

THE IMPACT OF TEACHER PERCEPTION OF STUDENT EFFORT AND
STUDENT ATTITUDE ON GRADING OF SECONDARY STUDENTS

A Dissertation
Presented to
the School of Education
Drake University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

by Jeffery P. Hawkins
October 2010

THE IMPACT OF TEACHER PERCEPTION OF STUDENT EFFORT AND
STUDENT ATTITUDE ON GRADING OF SECONDARY STUDENTS

by Jeffery P. Hawkins
October 2010

Approved by Committee:

David Darnell, PhD. chair

Elaine Smith Bright, EdD.

Jennifer Lindaman, EdD.

Janet M. McMahonill, PhD.
Dean of the School of Education

THE IMPACT OF TEACHER PERCEPTION OF STUDENT EFFORT AND STUDENT ATTITUDE ON GRADING OF SECONDARY STUDENTS

An abstract of a Dissertation by
Jeffery P. Hawkins
October 2010
Drake University
Advisor: David Darnell

The problem: Research has repeatedly noted the inclusion of nonacademic factors complicate the ability to interpret grades since these factors may directly conflict with each other and distort the meaning of a grade.

Procedure: This study investigated the impact two nonacademic factors, teacher perception of student effort and attitude toward subject matter, had on the final grade students earned in Algebra I. This study examined 851 students from 15 teachers' classrooms in and around the Des Moines, Iowa, metropolitan area.

Findings: Results of data analysis established a highly significant correlation between teacher perceptions of effort and grades in Algebra I ($r = .6267$) with a coefficient of determination of .3927. Likewise, a highly significant correlation ($r = .5743$) was established between perceptions of attitude toward subject matter and the final grade in Algebra, with a coefficient of determination of .3298. These findings supported previous research that nonacademic factors influence grading.

Recommendations: The findings have implications for pre service and in-service teacher training and professional development around grading and measurement, as well as for the reporting tools schools use to report grade

Table of Contents

	Page
List of Tables.....	vi
List of Figures.....	vii
Chapter	
1. INTRODUCTION.....	1
Statement of the Problem.....	1
Grades as Communication Tool.....	1
How Grades are Used.....	2
Grades as Tools of Miscommunication.....	5
Validity in Grading.....	7
Purpose of the Study.....	13
Research Questions.....	14
Significance of the Study.....	14
Definition of Terms.....	15
2. REVIEW OF THE LITERATURE.....	17
History of Grading.....	18
Various Meaning of Grades.....	19
Validity in Grading.....	21
Influence of Non-academic Factors on Grading.....	22
Lack of Training in Measurement.....	25
Impact of Invalid Grading Practices on Students.....	28
Summary.....	29

3. METHOD.....	31
Sample.....	31
Research Design.....	32
Materials.....	34
Process	35
Statistical Procedures	36
Limitations	37
4. ANALYSIS OF THE DATA	39
Descriptive Statistics.....	40
Research Questions	44
Summary	53
5. SUMMARY, CONCLUSIONS, DISCUSSION, AND RECOMMENDATIONS	55
Discussion	55
Review of Findings	58
Suggestions for Future Research.....	62
Summary and Conclusions.....	64
Bibliography.....	68
Appendices	
A. Survey Directions.....	75
B. Survey Instrument	76
C. Informed Consent Form	77

Tables

Table	Page
1. Descriptive Statistics for Dependent and Independent Variables	41
2. Descriptive Statistics for Teacher Perception of Effort for the Grade Distribution	42
3. Mean, Standard Error, Standard Deviation, Variance, and Count for Teacher Perception of Student Attitude Toward Subject Matter for the Grade Distribution	44
4. Single Factor ANOVA for Statistical Significance of Difference of Means of Teacher Perception of Student Effort.....	45
5. Scheffé Test for Statistical Significance between Means for the Grade Distribution for Effort.....	47
6. Single Factor ANOVA for Statistical Significance of Difference of Means of Teacher Perception of Student Attitude Toward Subject Matter	48
7. Scheffé Test for Statistical Significance between Means for the Grade Distribution for Attitude	49
8. Spearman <i>rho</i> for Correlation and Coefficient of Determination for Effort	51
9. Spearman <i>rho</i> for Correlation and Coefficient of Determination for Attitude.....	53

Figures

Figure	Page
1. Frequency Distribution of Grades	41
2. Means of Effort to Grades Line Graph	43
3. Means of Attitude to Grades Line Graph.....	44

Chapter 1

INTRODUCTION

Statement of the Problem

The act of assigning letter grades to represent a student's academic achievement in a classroom dates back to 1883 when a professor at Harvard marked a student's achievement as a "B" (Cuerton, 1971). Prior to this, measurement of a student's progress was through narratives written about the competencies of the student and their relative areas of weakness (Cuerton, 1971). Since this first B, the act of assigning grades has been surrounded in controversy (Cuerton, 1971; Ebel & Frisbie, 1986; Hopkins, Stanley, & Hopkins, 1990).

Grades as Communication Tools

To best understand this controversy, one must first recognize the fundamental purpose of grading, as well as its importance in determining future opportunities for those who receive grades. The primary purpose of assigning a student a grade is to communicate information about what a student knows and is able to do (Brookhart, 1993; Barnes, 1985; Marzano, 2000; Allen, 2005; Zoeckler, 2007). At this time, two primary voices are present in the literature about grading: those of measurement experts who study how to measure and report academic achievement, and those of practitioners who assign grades to students. And while these groups often differ on many aspects of grading, they agree that communication is a central purpose of grading. Also undisputed is the importance of grades. Allen (2005) stipulates that "assigning grades is probably the most important decision that classroom teachers make" (p. 221) because grades can open

up or close down important learning opportunities for students (Jasmine, 1999). Grades help to communicate information about students' performance. This directly impacts future learning opportunities for students. Others have suggested grades carry importance in other ways. If students receive grades lower than ones accurately depicting their true level of academic knowledge, it may lead them to believe they lack the ability to succeed academically and lower their sense of self-efficacy, as well as their motivation to learn (Pintrich & Schunk, 2002). However, beyond these fundamental premises of the importance of grades and that grades are used to communicate, much disparity exists.

What is communicated, and for whom this communication is directed, are often sources of murkiness. Are the grades students receive meant to speak to the knowledge and skills possessed, the effort put into work, the ability to be timely and organized, the motivation towards that subject area, behavior and attendance, or innate ability in that subject matter? Are the grades on a transcript meant to speak to the student who received the grade, the parent of that student, future teachers of that student, future employers of that student, or future educational institutions that may or may not allow that student to pursue further education? Unfortunately, the answer to these questions, as ascertained by an extensive review of the literature, is yes; yes to all of the questions, depending upon who is asking. Furthermore, there is a distinct difference in how these questions are answered between measurement experts and practitioners.

How Grades are Used

Complicating the issue of grades and the communication of those grades are competing philosophies about how grades should be used. "There is widespread agreement amongst measurement specialists that grades, at least in academic subjects,

should be based exclusively on measures of current achievement and that growth, ability, effort, conduct, and other non achievement factors should not be considered” (Cross & Frary, 1996, p. 53). Some measurement specialists posit that grades be used as academic achievement measures upon which valid determination of educational progress and career choices can be made. Other researchers, also measurement experts, have gone further to suggest that any other use of grades than that of a purely objective measurement of achievement makes the use grades invalid, unreliable, and difficult, if not impossible, to interpret. Allen (2005) summarized this thinking when he wrote:

If a teacher must summarize and communicate a student’s classroom progress in an academic area through a single report card grade, then there must be a consensus that the grade represents the most accurate statement of the student’s academic achievement, and only academic achievement. To include non-academic criteria, such as the student’s effort, compliance, attitude, or behavior, makes the grade impossible to interpret in any meaningful way. (pp. 221-222)

Cross and Frary (1996) also pointed to this philosophy of grading as exclusively the measurement of academic achievement when they stated:

Because of the importance placed on academic grades at the secondary level, either for educational or occupational decisions, grades should communicate as objectively as possible the levels of educational attainment in the subject. To encourage anything

less is to distort the meaning of grades as measures of academic achievement. (p. 56)

However, recommendations of measurement specialists have had little impact on the philosophies of teachers. “The measurement community has had little effect on grading practices” (Cross & Frary, 1996, p. 54). Teachers express their use of grades for the purpose of documenting and communicating achievement, but also to reward positive or compliant behavior (Frary, Cross, & Weber, 1993; Brookhart, 1994). Teachers also report grades are used as a motivator for students to put forth more effort (Karmel, 1970; Glasser, 1971; Hargis, 1990; Brookhart, 1994) or promote good study habits (Oosterhof, 2001). Others have concluded that not only do teachers find it acceptable and appropriate to ignore the suggestions of the measurement community (Cross & Frary, 1996; Frary et al., 1993; Allen, 2005), but they see those philosophies as incompatible with the realities of the classroom. Airasian and Jones (1993) noted this disparity when they concluded:

It is difficult for most teachers to separate their knowledge and perceptions of students from their grading judgment, so most do not. Context independent, best measurement practices...are generally dismissed by teachers as unrealistic, impractical, or more bluntly, not relevant to classroom needs (p. 241).

