have access to the confidential records.

Questions about the Study: If you have any questions about the study, you may contact me, Lisa Thibodeaux, at xxxxxxxxxxxxxxxx@gmail.com or my Supervising Investigator, Dr. Carol Wickstrom, at Carol.Wickstrom@unt.edu.

Review for the Protection of Participants: This research study has been reviewed and approved by the UNT Institutional Review Board (IRB). The UNT IRB can be contacted at (940) 565-3940 with any questions regarding the rights of research subjects.

Research Participants' Rights: Your participation in the survey confirms that you have read all of the above and that you agree to all of the following:

- Lisa Thibodeaux has explained the study to you and you have had an opportunity to contact her with any questions about the study. You have been informed of the possible benefits and the potential risks of the study.
- You understand that you do not have to take part in this study, and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits. The study personnel may choose to stop your participation at any time.
- You understand why the study is being conducted and how it will be performed.
- You understand your rights as a research participant and you voluntarily consent to participate in this study.
- You understand you may print a copy of this form for your records.

Agreement to Participate: Please indicate your agreement to participate in this study by clicking "Agree" below. If you choose not to participate, please indicate by clicking "Disagree" below, and then exit the survey. Participants must be 18 years of age or older. Thank you.

Agree  
 Disagree

## Question 2

### Definitions

For the purposes of this study, the following terms are used in the ways in which they are defined here:

Grades -- are the symbols that are assigned to student work. They may be numerical grades such as numbers on a 1-100 scale or 1-4 scale where the higher number represents the higher level of achievement, or they may be alpha grades which are most typically A, B, C, D, and F where A is the highest rating and F is the lowest. Grades will be used interchangeably to refer to grades given to a single assignment and grades that represent a marking period as recorded on a report card.

Grade book -- will be used in reference to the system a teacher uses to record assignments and students' grades on those assignments, whether it be a sophisticated electronic system or a simpler, more traditional pen-and-paper method for recording the information.

"Recorded as a grade" -- means that the work was assessed, assigned a grade, and recorded in a grade book

Grading period -- refers to the period of time determined by the campus or school district for which a single grade is recorded and reported to stakeholders. Grading periods can vary by district; most commonly grading periods are 6-weeks, 9-weeks, 12-weeks, or semester-long.

Best practice -- will be used when referring to instructional or grading practices that are considered by teachers and/or researchers to be the ones that are most likely to achieve the desired results.

English-language arts -- indicates a state-required (non-elective) course. In this study, this designation does NOT refer to reading only classes or other elective classes such as "Creative Writing" or "Literary Studies" or other standardized test preparation courses such as "SAT Prep" or state exam remediation courses.

Teachers -- will refer to those who are certified to teach the course and who are the designated record-keeper for the course. Co-teachers or support-teachers who do not bear the responsibility for grading, recording grades, and reporting grades to stakeholders should not participate in this study.

\_\_\_\_\_ I am a certified teacher of a state required English/language arts course in high school (grades 9-12).

\_\_\_\_\_ I am not a certified teacher of state required English/language arts course in high school (grades 9-12).

### Question 3

Think of the English class(es) you currently teach, and please answer this question with regards only to your own current and typical grading practices.

To what extent do you agree with each of the following statements?

3 = Completely agree

2 = Mostly agree, but with some reservations

1 = Mostly disagree, but with some reservations

0 = Completely disagree

\_\_\_\_\_ I sometimes adjust a student's report card grade if the final computation does not accurately reflect my personal knowledge of the student's success in the class.

\_\_\_\_\_ Before an assignment is graded, my students should have an opportunity to practice and master the skill that is being measured.

\_\_\_\_\_ Before beginning an assignment, my students are shown a well done model (e.g., a sample essay or project) in order to gain a better understanding of what is expected of them.

\_\_\_\_\_ Formative assessments – those assignments I use to let me know how the student is performing during the instruction and practice portion of the learning process – are recorded but not factored into the final grade.

\_\_\_\_\_ It is more important for me to know how well a student mastered the designated content than when; therefore, new evidence of understandings replace old evidence in my grade book.

\_\_\_\_\_ Only what has been taught in my class is graded.

\_\_\_\_\_ Student tardies, mis-behaviors, absences, or other conduct issues are considered when I determine a student's academic grade.

\_\_\_\_\_ Students benefit more from meaningful feedback (written or oral) than when I give a single letter or number grade.

\_\_\_\_\_ Students in my class are evaluated more heavily on their work completed at the end of a unit of study because it reflects their eventual understandings.

\_\_\_\_\_ Students in my class are given an opportunity to self-assess and revise their work using a criterion or grading guide before I grade it.

\_\_\_\_\_ Students in my class are given options for demonstrating their understandings.

\_\_\_\_\_ When a student in my class does not complete an assignment, the consequence is that they have to complete the assignment.

\_\_\_\_\_ I use grades to communicate a student's level of academic achievement to the student, his parents, and others (such as counselors, college admissions officers, etc.)

\_\_\_\_\_ In my class, I am ultimately in control of the grading decisions.

\_\_\_\_\_ Writing and reading achievement are equally reflected in my students' grades.

#### **Question 4**

Think of the English class(es) you currently teach, and please answer this question with regards only to your own current and typical grading practices.

To what extent do you find each of these assignments important when considering what should count toward your students' final grades in English?

3 = Very important; I always count this work toward a student's final grade

2 = Mostly important; I usually count this work toward a student's final grade

1 = Mostly unimportant; I usually do not count this work toward a student's final grade

0 = Completely unimportant; I never count this work toward a student's final grade

\_\_\_\_\_ First drafts of an essay/composition

\_\_\_\_\_ A fully revised and edited essay/composition

\_\_\_\_\_ Short answer responses to reading/literature questions

\_\_\_\_\_ Reading checks/quizzes

\_\_\_\_\_ Grammar worksheets

\_\_\_\_\_ Writing samples that measure grammar/usage

\_\_\_\_\_ Spelling tests

\_\_\_\_\_ Handwriting samples

\_\_\_\_\_ Multiple choice tests for reading/literature

\_\_\_\_\_ Multiple choice tests for grammar/usage

\_\_\_\_\_ True/false tests

\_\_\_\_\_ Classroom discussions

\_\_\_\_\_ Writer's notebook

\_\_\_\_\_ Class notes

\_\_\_\_\_ Annotations made on texts or in books

\_\_\_\_\_ Presentations (in front of an audience)

### Question 5

Please answer this question with regards to your personal beliefs about best grading practices that should be implemented by high school English teachers in general.

To what extent to you agree with each of the following statements?

3 = Completely agree

2 = Mostly agree, but with some reservations

1 = Mostly disagree, but with some reservations

0 = Completely disagree

\_\_\_\_\_ A teacher should adjust a student's report card grade if the final computation does not accurately reflect the teacher's personal knowledge of that student's success in the class.

\_\_\_\_\_ Before an assignment is graded, students should have an opportunity to practice and master the skill that is being measured.

\_\_\_\_\_ Before beginning an assignment, a student should be shown a well done model (e.g., a sample essay or project) in order to gain a better understanding of what is expected of him.