Beyond using grades as negative consequences for students, researchers have also documented teachers’ use of non-academic factors in determining grades to bolster student confidence (Friedman & Frisbie, 1995), or to mitigate the social complexities of failure. Brookhart (1991) found teachers often include non-academic factors for social reasons, understanding the implications of failure for their students and themselves.

Brookhart (1991) further suggested teachers may use non-academic factors in determining grades to mitigate consequences associated with inappropriate use of grades. This suggests some teachers are aware of the invalidating effects of including non-academic factors into grades, but choose to continue the practice of including non-academic factors in their grading to counteract their already invalid grading practices. Rather, teachers should simply base their measurement on academic achievement (Marzano, 2000; Allen, 2005; O'Connor, 2009).

The various uses of grades, beyond a simple reporting of academic achievement, has led measurement experts to conclude, “it would appear grades are more often measures of how well a student lives up to the teacher’s expectation of what a good student is than measuring the student’s academic achievement in the subject matter objectives” (Allen, 2005, p. 220). Others have concluded, “current grading practices don’t do the one thing they are meant to do, which is to provide an accurate indication of student achievement” (Clymer & Wiliam, 2007, p. 36).

Grades as Tools of Miscommunication

With such disparity in practitioner use of grades and such a variety of purposes, it is reasonable to conclude that not all who assign grades include similar factors when determining grades. Also reasonable is the conclusion that dissimilar processes are used in the calculation of grades. Kain (1996) found even within the same school, teachers often held very different views about the purpose of grades and failed to communicate with their colleagues about their grading practices.

If the variances between recommendations of measurement specialists and real world application of practitioners is so great so as to have caused disagreement and

confusion among professionals, imagine the muddled communication that must come to those who attempt to interpret a grade. One might think it impossible to look at any collection of grades and make a meaningful or valid interpretation about the student associated with those grades. When interviewed, teachers reported they rarely give credence to the prior grades of students entering their classes, even in the same subject area, due to the wide variability of how grades are determined (Waltman & Frisbie, 1994). Brookhart (1993) added: (check apa format)

Grades are simply another form of communication. And as such, it is the receiver of the information that graders must be concerned with. When a teacher assigns a grade, there are multiple parties who rely upon that grade to infer meaning about that student and their ability to perform in that academic subject area. Some may interpret a low grade to mean little effort was put forth, others that the student struggles with that academic content area. Still others may infer that a low grade represents a poor attitude about schooling or a refusal to be compliant. The same inferences, accurate or inaccurate, can be made from high grades. It is because of a history, and a general acceptance of the inclusion of various non-academic factors into grades that leads to this problem of miscommunication. (p. 285)

While both teachers and parents believe grades are very important, that may be as close as the two groups come together in their common understanding of student grades (Waltman & Frisbie, 1994). Researchers have concluded while teachers preferred to

include effort in grades and to assign higher grades to students who worked hard in and out of class, parents only knew how to interpret grades as achievement measures (Brookhart, 1994). Waltman and Frisbie (1994) found that even when parents expressed they understood and agreed grades often serve a dual purpose, they typically interpreted the grade assigned to be communication about academic achievement. Marzano (2000) summarized this hodgepodge approach to grading and its inherently blurred communication when he wrote, “in contrast to teachers’, students’, parents’, and community members’ assumption that grades are valid measures of student achievement...grades are so imprecise that they are almost meaningless” (p.1).

Validity in Grading

Validity is a complex, multifaceted measurement concept that contains many constructs. For the purpose of relating this concept to assigning and interpreting grades, one must understand the importance of internal validity. Internal validity is defined as indicating “the relationship between two or more variables should be unambiguous as to what it means rather than being due to ‘something else’” (Fraenkel & Wallen, 2006, p. 169). As this applies to grading, the variables to be considered are the letter grade and the knowledge and skills of the student in a subject matter. Grades given with high internal validity would have a strong correlation between the grade and the knowledge and skills of that student. Thus, the reason students would earn an A in a class would be because they evidenced superior knowledge and skills in that content area, rather than because they turned homework and projects in on time, put forth outstanding effort, and attended class regularly. The inclusion of non-academic factors such as organization,

effort, and attendance threaten the internal validity of grades and increase the likelihood the grade assigned for a class is more a result of something else.

The importance of valid grades cannot be overstated. The most fundamental measurement principle related to meaningful assessment and grading is the principle of validity (Gallagher, 1998; Gredler, 1999; Linn & Gronlund, 2000; Stiggins, 2001). Having grades reflect what students know and are able to do is critical in the proper development of their educational experience, as well as in the opportunities that may or may not be made available to them based upon those grades. Allen (2005) wrote, “Invalid grades that understate the student’s knowledge may prevent a student with ability to pursue certain educational or career opportunities” (p. 220).

The impact of invalid grading practices is something that has permeated all levels of the educational system, from elementary school through post graduate programming (Brookhart, 1993). Due to the wide variability in the criteria used in grading practices from teacher to teacher, the validity of student grades is unknown and thus grades have limited value as guides for planning the academic and career futures of students (Thorndike, 1997). This assertion, made by made researchers of grading, has been founded and reinforced time and time again throughout a century of research into the measurement of students. Researchers have found that grading systems used by teachers vary widely and unpredictably and often have low levels of validity due to the inclusion of non-academic criteria in the calculation of grades (Allen & Lambating, 2001; Brookhart, 1994, 2004; Frary et al., 1993; Olson 1989). Researchers have repeatedly noted the inclusion of non-academic factors complicate the ability to interpret grades since these factors may directly conflict with each other and distort the meaning of a

grade (Cross & Frary, 1996; Gusky, 1996; Linn & Gronlund, 2000; Nitko 2001; Stiggins, 2001; Stumpo 1997).

Understanding that determining, assigning, and reporting grades with internal validity is of critical importance to all involved in education, one may ask, how this is accomplished. The answer provided by measurement specialists is to simplify the process. Rather than attempting to communicate the academic achievement of students, as well as their internal motivation, aptitude for learning, alignment with the social mores of the institution or their subject matter teacher, their ability to persevere, and the expectations of parents and the community in a single letter grade, measurement experts suggest one, commonly agreed upon and understood message must be the intent of a letter grade. Since a single letter or numeric mark is commonly used to report grades, a single fact about the student must be the intended communication if the communication is to be valid or accurate. Thus, grades can not be a teacher's merged judgment of various factors (Allen, 2005).

This concept of simplicity, of reporting one fact with one reporting mark, seems so obvious that it begs the question, why the controversy? Why would teachers, knowing the potential consequences of reporting invalid grades, use a grading practice likely to invalidate their measurements of knowledge and skills? How is it educators have not simply worked hand-in-hand with educational measurement specialists to design systems to ensure the validity of the grades reported? The answers to these questions are anything but simple, and go far beyond the notion that teachers are either unable or unwilling to grade with validity.

While there are myriad reasons practitioners may not follow the suggestions of measurement experts, including the belief these suggestions are impractical and unrealistic for use in the classroom (Airasian & Jones, 1993), a review of the research suggests that at least three reasons exist for the disparity between the groups. Researchers have suggested that teachers' values and sense of right and wrong, response to expectations of others, and lack of knowledge and training most teachers have in using grading practices are all primary reasons why invalid grading practices are commonly used.

No other factor may influence grading more heavily than the individual teacher's personal values and sense of right and wrong. Zoeckler (2007) wrote, "teachers' ideas of right and wrong, of good and bad undoubtedly figure, perhaps significantly, in the evaluation of student work and in the grades such work receives" (p. 98). Buzzelli and Johnson (2002) concluded grading is heavily influenced by the values and beliefs of the teacher who grades. If a teacher values organization, turning work in on time, participating in class, putting forth what is perceived as good effort, or being compliant or easy to work with, those students who exhibit these characteristics will likely earn higher grades than those students who fail to reach those expectations. Researchers have found many teachers believe effort, student conduct, and attitude should influence final grades of students (Cross & Frary 1996; Frary et al., 1993).

If a teacher refuses to accept late work, or takes off points as a result of when work is turned in, rather than grading the work for its evidence of learning, then that teacher's values of organization and timeliness had outweighed the value of assessing learning. If a teacher grades homework based upon its completion, rather than scoring it

for evidence of learning, then that teacher's value of effort has outweighed the value of assessing learning. If a teacher awards points for those who speak in class and work well with groups, regardless of the value of their contributions or the subsequent reflection of learned knowledge and skills, then that teacher's value of participation has outweighed the value of assessing learning. If a teacher decides to adjust the grade of a student because that student has worked hard and is a good kid, then that teacher's values of effort and compliance have outweighed the value of assessing learning. The applications of teachers' personal values within their system of grading combine with the failure of the educational system to provide oversight or training in valid grading practices to create situations in which grades reported are invalid in their communication about knowledge and skills learned. Brookhart (1993) found non-academic factors are often used as criteria for assigning grades because some teachers consider the consequences of grades more important than the value of clear communication of information and the interpretability of the grades.

A second reason invalid grading practices are commonly used is teachers are simply responding to the expectations of others. "Teachers grade the way they do because they are responding to the expectations of parents, students, and their job as teachers" (Trough & Friedman, 1996, p.1). Teachers understand the grades they assign communicate information about each student, and these grades will influence the opportunities those students have after schooling. Thus, teachers are aware of and are influenced by the grading expectations of students, parents, the school, and the community. Furthermore, teachers understand the societal impact of student failure and dropout, and are therefore influenced by societal expectations. Noddings (2002) found

the following influences were all part and parcel of the everyday decision making of teachers:

- consideration for the common good, along with that of the individual teacher's desire to both encourage effort and reward achievement while promoting improvement and further development
- accountability to the larger community, the school organization, and parents' wishes
- the satisfaction of one's own conscience

Finally, teachers' widespread use of invalid grading practices can be closely linked to the lack of pre service training on valid grading practices. Lack of knowledge about measurement theory and application to grading practices is a pervasive problem with pre service teacher training at the college level (Goodwin, 2001; Schafer, 1991; Stiggins, 1991, 1999). Allen and Lambating (2001) found in a random sample of teacher education programs, less than one third required an assessment course, and many of those that did were not focused on classroom assessment and grading. Stiggins (1999) also noted concepts of reliability and validity related to classroom grading practices are not addressed in the courses which introduce these terms to pre service teachers.

Most often teachers are brought into a system that mandates their use of a single letter grade to report student learning in a content area without any pre service training on how to accurately measure this learning. Thus, teachers often adopt grading systems from colleagues or adapt grading systems based upon how they have been graded in the past. Allen (2005) writes, "Because grading is something that has been done to each of us

during our many years as students, it is hard to change the invalid grading schema that has become embedded in our minds” (p. 218).

Purpose of the Study

Despite longstanding concerns about the role of validity within grades and numerous attempts to help teachers understand the purpose and effective functions of grades within the overall evaluation system, there seems to be little progress in actual reform of classroom practice (Allen, 2005). Researchers and measurement experts have concluded for decades that grading practices used in classrooms do not validly reflect what a student knows and is able to do. Additionally, these grading practices result in poor, if not potentially damaging, communication tools due to the lack of clarity in what is communicated. Likewise, many well-respected researchers have called for broad scale change in grading practices in light of findings that non-academic factors are commonly included in grading. Unfortunately, measurement experts’ conclusions of invalidity and researchers’ calls for change have had little influence in the day-to-day work of practitioners.

The purpose of this study is to quantify the impact two non-academic factors (teacher perception of student attitude towards the subject matter, and teacher perception of student effort) have on the final grade assigned to secondary students. Each of these non-academic factors has been researched and has been deemed to be commonly included in teachers’ grades, and to complicate the meaning of the grade (Cross and Frary, 1996; Gusky, 1994; Linn & Gronlund, 2000; Nitko, 2001; Stiggins, 2001; Stumpo, 1997; Allen, 2005; Zoeckler, 2007).

Research Questions

- Are the differences in mean scores for teachers' perception of student effort statistically significant for students among grade distributions (As, Bs, Cs, Ds, and Fs)?
- Are the differences in mean scores for teachers' perception of student attitude toward subject matter statistically significant for students among grade distributions (As, Bs, Cs, Ds, and Fs)?
- What is the relationship between the final semester grade a student earns in Algebra I and the non-academic factor of student effort as perceived by the teacher?
- What proportion of variance within the final semester grade of a student in Algebra I can be attributed to teacher perception of student effort?
- What is the relationship between the final semester grade a student earns in Algebra I and the non-academic factor of student attitude toward the subject matter as perceived by the teacher?
- What proportion of variance within the final semester grade of a student in Algebra I can be attributed to teacher perception of student attitude toward the subject matter?

Significance of the Study

Because the grades students earn play an important role in determining the opportunities students have to pursue career and educational options, it is of critical

importance that grades on a student's transcript reflect accurately what the student knows and is able to do. To allow other factors to muddy the communication between those issuing grades and those interpreting those marks is inappropriate.

Despite long-standing concern over issues of valid grading practices, especially the inclusion of non-academic factors that invalidate grading processes, little change has occurred in teacher practice around grading. If relationships between non-academic factors can be quantified and the percentages of variance calculated attributed to those non-academic factors, and if those relationships show a significant level of variance (above .20), it may give measurement experts, teachers, and school leaders a way to emphasize the importance of validity in grading practices beyond the suggestions of measurement experts that have often been ignored.

Definition of Terms

Analysis of Variance (ANOVA): A statistical technique for determining the statistical significance of the differences among means; it can be used with two or more groups (Fraenkel & Wallen, 2006, p. G-1).

Coefficient of determination (r^2): The square of the correlation coefficient. It indicates the proportion of variance common to two variables (Fraenkel & Wallen, 2006, p. G-1).

Correlation Coefficient (r): a decimal number between .00 and -1.00 or 1.00 that indicates the degree to which two quantitative variables are related (Fraenkel & Wallen, 2006, p. G-2).

Criterion referenced: For the purposes of this study, this refers to the manner of determining student grades by comparing students' achievement to a set of standards for learning (Waltman & Frisbie, 1994).

Non-academic factor: any factor that is included in or influences the grade a student earns for a course that is not a measurement of students' knowledge or ability to meet the standards and benchmarks within a given course. Researched non-academic factors include: attendance, participation, behavior, compliance, organization, effort, attitude, motivation, and timeliness of completion of work.

Norm referenced: For the purposes of this study, this refers to the manner of determining student grades by comparing students' achievement to that of other students (Waltman & Frisbie, 1994).

Student attitude: the extent to which a student is positive about classroom assignments and shows and enjoyment of the curriculum (as perceived by the teacher).

Student effort: the extent to which a student works to meet course objectives and teacher expectations (as perceived by the teacher).

Chapter 2

REVIEW OF THE LITERATURE

Grading has been a controversial issue since the turn of the century (Cuerton, 1971; Ebel & Frisbie, 1986; Hopkins et al., 1990). As early as 1913, educators questioned the general acceptance of a letter grade given as a valid or precise tool in measuring student knowledge.

When we consider the practically universal use in all educational institutions of a system of marks, whether numbers or letters, to indicate the scholastic attainment of the pupils or students in these institutions, and when we remember how very great stress is laid by teachers and pupils alike upon these marks as real measures or indicators of attainment, we can be but astonished at the blind faith that has been felt in the reliability of the marking system. School administrators have been using with confidence an absolutely uncalibrated instrument...What faults appear in the marking systems that we are now using, and how can these be avoided or minimized? (Finkelstein, 1913, p.7)

Nearly 100 years after the concern about validity of grades was raised by Finkelstein, little change in practice has occurred to help ensure that a mark a student earns accurately reflects what that student knows and is able to do (Allen, 2005). To best understand the issues surrounding grading validity, one must also understand where grades came from and the complex and varied meanings communicated through grades.

History of Grading

The first documented attempts to record student achievement in a marking system come from early American universities (Smallwood, 1935). In 1877, at Harvard University, faculty classified students based upon their rank order on a 100 percent scale (Smallwood, 1935). Six divisions separated students from 100 percent to below 40 percent. It was not until 1883 that the first letter grade was used at Harvard (Cuerton, 1971; Smallwood, 1935) when a student was marked as having earned a “B”. The following year Harvard faculty, according to the annual report of the university president, “did away with the minute percentage system of marking, and substituted a classification of the students in each course of study in five groups, the lowest of which indicates those who have failed in the course” (Smallwood, 1935, p. 31). Other universities also worked on marking systems and classifications of students. However, in 1897, at Mount Holyoke, a marking system was introduced that became the cornerstone for school grading systems (Smallwood, 1935). The faculty at Mount Holyoke adopted the following marking and classification system: A: 95-100, B: 90-94, C: 85-89, D: 80-84, E: 75-79, F: Failed (Smallwood, 1935). By combining letter grades with percentages and including the descriptive adjective of failed, the university created a system for grading that has served as a foundation for marking systems for well over a century.

This trial and error method of comprising a grading system meant to differentiate students begins the conversation about validity. Were the intervals between letter grades and percentages used in establishing this system in 1897 equal? Did the intervals chosen in 1897, which still influence grading scales today, represent a normal distribution of

scores? If yes, has that normal distribution of scores remained consistent for over 100 years? The arbitrary nature of the letter grades themselves and the arbitrary percentages to which they are associated create a system seemingly based upon educators' best guess.

Various Meaning of Grades

Because grades are used by and for many people, the meaning of a grade varies according to the purpose for which it is being used and by who is using it (Brookhart, 1994). Four questions identified by Messick (1989) help to describe the varied meaning grades can have for teachers, students, parents, and others attempting to interpret them.

- What does the grade mean per se? – what factors constitute the make-up of a grade
- What does the grade mean when assigned to a student? – what meaning will be interpreted about a student by those who view the grade
- What does the grade mean when assigned to a student, and of what (social) value is it? – what does this grade mean in the context of how others are graded (fairness); and
- What does the grade mean when assigned to a student...and what will happen because of it? – what are the social consequences of how students' grades will be interpreted (p. 6)?

Waltman and Frisbie (1994) illustrated the multifaceted nature of grades and the varied potential meanings and interpretations of grades as follows.

One conceptualization of a grade symbol is that it is comprised of three separate facets. First, a grade compares or references a student's performance to either a relative standard (the performance of other

students) or an absolute standard (the teacher's explicit performance criteria). Second, a grade describes growth in academic achievement over time, or it indicates the status of an achievement at a certain point in time (e.g., the end of the quarter). Third, a grade focuses only on academic achievement in school subjects, or it also reflects such non academic characteristics as effort, disposition, deportment, or neatness. Because the different possible combinations of these three facets allow grading practices to take on a variety of meanings (some of which are illogical), the multifaceted nature of a grade symbol potentially contributes to miscommunication. If the grades of some students in a class are based upon comparisons with classmates, and the grades of others are based upon comparison with the teacher's standards, then the meaning of a given student's grade can only be guessed (p. 224).