\_\_\_\_\_ Formative assessments – those assignments used to let the teacher know how the student is performing during the instruction and practice portion of the learning process – should be recorded but not factored into the final grade.

\_\_\_\_\_ It is more important to know how well a student mastered the designated content than when; therefore, new evidence of understandings should replace old evidence in the gradebook.

\_\_\_\_\_ Only what has been taught in class should be graded.

\_\_\_\_\_ Student tardies, mis-behaviors, absences, or other conduct issues should be considered when determining a student's academic grade.

\_\_\_\_\_ Students benefit more from meaningful feedback (written or oral) than from a single letter or number grade.

\_\_\_\_\_ Students should be evaluated more heavily on their work completed at the end of a unit of study because it reflects their eventual understandings.

\_\_\_\_\_ Students should be given an opportunity to self-assess and revise their work using a criterion or grading guide before the teacher grades it.

\_\_\_\_\_ Students who are given options for demonstrating their understandings perform better than students who are not.

\_\_\_\_\_ The best consequence for students who do not complete an assignment is to require them to complete the assignment.

\_\_\_\_\_ The purpose of a grade is to communicate a student's level of academic achievement to students, parents, and others.

\_\_\_\_\_ The teacher should be ultimately in control of the grading decisions.

\_\_\_\_\_ Writing and reading achievement should be equally reflected in a student's grade.

### **Question 6**

Please answer this question with regards to your personal beliefs about best grading practices that should be implemented by high school English teachers in general.

To what extent is each of these assignments important when considering what should count toward a student's grade in English?

3 = Very important; this work should always count toward a student's final grade

2 = Mostly important; this work should usually count toward a student's final grade

1 = Mostly unimportant; this work should not usually count toward a student's final grade

0 = Completely unimportant; this work should never count toward a student's final grade

\_\_\_\_\_ First drafts of an essay/composition

\_\_\_\_\_ A fully revised and edited essay/composition

\_\_\_\_\_ Short answer responses to reading/literature questions

\_\_\_\_\_ Reading checks/quizzes

\_\_\_\_\_ Grammar worksheets

\_\_\_\_\_ Writing samples that measure grammar/usage

\_\_\_\_\_ Spelling tests

\_\_\_\_\_ Handwriting samples

\_\_\_\_\_ Multiple choice tests for reading/literature

\_\_\_\_\_ Multiple choice tests for grammar/usage

\_\_\_\_\_ True/false tests

\_\_\_\_\_ Classroom discussions

\_\_\_\_\_ Writer's notebook

\_\_\_\_\_ Class notes

\_\_\_\_\_ Annotations made on texts or in books

\_\_\_\_\_ Presentations (in front of an audience)

**Question 7**

To what extent are you familiar with each of the following organizations?

- I am a member of this organization
- I am not a member, but I am very familiar with the work of this organization.
- I am not a member, but I am somewhat familiar with this organization.
- I am not familiar with this organization whatsoever.

- \_\_\_\_\_ NCTE (National Council of Teachers of English)
- \_\_\_\_\_ IRA (International Reading Association)
- \_\_\_\_\_ ASCD (Association for the Supervision of Curriculum Development)

**Question 8**

To what extent are you familiar with the following educational authors, instructors, and researchers?

- I have an extensive understanding of his/her work.
- I have a limited understanding of his/her work.
- I know the name but am not otherwise familiar with his/her work.
- I do not recognize the name.

- \_\_\_\_\_ Jay McTighe
- \_\_\_\_\_ Ric Stiggins
- \_\_\_\_\_ Grant Wiggins
- \_\_\_\_\_ Carol Ann Tomlinson
- \_\_\_\_\_ Susan Brookhart
- \_\_\_\_\_ Alfie Kohn
- \_\_\_\_\_ Thomas Guskey
- \_\_\_\_\_ Douglas Reeves
- \_\_\_\_\_ Ken O'Connor
- \_\_\_\_\_ Robert Marzano

**Question 9**

Survey Respondent Information (for data collection purposes only)

- Age: \_\_\_\_\_
- Age when you began teaching: \_\_\_\_\_
- Years teaching: \_\_\_\_\_
- State in which your school is located: \_\_\_\_\_
- School name (to determine census classification only): \_\_\_\_\_
- # of total students whose grades you are responsible for this grading period: \_\_\_\_\_

**Question 10**

CURRENT teaching assignments (grades and courses). Please check all that apply:

- 9<sup>th</sup> grade English
- 9<sup>th</sup> grade English – Honors or PreAP
- 10<sup>th</sup> grade English
- 10<sup>th</sup> grade English – Honors or PreAP
- 11<sup>th</sup> grade English
- 11<sup>th</sup> grade English Honors or PreAP
- 12<sup>th</sup> grade English
- 12<sup>th</sup> grade English – Honors or PreAP
- AP Language and Composition
- AP Literature and Composition
- International Baccalaureate, 9-12
- Other: \_\_\_\_\_

**Question 11**

PREVIOUS teaching assignments (grades and courses). Please check all that apply:

- 6<sup>th</sup> grade English/Language Arts
- 6<sup>th</sup> grade English/Language Arts – Honors or PreAP
- 7<sup>th</sup> grade English/Language Arts
- 7<sup>th</sup> grade English/Language Arts – Honors or PreAP
- 8<sup>th</sup> grade English/Language Arts
- 9<sup>th</sup> grade English/Language Arts – Honors or PreAP
- 9<sup>th</sup> grade English
- 9<sup>th</sup> grade English – Honors or PreAP
- 10<sup>th</sup> grade English
- 10<sup>th</sup> grade English – Honors or PreAP
- 11<sup>th</sup> grade English
- 11<sup>th</sup> grade English Honors or PreAP
- 12<sup>th</sup> grade English
- 12<sup>th</sup> grade English – Honors or PreAP
- AP Language and Composition
- AP Literature and Composition
- International Baccalaureate, 6-8
- International Baccalaureate, 9-12
- Other: \_\_\_\_\_

**Question 12**

A separate component of this research study can only be conducted through an interview (in person, phone, or email).

Won't you please consider volunteering for this additional step in this research study?

\_\_\_\_\_ Yes. I am willing to participate. Please contact me to further explain the process and arrange a time that is convenient for me. My contact information is below.

\_\_\_\_\_ No. I am unable to further participate in this study.

Name and email: \_\_\_\_\_

## INTERVIEW/PORTFOLIO ANALYSIS: TEACHERS' GRADING PRACTICES IN HIGH SCHOOL ENGLISH

### **Question 1**

#### Informed Consent Notice

Before agreeing to participate in this research study, it is important that you read and understand the following explanation of the purpose, benefits and risks of the study and how it will be conducted.

Title of Study: A Study of High School English Teachers' Grade Books

Student Investigator: Lisa Thibodeaux, University of North Texas (UNT), College of Education.

Supervising Investigator: Dr. Carol Wickstrom

Purpose of the Study: You are being asked to participate in a research study which involves the study of grading practices used in high school English classrooms. The intention of this work is to invite teachers, as a whole, into a conversation around the role that grades and grading play in classrooms and schools and why these systems exist and are used thus. This study is designed to answer the following research question: What do high school English-language teachers' grades measure?