With this variety of possible meanings for assigned grades, it is no wonder that teachers express discomfort in the act of assigning grades to students (Lomax, 1996; Thorndike, 1997). Barnes (1985) found teachers viewed grading and evaluation as the most difficult part of their work. Assigning grades is such a complex and sometimes controversial issue that some have proposed their abolition (Kohn 1999; Marzano, 2000).

That being said, measurement experts agree that what is communicated through a grade should be a precise accounting of students' academic achievement (Brookhart, 1994; Allen, 2005; Baily & McTighe, 1996; Stiggins, 1999). Bailey and McTighe (1996) stated, "the primary purpose of secondary grades and reports is to communicate student achievement so that informed decisions can be made about the student's future (p. 120).

Validity in Grading

In order to have a precise accounting of academic achievement interpreted by those viewing the grade, what is considered when assigning a grade must have uniformity from teacher to teacher, subject to subject, and school to school. “The meaning of a score and the values implied in interpreting and using that score are intertwined in the concept of validity” (Brookhart, 1993, p. 123). It is through the inclusion of other considerations, these non-academic factors, that the meaning of the grade itself (Messick’s first question) is distorted and invalidated. Researchers agree, although ancillary information such as effort and attitude could be part of an overall student report, they should not be part of a grade that represents academic achievement (Tombari & Borich, 1999). Allen (2005) wrote, “since factors such a effort, motivation, and student attitude are subjective measures made by the teacher, their inclusion in a grade related to academic achievement increases the chance for the grade to be biased or unreliable, and thus invalid” (p. 220). Despite the varied meanings of grades and the inherent invalidity in the calculation of grades, parents report that they interpret grades as marks of academic achievement (Allen, 2005).

Waltman and Frisbie (1994) concluded the information between what a teacher intends to communicate through a grade and what parents interpret though grades is jumbled at best. Often teachers are not in agreement among themselves as to how to interpret grades from each other. Teachers state they rarely use the grades assigned to students by other teachers as information they can rely upon due to low levels of validity between teachers’ grading practices (Waltman & Frisbie, 1994).

Influence of Non-academic Factors on Grading

Measurement experts have for decades urged educators to understand and recognize the importance of the measurement principle of validity when issuing student grades (Allen, 2005). The validity of a grade is considered by measurement experts to be the most important concept when trying to ascribe meaning or interpretability to the grade (Gallagher, 1998; Gredler, 1999; Linn & Gronlund, 2000; Stiggins, 2001).

Researchers have noted grading practices by teachers rarely follow measurement principles or the grading practices recommended in measurement textbooks (Cross & Frary 1996; Frary et al., 1993). Griswold (1993) wrote “teachers indicate that achievement is important in determining student grades, but they use other indicators as well, even though they may know this to be contrary to the recommendations of measurement experts” (p. 311).

While it is recognized teachers do think about grading systems seriously and work to find ways to assign grades fairly, the lack of congruence between recommendations of measurement specialists and classroom teachers’ grading practices is disconcerting (Brookhart, 1994). The work of researchers and the recommendations of measurement experts have little effect on the grading practices of teachers (Allen, 2005). Brookhart (1993) stated, “measurement instruction can be expected to clarify teachers’ concepts of the meaning of grades, but there is no reason to expect that measurement instruction will change thinking about the values and social consequences [of grading].” (p. 140).

Brookhart (1993) also found due to most teachers’ child-centered orientations and desires to be an advocate for their children, teachers’ concerns about consequences of grading

may be expected to have more influence on grading practices than concern about interpretability.

Reasons highlighted in research literature that teachers do not align their grading practices with the recommendations of measurement are the use of grades to control behavior (Frary et al., 1993; Dockery, 1995), the use of grades as motivational tools (Brookhart, 2004; Glasser, 1971; Hargis, 1990; Karmel, 1970), or the use of grades to develop good study habits (Oosterhof, 2001). Research has concluded teachers see grading as more than reporting what a student knows and can do. Waltman and Frisbie (1994) found 50% of teachers surveyed indicated their grades reflected both achievement and non achievement factors. Brookhart (1993) found teachers believed the meaning of grades to be closely related to the idea of student work. “Grades are pay that students earn for activities they perform” (Brookhart, 1993, p. 123).

When presented with scenarios describing students as having either high effort and low aptitude or low effort and high aptitude, teachers at the secondary level indicated grading was influenced significantly by perception of effort (Griswold, 1993). Teachers surveyed were 53% more likely to give a higher grade to a student who was perceived to be showing high effort despite low aptitude, while 76% of teachers in the same study indicated they would give a lower grade to a student who showed low effort despite having a higher aptitude (Griswold, 1993). This finding was confirmed by Griswold (1993) when he compared the grading practices of pre service teachers with those of current practitioners. While 40% of secondary teachers indicated they would give a higher grade to a student who had high effort and low aptitude (compared with 61% of pre service teachers), 88% of the same group indicated they would give a lower grade to

those students who had low effort and high aptitude (as compared to 68% of pre service teachers). Both of these examples indicate the potential that secondary teachers may use grades to punish students for perceptions of low effort (especially when the corresponding perception for a student is they possess high aptitude) more frequently than they may use grades to reward perceptions of high student effort (Griswold, 1993). In fact, Griswold (1993) found teachers believed, “poor work habits should be punished by using a lower grade, even in light of academic achievement. Students must work hard for their grades” (p. 324). Griswold (1993) quoted a teacher interviewed for the study as saying, “It’s our job to reinforce daily habits that are needed on the job. If they are not going to work hard, how can they expect charity from me” (p. 324).

According to a 1999 NCES study, the majority of teachers said they based grades partly on absolute achievement (76.1%), but higher percentages of teachers indicated student effort (96.6%) and individual improvement (83.9%) factored into their grading decisions. Waltman and Frisbie (1994) found similar results as 88% of teachers surveyed said achievement should influence the grades of students. However, 94% of teachers believed effort should influence grading and 88% of teachers believed work completion should influence grading (Waltman and Frisbie, 1994). When asked about what factors influenced grades for final semester grades, Cizek, Fitzgerald, and Rancor (1995) found 89% of teachers surveyed considered achievement when determining grades. Cizek and Fitzgerald (1995) also found 52% of teachers considered attendance or participation in class and 61% of teachers considered effort and conduct when determining final grades for a course.

The prevalence of the inclusion of these subjective, non-academic factors led Zoeckler (2007) to conclude, “clearly, grading is heavily laden with moral considerations that go well beyond the requirements of achievement testing” (p. 84).

Lack of Training in Measurement

A final reason teachers do not align grading practices with measurement expert recommendations is a lack of knowledge about measurement concepts fundamental to sound grading practices. Fewer than half of the fifty states require specific coursework on assessment for initial teacher certification (Lomax, 1996; O’Sullivan & Chalnack, 1991; Stiggins, 1999). As a result of the pervasive lack of training pre service teachers receive on measurement theory (Stiggins, 1999) and the lack of ongoing training current practitioners receive on application of fundamental measurement concepts to grading systems, (Allen, 2005) teachers have developed practices that contradict their own grading philosophies (Cross & Frary, 1996).

Cross and Frary (1996) documented that because there is a clear distinction between how grades are used and how they should be used in an ideal situation, teachers often do not have grading practices that align with their own philosophies. Thirty-six percent of the teachers surveyed who stated they do not consider ability when issuing student grades indicated, in an ideal situation, ability should be considered (Cross & Frary, 1996). Likewise, 49% of teachers who stated they do consider ability when issuing student grades indicated that, ideally, ability should not be considered (Cross & Frary, 1996). Students too were confused over the role ability should play in determining grades. While 59% stated they felt ability should not be considered when determining grades, 55% stated they felt it was fair for their teachers to consider ability when

determining a student's grade (Cross & Frary, 1996). Waltman and Frisbie (1994) found that while 94% of teachers believed effort should be included as a factor that influences a student's grade, only 44% of teachers reported said they actually intended the grades to reflect a student's effort. Cross and Frary (1996) found that 25% of teachers indicated they raised students' grades "fairly often" as a result of high effort and that 39% of teachers take conduct and attitude into consideration when determining grades. In contrast to these findings, 81% of teachers in the same study reported their agreement with the assertion that attitude and effort should be excluded from consideration in determining assigned grades (Cross & Frary, 1996). The data support the thesis that while teachers may be clear about what factors they feel should influence grading, they do not know how to align their grading philosophies and practices. Waltman and Frisbie (1994) concluded teachers were unaware of how to align their grading philosophies with a plan for how to grade students.

Further research suggests teachers are unclear about some of the most fundamental measurement concepts related to grading. As noted earlier in this chapter, a valid grade can either be comparable to other students (Norm Referenced, NR) or can be comparable to a teacher or district standard (Criterion Referenced, CR), but not both. Waltman and Frisbie (1994) found 38% of teachers indicated they felt grades should be based both on comparison to peers (NR) and comparison to the standard (CR). Likewise, Cizek and Fitzgerald (1994) found 35% of teachers surveyed indicated final grades were reflective of students' achievement as compared to classroom objectives (CR) while 19% of teachers surveyed stated students' final grades were related to student achievement as compared to the performance of the class (NR). Nearly 30% of teachers surveyed stated

the students' final grades were reflective of both comparisons (Cizek & Fitzgerald, 1994). The final 16% stated students' final grades were reflective of student achievement compared to individual student goals (Cizek & Fitzgerald, 1994). In other words, 46% of the teachers surveyed by Cizek and Fitzgerald (1994) did not use a valid grading practice to allow for comparison of student results. These studies indicate a fundamental misunderstanding of basic principles of measurement. Because of the lack of exposure to measurement principles in pre service training, and because of a lack of discussion of grading philosophies and aligned grading practices in the profession, teachers are left to determine what factors they will consider when determining grades and what statistical philosophies will underpin their grading practices with little professional guidance. This lack of emphasis and discussion is not working in the best interest of students because too many teachers are not using grading practices founded upon viable measurement principles.

So long as pre service and in-service teachers take classes from education professors who base grading decisions on more than academic achievement, little change will occur with how teachers assign grades (Allen, 2005). Allen (2005) wrote:

Students in teacher education programs may be more influenced by the grading practices they have experienced as students in the past, as well as in their current courses taught by their education professors, than by what they learn about assessment and grading in their courses. (p. 221)

Impact of Invalid Grading Practices on Students

One can not quantify the detriment to students that has been imposed by the lack of response to measurement experts' suggestions for improved grading practices by teachers, school districts, and teacher preparatory institutions. Grades are seen as important by teachers, parents, and students (Waltman & Frisbie, 1994). Furthermore, often grades are used to determine which opportunities are available or unavailable to students (Allen, 2005). Allen (2005) stated, "Since important decisions are often based on a student's grade, invalid grades may result in dire consequences for the student" (p. 220). Grades influence educational plans (Rosenbaum, 2001), course-taking patterns in high school (Kelly, 2008), and eventual educational attainment (Rosenbaum, 2001). Grades in high school have been shown to affect earnings nine years after graduation (Rosenbaum, 2001). A student's early grades set the tone for his or her classroom experience and appear to have a causal effect on subsequent grades (Entwisle & Alexander, 1988).

Sinclair and Ghory (1987) found that because grades were seen by students as measures of merit, many youth felt marginal to the central school population partly because they were receiving messages in the form of failing grades that they did not belong in school. Cizek (1995) pointed out, "even as grades continue to be relied upon to communicate important information about academic performance and progress...they probably don't" (p. 104).

Concern has arisen around the grading practices used by teachers, the lack of validity such practices inherently bring to grading, and the miscommunication and contradictory in meaning surrounding grades for nearly a century. So too has concern

been expressed for nearly a century for how invalid grading will impact students.

Finkelstein (1913) wrote:

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...Nor may anyone seek refuge in the assertion that the marks of the students are of little real importance. The evidence is clear that the marks constitute a very real and a very strong inducement to work, that they are accepted as real and fairly exact measures of ability or of performance. Moreover, they not infrequently are the determiners of the student's career. (p. 19)

Summary

This chapter discussed the history of grading and how modern grading scales were founded upon arbitrary measures. The variety of potential meanings for grades and the various factors that influence grades were discussed, including research documenting teachers' inclusion of various non-academic factors in grades. The need for valid grading practices as suggested by measurement experts was illustrated and examples were shown of how educators often do not understand some of the most fundamental measurement principles required for valid measurement. The chapter then presented the argument that despite a vast body of research suggesting change in grading practice is needed, little change has occurred. Several research based reasons why teachers have not aligned their grading practices with the recommendations of measurement experts was documented.

Finally, the potential for invalid grading practices to negatively impact students was discussed and the century long-concern over this issue was documented.

Chapter 3

METHOD

This study investigated the relationship between the final semester grade a student earns in Algebra I and the non-academic factors of student effort and attitude as perceived by the teacher. The research was designed to determine if the differences in the mean of teacher perception of student effort and student attitude toward subject matter were statistically significant for each grade within a grade distribution. Finally, the research determine the proportion of variance in the final semester grade a student earned in Algebra I attributed to the non-academic factors of effort and attitude toward the subject matter.

Sample

For the purposes of this study, fifteen secondary Algebra I one teachers were selected for participation. Fifteen teachers (100%) completed the survey instrument. The study group consisted of 56.2% ($n=9$) female teachers and 43.8% ($n=6$) male teachers. When possible, teacher selection included one male and one female teacher from each selected school. Teachers in the study ranged from 3 years of experience to 38 years of experience, with a mean of teaching experience of 12.6 years. These teachers were a sample of Algebra I teachers from six school districts in and around the Des Moines, Iowa metropolitan area, including urban, suburban, and rural schools. Administrators in the chosen districts for the sample were asked to identify teachers with at least three years' teaching experience who taught at least two sections of Algebra I during the 2009-2010 school year.

Data were gathered from only those sections of Algebra I whose students were at the normal pace for taking Algebra I. Sections of students who were delayed in the math sequence for their district, or who had failed Algebra I previously and were retaking the class were excluded. Likewise, sections of students who were taking Algebra I at an advanced pace were excluded.

Upon being recommended by the school administration, teachers were contacted by the researcher to seek their cooperation in the study. The researcher met with each teacher agreeing to participate in the study, showed them the survey instrument to be used, and explained the data to be collected and the procedures to be used. During this meeting the researcher also presented, explained, and gathered signatures for consent to participate in the research.

Research Design

Because this study measured positive directional relationships between independent and dependent variables, a quantitative approach was appropriate. The research questions being studied lent themselves to be answered using correlational research. According to Fraenkel and Wallen (2006), correlational research is most appropriately used when the researcher wishes to describe the degree to which two or more quantitative variables are related.

Means were calculated for the data as a whole group, as well as for sub groups broken down by grade distributions (As, Bs, Cs, Ds, and Fs) for each dependent variable. An analysis of variance (ANOVA) was computed to determine if the differences in the means between the subgroups were statistically significant. A Scheffé test was then run

between each of the subgroups to determine exactly where statistical significance lay between the groups.

Because the data consisted of rank ordered data, Spearman *rho* correlations were used to determine the relationship between the dependent variable of the final semester grade a student earns in Algebra I with two different independent variables; teacher perception of student effort and teacher perception of student attitude towards subject matter. “Spearman rank correlation is used when one or both of the variables consist of ranks” (McDonald, 2009, p. 222). Because the data were rank-ordered, a scatter plot was not appropriate to show directional relationship of the variables. According to McDonald (2009) it would be misleading to put a linear regression line on a graph when it has been analyzed with rank correlation. He further stated that, “if you actually have true ranked data for both variables, you could plot a line through them” (McDonald, 2009, p. 222). Thus a line graph illustrating the linear relationship of the means of the groups was created, and compared to the line created when a typical linear regression was run. The results of this comparison are discussed in Chapter 4.

Correlation coefficients were calculated between each independent variable and grading to determine the strength of the relationship. The correlation coefficient for each independent variable was squared to establish the coefficient of determination. This allowed the researcher to be able to indicate the proportion of variance shared between each independent variable and grades awarded to students in the sample.

Materials

Once selected, teachers completed surveys to generate data points for the research. Teachers were given the survey instrument and directions, as well as copies of the research consent form on a flash drive and asked to complete the data electronically, so that data could be easily copied into Microsoft Excel for data analysis (see Appendices A and B, p. 82-83). The survey tool teachers used consisted of a four column spreadsheet. One column provided the teacher with an individual student code predetermined from a set of coded numbers generated by the researcher. This code allowed teachers to keep data organized for each student, while keeping individual student identity anonymous. The other three columns provided teachers a place to record data for the final semester grade for each student in one section of Algebra I, the teacher's perception of that student's effort in that class, and the teacher's perception of that student's attitude toward Algebra.

Teachers reported the final first semester grade for the fall semester of 2009 students earned in their Algebra I class. At the conclusion of the semester, they then ranked each selected student on a Likert scale in the areas of teacher perception of effort and teacher perception of attitude towards Algebra I. In teacher ranking of their perception of each student's effort, a ranking of 1 indicated perception of very little to no effort. A ranking of 2 indicated perception of below average (poor) effort. A ranking of 3 indicated perception of average effort. A ranking of 4 indicated perception of above average (good) effort. A ranking of 5 indicated perception of excellent effort.

Teachers then rated students on their perception of each student's attitude toward Algebra. A ranking of 1 indicated perception of an extreme dislike for Algebra. A

ranking of 2 indicated perception of a below average (poor) attitude toward Algebra. A ranking of 3 indicated perception of an average attitude toward Algebra. A ranking of 4 indicated perception of an above average (good) attitude toward Algebra. A ranking of 5 indicated perception of an excellent attitude toward Algebra. (Appendix B, p.77)

Process

During the fall of 2009 school districts were identified to participate in the study. Districts were chosen based upon size, with an attempt to find schools representing small, medium, and large schools. Districts were also chosen with the goal of including urban, suburban, and rural demographics. Upon gaining Institutional Review Board permission to conduct the study, the researcher contacted six Des Moines metro area school districts to seek their participation in the study. Applications for permission to conduct research within the districts were completed, and permission was granted from all six districts that were requested for research.