Study Procedures: You will be asked to complete an interview/portfolio analysis that will take approximately 20 minutes of your time.

Foreseeable Risks: No foreseeable risks are involved in this study.

Benefits to the Subjects or Others: This study is not expected to be of any direct benefit to you, but we hope to learn more about grading practices in high school English classes. Though much information is available about grading practices in general, this study will contribute to our field by identifying practices specific to the characteristics of assignments and instruction found only in high school English-language arts classes.

Compensation for Participants: None

Procedures for Maintaining Confidentiality of Research Records: This study will be kept confidential, and the records of this study will be kept private. No identifiers linking you to this study will be included in any sort of report that might be published. Research records will be stored securely and only Lisa Thibodeaux (the student investigator) and Dr. Carol Wickstrom (the supervising investigator) will have access to the confidential records.

Questions about the Study: If you have any questions about the study, you may contact me, Lisa Thibodeaux, at xxxxxxxxxxxxxxxx@gmail.com or my Supervising Investigator, Dr. Carol Wickstrom, at Carol.Wickstrom@unt.edu.

Review for the Protection of Participants: This research study has been reviewed and approved by the UNT Institutional Review Board (IRB). The UNT IRB can be contacted at (940) 565-3940 with any questions regarding the rights of research subjects.

Research Participants' Rights: Your participation in the survey confirms that you have read all of the above and that you agree to all of the following:

- Lisa Thibodeaux has explained the study to you and you have had an opportunity to contact her with any questions about the study. You have been informed of the possible benefits and the potential risks of the study.
- You understand that you do not have to take part in this study, and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits. The study personnel may choose to stop your participation at any time.
- You understand why the study is being conducted and how it will be performed.
- You understand your rights as a research participant and you voluntarily consent to participate in this study.
- You understand you may print a copy of this form for your records.

Agreement to Participate: Please indicate your agreement to participate in this study by responding with "agree" or "disagree."

Participants must be 18 years of age or older. Thank you.

\_\_\_\_\_ Agree  
\_\_\_\_\_ Disagree

## Question 2

### Definitions

For the purposes of this study, the following terms are used in the ways in which they are defined here:

Grades -- are the symbols that are assigned to student work. They may be numerical grades such as numbers on a 1-100 scale or 1-4 scale where the higher number represents the higher level of achievement, or they may be alpha grades which are most typically A, B, C, D, and F where A is the highest rating and F is the lowest. Grades will be used interchangeably to refer to grades given to a single assignment and grades that represent a marking period as recorded on a report card.

Grade book -- will be used in reference to the system a teacher uses to record assignments and students' grades on those assignments, whether it be a sophisticated electronic system or a simpler, more traditional pen-and-paper method for recording the information.

"Recorded as a grade" -- means that the work was assessed, assigned a grade, and recorded in a grade book

Grading period -- refers to the period of time determined by the campus or school district for which a single grade is recorded and reported to stakeholders. Grading periods can vary by district; most commonly grading periods are 6-weeks, 9-weeks, 12-weeks, or semester-long.

Best practice -- will be used when referring to instructional or grading practices that are considered by teachers and/or researchers to be the ones that are most likely to achieve the desired results.

English-language arts -- indicates a state-required (non-elective) course. In this study, this designation does NOT refer to reading only classes or other elective classes such as "Creative Writing" or "Literary Studies" or other standardized test preparation courses such as "SAT Prep" or state exam remediation courses.

Teachers -- will refer to those who are certified to teach the course and who are the designated record-keeper for the course. Co-teachers or support-teachers who do not bear the responsibility for grading, recording grades, and reporting grades to stakeholders should not participate in this study.

\_\_\_\_\_ I am a certified teacher of a state required English/language arts course in high school (grades 9-12).

\_\_\_\_\_ I am not a certified teacher of state required English/language arts course in high school (grades 9-12).

**Question 3**

Grade Report

Please select one class period of the day that is a typical representation of your English classes as a whole.

Look at the grade report for that class only for the most recently concluded grading period in this school year.

For example: 2<sup>nd</sup> period, 3<sup>rd</sup> six weeks

Have you located the grade report for the class period and the grading period you have selected?

**Question 4**

What is the length of the grading period for which you selected the grade report?

- \_\_\_\_\_ 6 week
- \_\_\_\_\_ 9 weeks / quarters
- \_\_\_\_\_ 12 weeks
- \_\_\_\_\_ 18 weeks / semester
- \_\_\_\_\_ Other: \_\_\_\_\_

**Question 5**

What grade level and course are represented by this grade report?

\_\_\_\_\_

**Question 6**

How many assignments total did you record in the grade book for this class during this grading period?

\_\_\_\_\_

**Question 7**

In this class and during this grading period, how many times did you replace a student's final report card score with a grade different than what was computed by the grade book?

\_\_\_\_\_

**Question 8**

Consider all of your students' grades in this class during this grading period. How many grades represent a student's first attempt at trying out a new skill or demonstrating knowledge learned that day or the previous day?

---

**Question 9**

Consider all of your student's grades in this class during this grading period. How many times did you delete a score you had already recorded for an assignment in order to record a higher score?

---

**Question 10**

How many assignments in the grade book measured a skill or knowledge that you believe should have been taught and learned in a previous class but not necessarily in your class?

---

**Question 11**

Which of these things might cause a student to earn a grade of a "0" in the grade book?

---

**Question 12**

How many students in this class in this grading period earned a score of a "0" on an assignment?

---

**Question 13**

Approximately how many students in this class in this grading period earned a deduction in score on an assignment for not following procedures (such as using a complete heading, submitting through turnitin.com, etc.) or for submitting late, sloppy, or incomplete work?

---

**Question 14**

How many assignments for this class during this grading period were collected and graded after you had a chance to provide the student with some initial feedback on his performance?

\_\_\_\_\_

**Question 15**

In this class during this grading period, for which of the following reasons did you use grades? (Check all that apply.)

- \_\_\_\_\_ To emphasize the importance of the assignment
- \_\_\_\_\_ To motivate students to pay attention and work hard
- \_\_\_\_\_ To reward students for completing the assignment
- \_\_\_\_\_ To communicate student achievement to the student and/or others
- \_\_\_\_\_ To reward students who tried their best

Additional answers or comments: \_\_\_\_\_

**Question 16**

In this class during this grading period, who decided which assignments you were going to collect and grade? (Check all that apply.)

- \_\_\_\_\_ The school district
- \_\_\_\_\_ The principal or other campus administrator
- \_\_\_\_\_ My team leader or department chair
- \_\_\_\_\_ I did

Additional answers or comments: \_\_\_\_\_

**Question 17**

In this class during this grading period, who decided how many grades you will have in the grading period? (Check all that apply.)