During the fall of 2009 building administrators in selected and confirmed districts were asked to identify two Algebra I teachers who met the study criteria. Individual teachers were contacted in December of 2009 and January of 2010. The researcher met individually with each prospective subject of the study. Teachers were given the documents to be completed for the study, had the data collection process explained to them, and had the consent forms described. Teachers were given a flash drive that contained electronic versions of the survey directions and the survey tool to be completed. By having identical and unmarked flash drives from which teachers could randomly draw from a bag, the districts and teachers were assured that the researcher did not know the individual student codes assigned for any given teacher or school. The

electronic versions of the survey were then emailed back to a research assistant who aggregated the data. This process ensured that:

- Each student was assigned an individual identification code to aid in aggregating data without worry that codes were duplicated
- Each student's identity was kept completely anonymous from the researcher
- Each teacher's identity was kept completely anonymous from the researcher
- The researcher did not know which district submitted which data

Teachers signed consent forms during the presentations about the research (see Appendix C, p. 84). Teachers were asked to have data completed and emailed to the research assistant by the end of January 2010.

Upon receiving the data from the teachers, the research assistant aggregated the data to an Excel file. This step prevented the researcher from being able to associate any particular set of data with any particular school or teacher. Because the researcher was concerned only with aggregate data, there was no need for the researcher to know which teacher or school reported which data.

Statistical Procedures

Basic descriptive statistics are beneficial in defining the nature of the population for the reader. The mean of each independent variable and the standard error of the mean give readers information about the population's baseline data. A frequency distribution of the dependent variable aids in determining whether a normal curve exists

within the data being examined. Descriptive statistics were calculated for each variable to aid in the interpretation of the findings.

Students were grouped into like grades of As, Bs, Cs, Ds, and Fs. Means for each group were calculated for each teacher perception of student effort and student attitude toward Algebra I. A line graph was created to show the directional relationship between each independent variable and grading. An ANOVA was run to determine if the difference within and between these means was statistically significant. A Scheffé test was then run between each of the various subgroups to determine exactly where statistical significance lay between the groups

As stated earlier in the chapter, Spearman *rho* correlations were then run to determine the strength of the relationship between each independent variable and grading. Those correlation coefficients were then squared to calculate the coefficient of determination so as to measure the percentage of variance shared between each independent variable and grading.

Limitations

While attempts were made in the design of this study to minimize limitations, including strategies in selecting schools and teachers and the attempt to include a broad number of data points for each variable, limitations still existed. One limitation of this study was that both independent variables were self-reported data based upon each subject teacher's personal perceptions. Teachers may have been influenced by the fact data were collected on students' grades and the teacher's perception of student effort and attitude in class. Teachers' answers about their perceptions of student effort and attitude may have been skewed if they perceived the study to view these as positive or negative

factors impacting grading as seen by the researcher. Furthermore, because building administrators were asked to choose teachers appropriate for the study, and because the researcher needed to explain the purpose of the study to those administrators, the possibility existed that administrators biased the study by choosing teachers whom they believed would positively reflect their beliefs about the appropriate relationship between grades and the independent variables studied.

Another limitation of this study was in its generalizability. Because the study included data from 15 Algebra I teachers in and around the Des Moines, Iowa area, the results were limited in their generalizability. Further research should be conducted to determine if the results of this study apply to other content areas or other school districts in other areas. Further research should be done to determine if the results of this study apply to schools other than secondary schools. Finally, further research should be done to determine if the results of this study apply to other non-academic factors identified in the research as impacting grading of students, but not studied here, such as behavior, attendance, and organization / completion of work on time.

Chapter 4

ANALYSIS OF THE DATA

As discussed in chapter one, the purpose of this study was to quantify the impact that two non-academic factors (teacher perception of student attitude towards the subject matter, and teacher perception of student effort) had on the final grade assigned to ninth grade students Algebra I. This study was designed to answer the following questions:

- Are the differences in mean scores for teachers' perception of student effort statistically significant for students among grade distributions (As, Bs, Cs, Ds, and Fs)?
- Are the differences in mean scores for teachers' perception of student attitude toward subject matter statistically significant for students among grade distributions (As, Bs, Cs, Ds, and Fs)?
- What is the relationship between the final semester grade a student earns in Algebra I and the non-academic factor of student effort as perceived by the teacher?
- What proportion of variance within the final semester grade of a student in Algebra I can be attributed to teacher perception of student effort?
- What is the relationship between the final semester grade a student earns in Algebra I and the non-academic factor of student attitude towards the subject matter as perceived by the teacher?

- What proportion of variance within the final semester grade of a student in Algebra I can be attributed to teacher perception of student attitude toward subject matter?

The researcher hypothesized that there would be statistically significant differences in the means of teachers' perception of students' effort and students' attitude toward Algebra between those students earning passing grades and those failing the course. It was further hypothesized that there would be a significant level of practical significance for each non-academic factor as it related to its shared variance with the final grade students earned in the course. In these next sections, the results of the relationship between teachers' perception of student effort and teachers' perception of student attitude toward subject matter are discussed and summarized as they pertain to the aforementioned research questions.

Descriptive Statistics

Before delving into a statistical analysis of the research questions, a better understanding of the variables is pertinent to better appreciate the composition of the data of the study. Table 1 illustrates that from the 851 student data points, the mean grade students earned in Algebra I in this study was a 2.77 on a 4.0 scale or a C+. For the group as a whole, the teachers perceived the students to be exhibiting somewhere between average effort (3) and above average or good effort (4) as illustrated by the 3.51 mean for effort. Likewise, teachers perceived that the students in the study overall exhibited somewhere between average attitude (3) and above average or good attitude (4) as shown by the 3.51 mean for attitude toward the subject matter. The standard error of the mean of the grade ($SE= 2.779$) indicates that the distribution of grades is not that of a

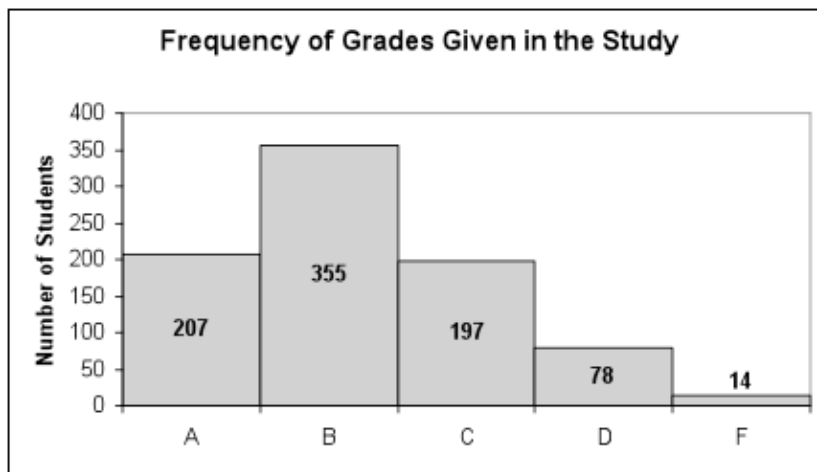
normal curve. Thus, a frequency distribution of the grades given to students in this study became relevant to its analysis, as shown in Figure 1.

Table 1 – Descriptive Statistics for Dependent and Independent Variables

	Mean	SE	SD	Variance	Count
Grade	2.7791	2.7791	0.9704	0.9417	851
Effort	3.5076	0.0378	1.1022	1.2149	851
Attitude	3.5065	0.0342	0.9969	0.9938	851

Figure 1 indicates that the grade distribution for the 851 students who had grades reported is not a normal curve. Rather, the lack of students earning an F ($n=14$) or D ($n=78$) 10.8%, is skewed compared the number of students earning As ($n= 207$) and Bs ($n= 355$) 66.0%. This lack of a normal curve in the grade distribution of the sample was important to keep in mind as the results were analyzed.

Figure 1 – Frequency Distribution of Grades



Basic descriptive statistics were calculated for each of the independent variables.

Table 2 shows the descriptive statistics for teacher perception of effort. Students who

earned As ($M = 4.50$) evidenced a difference in their average rating of effort from did those students earning other grades: Bs ($M = 3.60$), Cs ($M = 2.91$), Ds ($M = 2.34$), and Fs ($M = 1.43$). The low range of standard error of the mean ($SE = .044 - .202$) and the low range of variance ($.514 - .885$) suggested this distribution of ranks indicated a directional relationship.

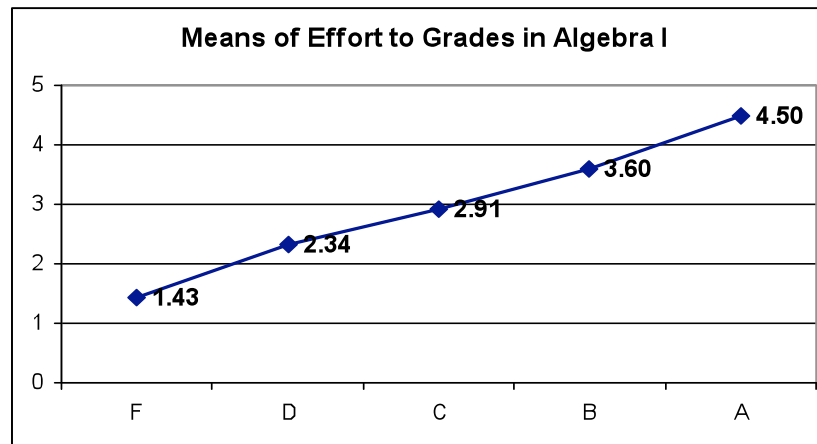
Table 2- Descriptive Statistics for Teacher Perception of Effort for the Grade Distribution

	Mean	SE	SD	Variance	Count
A	4.4979	.0466	.7168	.5138	237
B	3.6020	.0438	.8843	.7821	407
C	2.9083	.0533	.7863	.6183	218
D	2.3373	.1032	.9406	.8848	83
F	1.4286	.2020	.7559	.7559	14

Noting the differences in the means of students earning differing grades, one can see a positive directional relationship between the means of these ranked scores of teacher perception, and the ranking of grades given to students within this distribution, as seen in Figure 2.

Figure 2 depicts a positive directional relationship between grades earned in Algebra I and teachers' perception of student effort. As student grades rose, so did teacher perception of effort. As stated above, the lower standard error of the mean and low of variance suggested that the ranking of effort approximated this relationship. Thus analysis of the statistical significance of the difference between means was a valid question. An ANOVA was run to determine if the differences between the means were significant.

Figure 2 – Means of Teacher Perception of Effort compared to Grades in Algebra I



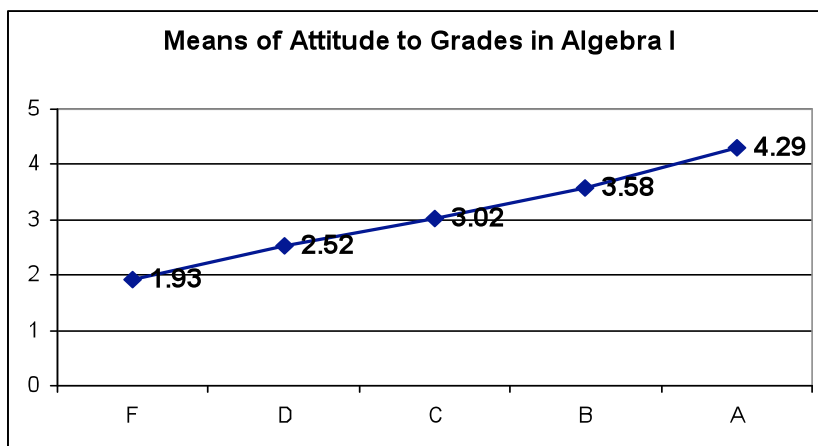
Descriptive statistics for teacher perception of attitude toward subject matter were calculated. Table 3 shows those results. Much like the data from the independent variable for effort, students who earned an A ($M = 4.29$) evidenced a difference in their average rating of attitude toward subject matter, than did those students earning other grades: Bs ($M = 3.58$), Cs ($M = 3.02$), Ds ($M = 2.52$), and Fs ($M = 1.93$). Again, the low range of standard error of the mean ($SE = .042 - .221$) and the low range of variance (.472 – .717) suggested that this distribution of ranks were indicative of a directional relationship. This provided validity to running an ANOVA to determine if the means of teacher perception of student attitude toward subject matter were statistically significantly different.

Table 3 – Descriptive Statistics for Teacher Perception of Attitude towards Algebra I for the Grade Distribution

	Mean	SE	SD	Variance	Count
A	4.2911	.0476	.7333	.5378	237
B	3.5799	.0420	.8468	.7171	407
C	3.0229	.0557	.8227	.6769	218
D	2.5181	.0754	.6872	.4722	83
F	1.9286	.2215	.8287	.6868	14

The differences in the means of the distribution of ranks depicted a positive directional relationship between grades and teacher perception of attitude. Much like teacher perception of effort, as teachers perceived students to have a more positive attitude towards their subject matter, their grades also increased. This directional relationship is illustrated in Figure 3 below.

Figure 3 - Means of Teacher Perception of Attitude compared to Grades in Algebra I



Research Questions

Upon gaining an understanding of the composition of each variable, one is ready to explore the specific research questions addressed within this study. While the distribution of grades was not indicative of a normal curve, the low range of standard

error and low variance for both independent variables, combined with the positive directional relationship of the rank ordered data for both teacher perception of effort and attitude suggests the questions of the study were validly explored through correlational research.

Question 1: Are the differences in mean scores for teachers' perception of student effort statistically significant for students among grade distributions (As, Bs, Cs, Ds, and Fs)?

Null hypothesis – differences in the means of student effort as perceived by the teacher for students earning different grades are not statistically significant.

As revealed in Table 3 (page 44), there were differences in the mean ranks given to students earning different grades within the distribution for teacher perception of effort. Those means illustrated a positive directional relationship. Table 4 evidences the results of a single factor ANOVA run to determine if the difference of those means was statistically significant.

Table 4 explains that the F value of the variance between groups in this distribution was 146.79. The F critical value of 2.38 determined that there was a statistically significant difference between the means of the groups within this distribution. Furthermore, the P value of 2.55 E-95 indicated that this difference did not occur by chance.

Table 4 – ANOVA for Statistical Significance of Difference of Effort Means

Anova: Single
Factor

SUMMARY					
Groups	Count	Sum	Average	Variance	
A	207	923	4.4589	0.5408	
B	355	1283	3.6141	0.8252	

C	197	575	2.9188	0.6464
D	78	184	2.3590	0.9344
F	14	20	1.4286	0.5714

ANOVA						
Source of Variation	SS	df	MS	F	P value	F crit
Between Groups	423.092	4	105.7730	146.7893	2.55E-95	2.3825
Within Groups	609.6083	846	0.7206			
Total	1032.7	850				

A Scheffé test was run against each possible pairing of groups within this distribution to determine where the statistically significant differences were. The results of the Scheffé test are made clear in Table 5.

Analysis of Table 5 shows that all pairings of grade rankings were statistically significantly difference from each other. *F* values established from the Scheffé test ranged from 14.26 when comparing means for teacher perception of effort between students who earned Ds compared to those who earned Fs, to 346.71 when comparing means for teacher perception of effort for students who earned As compared to those who earned Ds. When compared to the *F* critical value of 2.38 as determined by the ANOVA and through the process of running the Scheffé test, it was determined the there was a statistically significant difference between each and every mean for teacher perception of effort. In other words, there were actual differences in how teachers perceived the effort of the students in this study who earned different grades, and these differences were not due to chance. As a result of these findings, the researcher rejected the null hypothesis.

Table 5 – Scheffé Test for Statistically Significant Differences in Means of Effort

	A			
B	129.5221	B		
C	332.2797	85.00078	C	
D	346.7071	139.8036	86.81928	D
F	167.1151	89.28044	40.28343	14.25924

Question 2: Are the differences in mean scores for teachers' perception of student attitude toward subject matter statistically significant for students among grade distributions (As, Bs, Cs, Ds, and Fs)?

Null hypothesis - the difference in the means of teacher perception of student attitude toward subject matter for students earning different grades are not statistically significantly different.

As described in Table 4 (page 45), there were differences in the mean ranks given to students earning different grades within the distribution for teacher perception of student attitude toward subject matter. Those means indicated a positive directional relationship. Table 6 depicts the results of a single factor ANOVA run to determine if the difference of those means was statistically significant. Table 6 indicates that the F value of the variance between groups in this distribution was 103.96. The F -critical value of 2.38 determined that there was a statistically significant difference between the means of the groups within this distribution. Furthermore, the P value of 5.03 E-72 indicated that this difference did not occur by chance.

Table 6 – ANOVA for Statistical Significance of Difference of Attitude Means

Anova: Single Factor

SUMMARY					
Groups	Count	Sum	Average	Variance	
A	207	883	4.2657	0.5650	
B	355	1280	3.6056	0.7536	
C	197	598	3.0355	0.6977	
D	78	196	2.5128	0.4868	
F	14	27	1.9286	0.6868	

ANOVA						
Source of Variation	SS	df	MS	F	P value	F crit
Between Groups	278.3722	4	69.5931	103.9579	5.03E-72	2.3825
Within Groups	566.3422	846	0.6694			
Total	844.7145	850				

A Scheffé test was run against each possible pairing of groups within the grade distribution to determine exactly where the statistically significant difference was. The results of the Scheffé test are described in Table 7. Analysis of Table 7 shows that all pairings of grade rankings were statistically significantly different. F values from the Scheffé test range from 6.05 when comparing means for teacher perception of student attitude towards subject matter between students who earned Ds compared to those who earned Fs, to 260.02 when comparing means for teacher perception of student attitude toward subject matter for students who earned As compared to those who earned Ds. When compared to the F -critical value of 2.38 as determined by the ANOVA and through the process of running the Scheffé test, it was determined the there was a statistically significant difference between each mean for teacher perception of student attitude toward Algebra I. In other words, there were actual differences in how teachers perceived the attitude of students toward their subject matter who earned different grades,

and it was highly unlikely that this difference was due to chance. As a result, the researcher rejected the null hypothesis.