- \_\_\_\_\_ The school district
- \_\_\_\_\_ The principal or other campus administrator
- \_\_\_\_\_ My team leader or department chair
- \_\_\_\_\_ I did

Additional answers or comments: \_\_\_\_\_

**Question 18**

For this class in this grading period, please indicate which of the following types of assignments are represented in the grades in your grade book:

- \_\_\_\_\_ A first draft of an essay/composition
- \_\_\_\_\_ A fully revised and edited essay/composition
- \_\_\_\_\_ A short answer response to reading/literature questions
- \_\_\_\_\_ A reading check/quiz
- \_\_\_\_\_ A grammar worksheet
- \_\_\_\_\_ A writing sample that measures grammar/usage
- \_\_\_\_\_ A spelling test
- \_\_\_\_\_ A handwriting sample
- \_\_\_\_\_ A multiple choice test for reading/literature
- \_\_\_\_\_ A multiple choice test for grammar/usage
- \_\_\_\_\_ A true/false test
- \_\_\_\_\_ A class discussion
- \_\_\_\_\_ A writer's notebook
- \_\_\_\_\_ Class notes
- \_\_\_\_\_ An annotation made on texts or in books
- \_\_\_\_\_ A presentation (in front of an audience)

Additional answers or comments: \_\_\_\_\_

SURVEY ON CAUSES FOR DISCREPANCIES BETWEEN HIGH SCHOOL  
ENGLISH TEACHER BELIEFS AND PRACTICES, SURVEY 2

**Question 1**

Informed Consent Notice

Before agreeing to participate in this research study, it is important that you read and understand the following explanation of the purpose, benefits and risks of the study and how it will be conducted.

Title of Study: A Study of High School English Teachers' Grade Books

Student Investigator: Lisa Thibodeaux, University of North Texas (UNT), College of Education.

Supervising Investigator: Dr. Carol Wickstrom

Purpose of the Study: You are being asked to participate in a research study which involves the study of grading practices used in high school English classrooms. The intention of this work is to invite teachers, as a whole, into a conversation around the role that grades and grading play in classrooms and schools and why these systems exist and are used thus. This study is designed to answer the following research question: What are the perceived causes for the discrepancies between what high school English teachers believe grades should measure and what grades in the high school English class actually measure?

Study Procedures: You will be asked to complete an interview/portfolio analysis that will take approximately 20 minutes of your time.

Foreseeable Risks: No foreseeable risks are involved in this study.

Benefits to the Subjects or Others: This study is not expected to be of any direct benefit to you, but we hope to learn more about grading practices in high school English classes. Though much information is available about grading practices in general, this study will contribute to our field by identifying practices specific to the characteristics of assignments and instruction found only in high school English-language arts classes.

Compensation for Participants: None

Procedures for Maintaining Confidentiality of Research Records: This study will be kept confidential, and the records of this study will be kept private. No identifiers linking you to this study will be included in any sort of report that might be published. Research records will be stored securely and only Lisa Thibodeaux (the student investigator) and Dr. Carol Wickstrom (the supervising investigator) will have access to the confidential records.

Questions about the Study: If you have any questions about the study, you may contact me, Lisa Thibodeaux, at xxxxxxxxxxxxxxxx@gmail.com or my Supervising Investigator, Dr. Carol Wickstrom, at Carol.Wickstrom@unt.edu.

Review for the Protection of Participants: This research study has been reviewed and approved by the UNT Institutional Review Board (IRB). The UNT IRB can be contacted at (940) 565-3940 with any questions regarding the rights of research subjects.

Research Participants' Rights: Your participation in the survey confirms that you have read all of the above and that you agree to all of the following:

- Lisa Thibodeaux has explained the study to you and you have had an opportunity to contact her with any questions about the study. You have been informed of the possible benefits and the potential risks of the study.
- You understand that you do not have to take part in this study, and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits. The study personnel may choose to stop your participation at any time.
- You understand why the study is being conducted and how it will be performed.
- You understand your rights as a research participant and you voluntarily consent to participate in this study.
- You understand you may print a copy of this form for your records.

Agreement to Participate: Please indicate your agreement to participate in this study by responding with "agree" or "disagree."

Participants must be 18 years of age or older. Thank you.

\_\_\_\_\_ Agree  
\_\_\_\_\_ Disagree

## Question 2

### Definitions

For the purposes of this study, the following terms are used in the ways in which they are defined here:

Grades -- are the symbols that are assigned to student work. They may be numerical grades such as numbers on a 1-100 scale or 1-4 scale where the higher number represents the higher level of achievement, or they may be alpha grades which are most typically A, B, C, D, and F where A is the highest rating and F is the lowest. Grades will be used interchangeably to refer to grades given to a single assignment and grades that represent a marking period as recorded on a report card.

Grade book -- will be used in reference to the system a teacher uses to record assignments and students' grades on those assignments, whether it be a sophisticated electronic system or a simpler, more traditional pen-and-paper method for recording the information.

"Recorded as a grade" -- means that the work was assessed, assigned a grade, and recorded in a grade book

Grading period -- refers to the period of time determined by the campus or school district for which a single grade is recorded and reported to stakeholders. Grading periods can vary by district; most commonly grading periods are 6-weeks, 9-weeks, 12-weeks, or semester-long.

Best practice -- will be used when referring to instructional or grading practices that are considered by teachers and/or researchers to be the ones that are most likely to achieve the desired results.

English-language arts -- indicates a state-required (non-elective) course. In this study, this designation does NOT refer to reading only classes or other elective classes such as "Creative Writing" or "Literary Studies" or other standardized test preparation courses such as "SAT Prep" or state exam remediation courses.

Teachers -- will refer to those who are certified to teach the course and who are the designated record-keeper for the course. Co-teachers or support-teachers who do not bear the responsibility for grading, recording grades, and reporting grades to stakeholders should not participate in this study.

\_\_\_\_\_ I am a certified teacher of a state required English/language arts course in high school (grades 9-12).

\_\_\_\_\_ I am not a certified teacher of state required English/language arts course in high school (grades 9-12).

**Instructional Note:**

For the remainder of the survey, you are trying to help us understand why some teachers do not engage in the same effective practices that other teachers use.

You may select as many answers as apply.

**Question 3**

Some high school English teachers report that they believe they should adjust a student’s report card grade if the original computation does not accurately reflect their personal knowledge of that student’s achievement during the grading period.

Many teachers employ this practice. This question is trying to better understand why some teachers might not.

What do you think might be the reason that some teachers do not adjust a student’s grade?

- Their district and/or campus administration does not allow for this practice.
- Their team leader or other teacher-supervisor does not allow for this practice.
- The electronic grade book system does not allow for this practice.
- This practice is time-consuming.
- This practice is too hard to justify to parents.
- Other (please specify) \_\_\_\_\_

**Question 4**

Some high school English teachers report that they believe students should have time to practice and master a skill or knowledge before it is measured and a grade recorded.

Many teachers practice these behaviors. This question is trying to better understand why some teachers might not.

What do you think might be the reason that some teachers do not provide time for students to practice and master a skill or knowledge before it is graded and recorded?

- Their district and/or campus administration does not allow for this practice.
- Their team leader or other teacher-supervisor does not allow for this practice.
- This practice takes up too much class time.
- Many teachers do not realize how much practice is actually needed before mastery.
- Many teachers want to record early grades so they can be compared with later grades.
- Other (please specify) \_\_\_\_\_

**Question 5**

Some high school English teachers report that they believe that formative assessment information should be recorded but not factored into the student’s final grade.

Many teachers employ this practice. This question is trying to better understand why some teachers might not.

What do you think might be the reason that some teachers do not record formative assessment information outside of the students’ final grades?

- Their district and/or campus administration does not allow for this practice.
- Their team leader or other teacher-supervisor does not allow for this practice.
- The electronic grade book system does not allow for this practice.
- This practice takes is time-consuming.
- This practice is too hard to justify to parents.
- This practice is too hard to explain to students.
- Other (please specify) \_\_\_\_\_

**Question 6**

Some high school English teachers report that they believe that when a student shows that he has grown in his understandings, his previously recorded (lower) scores should be replaced with new (higher) scores in the grade book.

Many teachers employ this practice. This question is trying to better understand why some teachers might not.

What do you think might be the reason that some teachers do not replace old scores with new ones?

- Their district and/or campus administration does not allow for this practice.
- Their team leader or other teacher-supervisor does not allow for this practice.
- The electronic grade book system does not allow for this practice.
- This practice takes is time-consuming.
- This practice is too hard to justify to parents.
- Other (please specify) \_\_\_\_\_

**Question 7**

Some high school English teachers report that they believe that a student should only be graded on what he has been taught in his current English class.

Many teachers employ this practice. This question is trying to better understand why some teachers might not.

What do you think might be the reason that some teachers grade students on skills or content not actually taught in their current English class?

- Their district and/or campus administration does not allow for this practice.
- Their team leader or other teacher-supervisor does not allow for this practice.
- The electronic grade book system does not allow for this practice.
- It is easy to run out of time to teach everything, though it must be measured.
- Teachers expect that students learned pre-requisite concepts in their previous English class.
- Other (please specify) \_\_\_\_\_

**Question 8**

Some high school English teachers report that they believe that they should not include evidence of student conduct/behavior issues in their students' academic scores.

Many teachers employ this practice. This question is trying to better understand why some teachers might not.

What do you think might be the reason that some teachers do not avoid including conduct/behavior issues in students' academic scores?

- Their district and/or campus administration does not allow for this practice.
- The electronic grade book system does not allow for this practice.
- It is time-consuming to make this distinction between academic and behavior-related grades.
- Teachers do not necessarily know the difference between academic and behavior-related grades.
- Parents expect behavior grades to be reported in academic grades.
- Classroom management is improved when behavior grades are reported in academic grades.
- Other (please specify) \_\_\_\_\_

**Question 9**

Some high school English teachers report that they believe that when students do not complete their assignments, they should be given the “consequence” of having to complete it.

Many teachers employ this practice. This question is trying to better understand why some teachers might not.

What do you think might be the reason that some teachers do not require students to complete their missing work?

- Their district and/or campus administration does not allow for this practice.
- Their team leader or other teacher-supervisor does not allow for this practice.
- The electronic grade book system makes it too difficult to monitor this practice.
- This practice takes is time-consuming.
- This practice is too hard to justify to parents.
- Other (please specify) \_\_\_\_\_

**Question 10**

Some high school English teachers report that they believe that the primary purpose of grades should be to communicate student achievement to students and others.

Many teachers employ this practice. This question is trying to better understand why some teachers might not.

What do you think might be the reason that some teachers use grades to motivate or punish students?

- Their district and/or campus requirements force the use of grades for other purposes.
- Teachers do not know how to otherwise motivate students to do their work.
- There are not a lot of other ways teachers can motivate students to do their work.
- Parents expect teachers to use grades to motivate or punish.
- Other (please specify) \_\_\_\_\_

**Question 11**

Some high school English teachers report that they believe that the classroom teacher should ultimately be in control of grading decisions such as how many grades they should take during a single grading period.

Many teachers report that they *are* in control of these decisions. This question is trying to better understand why some teachers are not.

What do you think might be the reason that some teachers are not in control of their own grading decisions?

\_\_\_\_\_ Their district and/or campus administration takes control of these ideas to create equity and fairness among the classes.

\_\_\_\_\_ Their district and/or campus administration takes control of these ideas to minimize complaints from parents.

\_\_\_\_\_ Their team leader or other teacher-supervisor takes control of these decisions.

\_\_\_\_\_ Parents expect teachers to have the same grades as other teachers.

\_\_\_\_\_ Other (please specify) \_\_\_\_\_

**Question 12**

Some high school English teachers report that they believe that grades from student writing samples showing students' grammar and usage abilities should be recorded in the grade book and includes in students' final grades.

Many teachers employ this practice. This question is trying to better understand why some teachers might not.

What do you think might be the reason that some teachers do not include measures of grammar and usage via student writing samples in students' final grades?

\_\_\_\_\_ Their district and/or campus administration does not allow for this practice.

\_\_\_\_\_ Their team leader or other teacher-supervisor does not allow for this practice.

\_\_\_\_\_ This practice takes is time-consuming.

\_\_\_\_\_ This practice is too hard to justify to parents.

\_\_\_\_\_ It is easier to use a different method of measuring grammar skills, like grammar worksheets.

\_\_\_\_\_ Other (please specify) \_\_\_\_\_

**Question 13**

Some high school English teachers report that they believe grades from classroom discussions should be included in students' final grades.

Many teachers employ this practice. This question is trying to better understand why some teachers might not.

What do you think might be the reason that some teachers do not use grades from classroom discussions in students' final scores?

- Their district and/or campus administration does not allow for this practice.
- Their team leader or other teacher-supervisor does not allow for this practice.
- The electronic grade book system makes this difficult.
- This practice takes is time-consuming.
- This practice is too hard to justify to parents.
- Other (please specify) \_\_\_\_\_

**Question 14**

Some high school English teachers report that they believe grades from students' writer's notebooks should be included in students' final grades.

Many teachers employ this practice. This question is trying to better understand why some teachers might not.

What do you think might be the reason that some teachers do not include grades from writer's notebooks in students' final grades?

- Their district and/or campus administration does not allow for this practice.
- Their team leader or other teacher-supervisor does not allow for this practice.
- The electronic grade book system makes this difficult.
- This practice takes is time-consuming.
- This practice is too hard to justify to parents.
- Teachers don't necessarily understand how to measure this.
- It is easier to use a different method of measuring student work.
- Other (please specify) \_\_\_\_\_

**Question 15**

Some high school English teachers report that they believe that grades from students' presentations in front of a class should be included in students' final grades.

Many teachers employ this practice. This question is trying to better understand why some teachers might not.

What do you think might be the reason that some teachers do not include grades from students' presentations in students' final grades?

- Their district and/or campus administration does not allow for this practice.
- Their team leader or other teacher-supervisor does not allow for this practice.
- This practice takes is time-consuming.
- This practice is too hard to justify to parents.
- It is easier to use a different method of measuring grammar skills, like grammar worksheets.
- Other (please specify) \_\_\_\_\_

APPENDIX E  
REQUESTIONS FOR PARTICIPATION

## INVITATION TO PARTICIPATE IN SURVEY 1 PILOT STUDY

Hello, friends:

You may or may not know that I am a doctoral student working under the supervision of Dr. Carol Wickstrom in the College of Education at the University of North Texas.

I would like to invite you to participate in a pilot study that will inform my research on grading practices in high school English classes.

As a participant, you will be asked to complete a short on-line survey. This survey is comprised of the actual research survey I've built for the grading study, plus a request for you to measure the time it takes you to complete it, and a 4-question post-survey that will provide me with additional information about the survey content and your experience.

The responses you provide will not be made part of the study and will not be analyzed for content. There are no risks involved in this process.

If you are willing and able to participate, please access the survey at: \_\_\_\_.

If you have any questions now or during or after the completion of the survey, please contact me at this email address (xxxxxxxxxxxxxxxx@yahoo.com), or you may contact my advisor Dr. Carol Wickstrom at Carol.Wickstrom@unt.edu.

Thank you,  
Lisa Thibodeaux

## INVITATION TO PARTICIPATE IN SURVEY 2 PILOT STUDY

hello, friends:

You may or may not know that I am a doctoral student working under the supervision of Dr. Carol Wickstrom in the College of Education at the University of North Texas.

I would like to invite you to participate in a pilot study that will inform my research on grading practices in high school English classes.

If you previously participated in the pilot study for Survey 1, please know that this is a different survey altogether, and so you may still participate in this pilot study if you wish to.

As a participant, you will be asked to complete a short on-line survey, which includes 15 open-ended response questions, plus a request for you to measure the time it takes you to complete it, and a 4-question post-survey that will provide me with additional information about the survey content and your experience.

The responses you provide will not be made part of the study and will not be analyzed for content. There are no risks involved in this process.

If you are willing and able to participate, please access the survey at: \_\_\_\_.

If you have any questions now or during or after the completion of the survey, please contact me at this email address (xxxxxxxxxxxxxxxx@yahoo.com), or you may contact my advisor Dr. Carol Wickstrom at Carol.Wickstrom@unt.edu.

Thank you,  
Lisa Thibodeaux

## INVITATION TO PARTICIPATE IN SURVEY 1

### Email Recruitment / Direct Contact of Known Persons

ello friends:

You may or may not know that I am a doctoral student working under the supervision of Dr. Carol Wickstrom in the College of Education at the University of North Texas.

I would like to invite you to participate in my research study on grading practices in high school English classes. You may participate if you are currently a certified high school English teacher teaching at least one state required high school English class.

As a participant, you will be asked to complete a 15-20 minute on-line survey. There are no risks involved in this process, and your responses will remain confidential and anonymous. If you are willing and able to participate, please access the survey at: \_\_\_\_\_.

If you have any questions now or during or after the completion of the survey, please contact me at this email address (xxxxxxxxxxxxxxxx@yahoo.com), or you may contact my advisor Dr. Carol Wickstrom at Carol.Wickstrom@unt.edu.

Thank you,  
Lisa Thibodeaux

### Facebook Recruitment

Hello Facebook Friends: I am currently working under the supervision of Dr. Carol Wickstrom in the College of Education at the University of North Texas to conduct a research study on grading practices in high school English classes. (a) If you are currently a certified high school English teacher teaching at least one state required high school English class, and if you are willing to participate in this study, please access the survey at: \_\_\_\_\_. (b) Or – if you know someone who would qualify as a potential participant, will you please invite him/her to access the survey at the above web address by sharing this entire email with him/her? Page one of the survey provides additional information regarding risks, confidentiality, and purpose of the study. Thank you!

Twitter Recruitment / 140 character limit

HS English teachers willing to participate in my research study/survey on grading practices – please visit survey at: \_\_ All info on pg 1 of survey. Thanks!

Email Recruitment / Indirect Contact of Unknown Persons (via Direct Contact of Known Persons)

Hello friends:

You may or may not know that I am a doctoral student working under the supervision of Dr. Carol Wickstrom in the College of Education at the University of North Texas.

I would like to invite currently certified high school English teachers teaching at least one state required high school English class to participate in my research study on grading practices in high school English classes.

As you may know someone who would be a potential participating in this study, I am contacting you and asking you to please forward this entire email to those persons.

If you have any questions, please contact me at this email address (ThibodeauxGrading@yahoo.com), or you may contact my advisor Dr. Carol Wickstrom at Carol.Wickstrom@unt.edu.

Thank you,  
Lisa Thibodeaux

---

Hello potential participants:

In addition to the information above, I would like you to know that as a participant, you will be asked to complete a 15-20 minute on-line survey. There are no risks involved in this process, and your responses will remain confidential and anonymous.

If you are willing and able to participate, please access the survey at: \_\_\_\_\_.

If you have any questions now or during or after the completion of the survey, please contact me at this email address (ThibodeauxGrading@yahoo.com), or you may contact my advisor Dr. Carol Wickstrom at Carol.Wickstrom@unt.edu.

Thank you,  
Lisa Thibodeaux

## INVITATION TO PARTICIPATE IN INTERVIEW/PORTFOLIO ANALYSIS

Dear \_\_\_\_\_:

Thank you for indicating that you would be willing to participate further in the research study on grading practices.

If you are still willing and able, please respond to this email to let me know the best phone number for reaching you and any times in the next week that would be most convenient for you.

The interview/grade book analysis takes approximately 20 minutes to complete. You will need access to your grade book so that you can refer to assignments and recorded grades. You will not be asked to disclose any personal information about yourself or any of your students.

There are no risks involved in this process, and your responses will remain confidential.

If you have any questions now or during or after the completion of the interview/grade book analysis, please contact me at this email address (xxxxxxxxxxxxxxxxx@yahoo.com), or you may contact my advisor Dr. Carol Wickstrom at Carol.Wickstrom@unt.edu.

Thank you,  
Lisa Thibodeaux

## INVITATION TO PARTICIPATE IN SURVEY 2

### Facebook Recruitment

Hello Facebook Friends: I am currently working under the supervision of Dr. Carol Wickstrom in the College of Education at the University of North Texas to conduct a research study on grading practices in high school English classes. (a) If you are currently a certified high school English teacher teaching at least one state required high school English class, and if you are willing to participate in this study, please access the survey at: \_\_\_\_\_. (b) Or – if you know someone who would qualify as a potential participant, will you please invite him/her to access the survey at the above web address by sharing this entire email with him/her? Page one of the survey provides additional information regarding risks, confidentiality, and purpose of the study. Thank you!

### Twitter Recruitment / 140 character limit

HS English teachers willing to participate in my research study/survey on grading practices – please visit survey at: \_\_\_ All info on pg 1 of survey. Thanks!

### NCTE Teaching and Learning Forum Recruitment

Hello NCTE Friends: I am currently working on a research study on grading practices in high school English classes. If you are currently a certified HS Eng teacher and are willing to participate, please access the survey at: \_\_\_\_\_ (Page 1 of the survey provides more info.) Many thanks.

Email Recruitment / Indirect Contact of Unknown Persons (via Direct Contact of Known Persons)

Hello friends:

You may or may not know that I am a doctoral student working under the supervision of Dr. Carol Wickstrom in the College of Education at the University of North Texas.

I would like to invite currently certified high school English teachers teaching at least one state required high school English class to participate in my research study on grading practices in high school English classes.

As you may know someone who would be a potential participating in this study, I am contacting you and asking you to please forward this entire email to those persons.

If you have any questions, please contact me at this email address (xxxxxxxxxxxxxxxx@yahoo.com), or you may contact my advisor Dr. Carol Wickstrom at Carol.Wickstrom@unt.edu.

Thank you,  
Lisa Thibodeaux

---

Hello potential participants:

In addition to the information above, I would like you to know that as a participant, you will be asked to complete a 15-20 minute on-line survey. There are no risks involved in this process, and your responses will remain confidential and anonymous.

If you are willing and able to participate, please access the survey at: \_\_\_\_\_.

If you have any questions now or during or after the completion of the survey, please contact me at this email address (xxxxxxxxxxxxxxxx@yahoo.com), or you may contact my advisor Dr. Carol Wickstrom at Carol.Wickstrom@unt.edu.

Thank you,  
Lisa Thibodeaux

## Comprehensive Reference List

- Aaronson, E., Foley, T., Holmes, J. & Wallowitz, J. (1994). *Teacher-pleasing: Traditional grading--and learning? A collaborative qualitative study*. Paper presented at the Conference of the International Society for Exploring Teaching Alternatives.
- Airasian, P. W. (1984). *Classroom assessment and educational improvement*. Paper presented at Classroom Assessment: A Key to Educational Excellence, Northwest Regional Educational Laboratory, Portland, Oregon.
- Airasian, P. W. (2000). *Assessment in the classroom: A concise approach* (2<sup>nd</sup> ed.). Boston: McGraw-Hill.
- Allen, J. D. & Lambating, J. (2001, April). *Validity and reliability in assessment and grading: Perspectives of preservice and inservice teachers and teacher education professors*. Paper presented at the Annual Meeting of the American Educational Research Association, Seattle, WA.
- Anderson, L. W. (2003). *Classroom assessment: Enhancing the quality of teacher decision making*. Mahwah, NJ: Lawrence Erlbaum Publishers.
- Applebee, A. & Langer, J. (2009). What is happening in writing instruction. *English Journal*, 98 (5), 18-28.
- Barnwell, P. (2008). Could standard grading practices be counterproductive? *Education Week*. Retrieved from [http://www.edweek.org/ew/articles/2008/06/30/43barnwell-com\\_web.h27.html](http://www.edweek.org/ew/articles/2008/06/30/43barnwell-com_web.h27.html).

- Barnwell, P. (2009). "What's in an 'A'"? *NEA Today*. Retrieved from [www.nea.org/home/20705.htm](http://www.nea.org/home/20705.htm).
- Barnwell, P. (2010, February 25). Class, take our your cellphones! [*Education Week* Commentary]. Retrieved from [http://questionsforschools.org/?page\\_id=18](http://questionsforschools.org/?page_id=18).
- Bishop, J. H. (1992). "Why U.S. students need incentives to learn." *Educational Leadership* 49(6), 15–18.
- Black, P. & William, D. (1998). Assessment and classroom learning. *Educational Assessment: Principles, Policy, and Practice*, 5(1), 7–74.
- Black, P., C. Harrison, C. Lee, B. Marshall, and D. Williams. (2003). *Assessment for learning*:
- Bonesronning, H. (2004). Do teachers' grading practices affect student achievement? *Educational Economics*, 12(2), 151-167.
- Boone, H. & Boone, D. (2012, April). Analyzing Likert data. *Journal of Extension*, 50(2).
- Brantlinger, E. A. (1993). *The politics of social class in secondary school: Views of affluent and impoverished youth*. New York: Teachers College Press.
- Britzman, D. (1991). *Practice makes practice: A critical study of learning to teach*. Albany, New York: State University of New York Press.
- Brookhart, S. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice*, 10(1), 35-36.
- Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement*, 30(2), 123-142.
- Brookhart, S. M. (1994). Teachers grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279-301.

Brookhart, S. M. (1998, April). *Teaching about grading and communicating assessment results*.

Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Diego, CA.

Bruner, J. S. (1966). *Towards a theory of instruction*. Cambridge, MA: Harvard University Press.

Carlson, L. A. (2003). Beyond assessment to best grading practice: Practical guidelines. In J. E.

Wall & G. W. Walz (Eds.), *Measure up: Resources on testing for teachers, counselors, and administrators*, (pp. 507-515). Austin, TX: PRO-ED.

Cizek, G. J. (1997). Learning, achievement, and assessment: Constructs at a crossroads. In G. D.

Phye (Ed.), *Handbook of classroom assessment*. NY: Academic Press.

Clymer, J. B. & William, D. (2007). Improving the way we grade science. *Educational Leadership*,

64(4), 36-42.

College Board Office of Research and Development (1998, May). *High school grading practices*.

Retrieved from [http://professionals.collegeboard.com/profdownload/pdf/high\\_school\\_grading\\_10506.pdf](http://professionals.collegeboard.com/profdownload/pdf/high_school_grading_10506.pdf).

College Board. (2011). *Class rank & college admissions*. Retrieved from

<http://professionals.collegeboard.com/guidance>.

Common Core State Standards Initiative (2011). Retrieved from

<http://www.corestandards.org/resources/frequently-asked-questions>.

Cross, L. H. & Frary, R. B. (1999). Hodgepodge grading: Endorsed by students and teachers alike.

*Applied Measurement in Education*, 12(1), 53-72.

Cross, L. H. (1995). Grading students. *Practical Assessment, Research & Evaluation*, 4(8), 1-4.

District Grading Policy, S. B. No. 2033, Enacted by the Legislature of the State of Texas, Section 1, Subchapter B, Chapter 28, Education Code (2009). Retrieved from

<http://www.legis.state.tx.us/tlodocs/81R/billtext/html/SB02033F.htm>.

Dornbusch, S. M. & Ritter, P. L. (1987). When effort in school does not produce better grades. *Family Perspective*, 21, 285-298.

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

Erickson, J. A. (2010). Grading practices: The third rail. *Principal Leadership*, (10)7, 22-24.

Feldman, A., Kropf, A., & Alibrandi, M. (1996, April). *Making grades: How high school science teachers determine report card grades*. Paper presented at the annual meeting of the American Educational Research Association, New York.

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

Francis, R. W. (2006, Fall). Common errors in calculating final grades. *NEA Higher Educational Journal*, 11, 15-22.

Frary, R. B., Cross, L. H., & Weber, L. J. (1992, April 21-23). *Testing and grading practice and opinions in the nineties: 1890s or 1990s?* Paper presented at the annual meeting of the National Council on Measurement in Education. San Francisco, CA.

Goals 2000: Educate America Act, H.R. 1804, 103d Cong. (1994)

Guskey, T. (2010). Lessons of mastery learning. *Educational Leadership*, 68(2), 53-57.

Guskey, T. R. (Ed.) (2009). *Practical solutions for serious problems in standards-based grading*. Thousand Oaks, CA: Corwin Press.

- Guskey, T.R. and Bailey, J. M. (2001) *Developing Grading and Reporting Systems for Student Learning*. Corwin Press; Thousand Oaks, California.
- Jamison, S. (2004). Likert scales: How to (ab)use them. *Medical Education*, 38, 1217-1218.
- Kirschenbaum, H., Simon, S., & Napier, R. W. (1971). *Wad-ja-get?* NY: Hart.
- Kohn, A. (1999, March). From grading to de-grading. *High School Magazine*, 38-43.
- Krosnick, J. A., et al. (2002). The impact of 'no opinion' response options on data quality: Nonattitude reduction or invitation to satisfice? *Public Opinion Quarterly*, 66, 371–403.
- Kuzon, W.M., Jr., Urbanchek, M.G., & McCabe, S. (1996). The seven deadly sins of statistical analyses. *Annals of Plastic Surgery*, 37, 265-272.
- Lamb, T., Allen, G., Green, K. (2010, May). Is 'neutral' on a Likert scale the same as 'don't know' for informed and uninformed respondents? Effects of serial position and labeling on selection of response options. Paper presented at the annual meeting of the *National Council on Measurement in Education*. Denver, CO.
- Likert, R. (1932). A Technique for the Measurement of Attitudes. *Archives of Psychology*, 140, 1–55.
- Lipman, P. (1997). Restructuring in context: A case study of teacher participation and the dynamics of ideology, race, and power. *American Educational Research Journal*, 31, 3-37.
- Mansfield, H. C. (2001). Grade inflation: It's time to face the facts. *Chronicle of Higher Education*. Retrieved from <http://chronicle.com/free/v47/i30/30b02401.htm>.
- Marzano, R. J. (2001). *Classroom instruction that works*. Alexandria, VA: Association for Supervision and Curriculum Development.

- Marzano, R. J. (2009). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.
- McMillan, J. H. & Workman, D. J. (1998). *Classroom assessment and grading practices: A review of the literature*. Paper presented at the Metropolitan Educational Research Consortium. Richmond, VA.
- McMillan, J. H. (1997). *Classroom assessment: Principles and practice for effective instruction*. Boston: Allyn & Beacon.
- McMillan, J. H. (2008) *Assessment essentials for standards-based education*. Thousand Oaks, CA: Corwin Press.
- McTighe, J. & O'Connor, K. (2005). Seven practices for effective learning. *Educational Leadership*, 63(3), 10-17.
- McTighe, J. & Tomlinson, C. A. (2006). *Integrating differentiated instruction and understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.
- National Association of Secondary School Principals. (2010, March). *Class rank, GPA, and grading* (NASSP Board Position Statements). Retrieved from [www.nassp.org/Content.aspx?topic=61042](http://www.nassp.org/Content.aspx?topic=61042).
- National Council on Measurement in Education. (1995). *Code of professional responsibilities in educational measurement*. Retrieved from [www.natd.org/Code\\_of\\_Professional\\_Responsibilities.html](http://www.natd.org/Code_of_Professional_Responsibilities.html).
- NCTE High School Matters. (2008, March 28). What do grades really measure? [Web log post]. Retrieved from <http://nctesecondary.blogspot.com/2008/03/what-do-grades-really-measure.html>.

Nitko, A. J. (1991). Editorial: What are we teaching teachers about assessment and why?

*Educational Measurement: Issues & Practice*, 10(1), 2.

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, Stat. 1425 (2002).

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York: McGraw-Hill.

O'Connor, K. (2006). *A repair kit for grading: 15 fixes for broken grades*. Portland: Educational Testing Institute.

O'Connor, K. (2009). *How to grade for learning*. Thousand Oaks, CA: Corwin Press.

Pearson-Casanave, C. & Sosa, M. (2007). *Respite for teachers: Reflection and renewal in the teaching life*. Ann Arbor, MI: University of Michigan Press.

Pink, D. (2011). *Drive: The surprising truth about what motivates us*. NY: Riverhead Books.

Pollio, H. R. & Beck, H. P. (2000). When the tail wags the dog: Perceptions of learning and grade orientation in and by contemporary college students and faculty. *Journal of Higher Education*, 7(1), 84-102.

Reeves, D. (2008, Feb.) Effective grading practices. *Educational Leadership*, (65)5, 85-87.

Rich, R. H. (2001). Hidden factors in teachers' secondary grading practices. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses.

Robinson, G. E., and Craver, J. M. (1989). *Assessing and grading student achievement*. Arlington, VA: Educational Research Service, 1989.

Snowman, J., Biehler, R. F., & Bonk, C. J. (2000). *Psychology applied to teaching*, (9<sup>th</sup> ed.). Boston: Houghton Mifflin.

Sousa, D. A. (1998). Is the fuss about brain research justified? *Education Week*, (18)16, 35-52.

Retrieved from <http://www.edweek.org/ew/articles/1998/12/16/16sousa.h18.html>.

- Southeast Comprehensive Center (SECC) at the Southwest Educational Development Laboratory (SEDL). (2009, August 12). *Rapid response: Statewide uniform grading practices*. Retrieved from [http://secc.sedl.org/orc/rr/secc\\_rr\\_00094.pdf](http://secc.sedl.org/orc/rr/secc_rr_00094.pdf).
- Standards for Teacher Competence in Educational Assessment of Students. (1990), pp. 2-3. Retrieved from <http://www.eric.ed.gov/PDFS/ED323186.pdf>.
- Stanford University. (1994). School tracking harms millions, sociologist finds [Press release]. Retrieved from <http://news.stanford.edu/pr/94/940302Arc4396.html>
- Starch, D., & Elliott, E. C. (1912). Reliability of the grading of high-school work in English. *School Review*, 20, 442-457.
- Stiggins, R. (2007). Assessments through the student's eyes. *Educational Leadership*, 64(8), 22-26.
- Stiggins, R. J., & Conklin, N. F. (1992). *In teachers' hands: Investigating the practices of classroom assessment*. Albany, NY: State University of New York Press.
- Stiggins, R. J., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, 8(2), 5-14.
- Tomlinson, C. A. (2001). Grading for success. *Educational Leadership*, 58(6), 12-15.
- Tomlinson, C. A. and McTighe, J. (2006). *Integrating differentiated instruction & understanding by design: Connecting content and kids*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Truog, A. L. & Friedman, S. J. (1996). *Evaluating high school teachers' written grading policies from a measurement perspective*. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

TwentySided. (2011, September 30). Autobiography Part 21: Homework [Web log post.]

Retrieved from <http://www.shamusyoung.com/twentysidedtale/?p=13351>.

U.S. Department of Education, Office of Educational Research and Improvement (1988).

*National Education Longitudinal Study of 1988*. Retrieved from

<http://nces.ed.gov/surveys/nels88/>.

Wiggins, G. & McTighe, J. (2005). *Understanding by design* (2<sup>nd</sup> ed.). Alexandria, VA: Association for Supervision and Curriculum Development.

Wiggins, G. & McTighe, J. (2008). Put understanding first. *Educational Leadership*, 65(8), 36-41.

Willingham, W. W., Pollack, J. M., and Lewis, C. (2002). Grades and test scores:

Accounting for observed differences. *Journal of Educational Measurement*, 39(1), 1-37.

Wormelli, R. (2006). Turning zeroes to 60s. *Teaching in the Middle*, 9(3), 12-15.