Table 7 – Scheffé Test for Statistically Significant Differences in Means of Effort

	A			
B	85.1000	B		
C	228.1787	31.0477	C	
D	260.0255	114.0823	22.8058	D
F	106.9949	56.5874	23.9259	6.0523

When both Scheffé tests run for each independent variable were compared it was interesting to note that the lowest F value was the same comparison for each independent variable (the comparison of Ds to Fs), as was the highest F value (As to Ds). Further analysis of both Scheffé tests informs that five of the ten comparisons between groups for both independent variables shared the same ranking of F value strength relative to the other comparisons. This, along with the similarities in the line graphs of the means of the groups having similar directional relationships and the consistently low standards of error and variance for both variables, caused the researcher to wonder about the relationship between teacher perception of student effort and teacher perception of student attitude toward subject matter. It appeared as if the two variables were closely related.

Question 3: What is the relationship between the final semester grade a student earns in Algebra I and the non-academic factor of student effort as perceived by the teacher?

Null hypothesis – there is no relationship between the final semester grade a student earns in Algebra I and the non-academic factor of student effort as perceived by the teacher.

The ranks of one variable do not co-vary with the ranks of the other variable; in other words, as the ranks of one variable increase, the ranks of the other variable are not more likely to increase (or decrease).

In order to determine if a relationship existed between variables, typically a scatter plot would be used to determine the strength of a directional relationship to justify the use of correlations. However, because both the grades given to students in this study, as well as the Likert scale options given to teachers to evaluate their perception of student effort were rank-ordered data sets, the use of a scatter plot to determine a linear relationship was invalidated (McDonald, 2009). However, a linear relationship was seen by charting the relationship of the means of the variables and the grades, as seen in Figure 2 (page 43).

After determining that a positive linear relationship did exist between grades earned and teacher perception of effort, a Spearman *rho* was run on the data set to determine the strength of the relationship. Again, the Spearman *rho* was a more appropriate correlative test due to the rank-ordered nature of the data set (McDonald, 2009). Table 8 illustrates the results of the Spearman *rho* for teacher perception of student effort. Table 8 depicts that a statistically strong relationship existed between teachers' perception of student effort and the letter grade earned in Algebra I. According to Fraenkel and Wallen, (2006) any correlation coefficient (r) above .40 is generally significant enough to warrant consideration for practical and theoretical use. They further stated that correlation coefficients that show a strength of relationship above the .60 level indicate a very important relationship between the variables. Teacher perception of student effort had a very strong correlation ($r = .6267$). This correlation was run at the 0.05 level of significance. The P value generated by the Spearman's *rho* evidences that the significance of the relationship was unlikely to have occurred by chance ($p = 4.86E-94$). The results of the Spearman *rho* make clear that

there is most definitely co-variance between grades and teacher perception of student effort, and thus the null hypothesis was rejected.

Table 8 - Spearman *rho* correlation between teacher perception of student effort and grading, with coefficient of determination

Spearman's <i>rho</i> :	.6267
degrees of freedom:	849
<i>P</i> value:	4.86E-94
Coefficient of Determination:	.3927

Question 4: What proportion of variance within the final semester grade of a student in Algebra I can be attributed to teacher perception of student effort?

Using the data from Table 8 (page 51), one can see that the coefficient of determination, or the percentage of variance shared between the variables of grading and teacher perception of student effort was also very strong ($r^2 = .3927$). In other words, up to 39 percent of the final letter grade that a student in this study earned in Algebra I may have been influenced by the teacher's perception of that student's effort. These findings are similar in strength of relationship to data found by other researchers who have quantified the relationship between effort and grading (Rich, 2001).

Question 5: What is the relationship between the final semester grade a student earns in Algebra I and the non-academic factor of student attitude towards the subject matter as perceived by the teacher?

Null hypothesis – there is no relationship between the final semester grade a student earns in Algebra I and the non-academic factor of student attitude toward subject matter as perceived by the teacher. The ranks of one variable do not co-vary with the ranks of the

other variable; in other words, as the ranks of one variable increase, the ranks of the other variable are not more likely to increase (or decrease).

Just as with the independent variable of teacher perception of student effort, because both the dependent variable of grade and the independent variable of teacher perception of student attitude toward subject matter were rank-ordered data the use of a scatter plot to determine a directional relationship upon which to evaluate the validity of a correlational relationship was compromised. Thus, the researcher used the positive directional relationship suggested by Figure 3 (page 44) when comparing the letter grades earned by students in the study to the means earned by the different groups within the grade distribution for teacher perception of student attitude toward subject matter. This relationship suggested that an examination of the correlation between these variables would be appropriate.

After determining that a positive linear relationship did exist between grades earned and teacher perception of student attitude toward subject matter, a Spearman *rho* was run on the data set to determine the strength of the relationship. Again, the Spearman *rho* was a more appropriate correlative test due to the rank-ordered nature of the data set (McDonald, 2009). Table 9 illustrates the results of the Spearman *rho* for teacher perception of student attitude toward subject matter. Table 9 evidences that a statistically strong relationship existed between teachers' perception of student effort and the letter grade earned in Algebra I as this correlation coefficient is above the .40 level declared as generally significant enough to warrant consideration for practical and theoretical use by Fraenkel and Wallen (2006). The correlation coefficient for teacher perception of students' attitude, while strong ($r = .5743$), was not as strong as was teacher

perception of student effort ($r = .6267$). The correlation for grades and attitude was run at the 0.05 level of significance. The P value generated by the Spearman's ρ clarifies that the significance of the relationship shown here is unlikely to have occurred by chance ($p = 7.73E-76$). The results of the Spearman ρ indicate that there is co-variance between grades and teacher perception of student attitude, and thus the null hypothesis was rejected.

Table 9 - Spearman ρ correlation between teacher perception of student attitude toward subject matter and grading, with coefficient of determination

Spearman's ρ :	.5743
degrees of freedom:	849
P value:	7.73E-76
Coefficient of Determination	.3298

Question 6: What proportion of variance within the final semester grade of a student in Algebra I can be attributed to teacher perception of student attitude toward subject matter?

Using the data from Table 9, one can see that the coefficient of determination, or the percentage of variance shared between the variables of grading and teacher perception of student attitude toward subject matter is very strong ($r^2 = .3298$). In other words, up to 33 percent of the final letter grade that students in this study earned in Algebra I may have been influenced by the teacher's perception of that student's attitude toward the subject matter.

Summary

The findings of this research demonstrated that there were statistically significant differences in the mean rankings of teacher perception of student effort and student

attitude toward Algebra I. This indicated that for the students in this study, teachers perceived the effort and attitude of those earning As differently than those earning other letter grades. Furthermore, this difference in perception was not by chance. Teachers truly saw these students as different. These research findings also substantiated that the teacher perception of these two independent variables played a significant role in determining the overall grade students in this study earned in Algebra I.

Chapter 5

SUMMARY, CONCLUSIONS, DISCUSSION AND RECOMMENDATIONS

Discussion

“Grades are probably the most common and public use of educational measurement. They are an integral part of classroom instruction” (Brookhart, 1994, p. 299). Yet unfortunately, these common and public uses of measurement are unclear. Because of the inclusion of a variety of non-academic factors, such as attendance, behavior, effort, attitude, compliance, or organization, what is expressed by a letter grade about what a student knows and is able to do within a subject is unclear. And because the grade issued has a variety of intended audiences and intended messages; including providing information to the student about their performance, giving documentation to parents about achievement, providing colleges or future employers with information needed for admissions or acceptance to a job, or giving information to future teachers about the skills and abilities of students they are educating, what is communicated through a single letter mark can not help but be complicated with a lack of clarity.

What is clear, however, is that both teachers and parents believe that grades are very important (Waltman & Frisbie, 1994). Even though teachers express they do not enjoy grading as a responsibility of their job (Lomax, 1996) and find it to be a difficult aspect of their job (Barnes, 1985), they understand well the grades students earn are very important. The grades students earn have power to impact students in a number of ways. Certainly, the post-secondary educational options and employment options of students are impacted by the grades they receive. Failing grades create obstacles to successful

completion of high school, as well as send the message to those failing that they may not belong in school. It is for these reasons Brookhart (1994) wrote, “grades do not merely measure the instructional process, they participate in it” (p. 299).

Unfortunately, the role of that participation is not always what one would desire for students. Considerable evidence asserts that many common grading practices actually lower student performance (Black & Wiliam, 1998). The arbitrary nature of invalidity in grading may cause students to disengage with the work of earning good grades. Studies of talented and gifted students are abundant with documentation of failure as a result disenfranchised students (Csikszentmihályi, Rathunde, & Whalen, 1997). Students are able to detect the unfairness of invalid grading practices. While some students become adept at working within the system to earn good grades, others shut down and refuse to comply.

It is for these reasons that researchers and measurement experts have argued the importance of validity in grading practices. As discussed in chapters one and two, validity in grading is both essential for clear communication, and has been a topic of controversy for some time. Research studies from a number of well respected researchers have concluded that much is included when teachers grade beyond what students know and are able to do. Brookhart (1993) found that the meaning of grades varied extensively among teachers. If grades are to be meaningful to any of the above mentioned audiences, practitioners must grade with validity. Basing grades solely on what students know and are able to do (achievement) is essential to validity in grading. If one letter is to be used to represent the achievement of a student in a course, then only factors based upon what students know and are able to do should influence the grade. To include anything else in

the factoring of that grade is to invalidate the grade itself. As Tomlinson (1994) stated, grades that are based upon factors beyond achievement may constitute an easy lie that tells nothing about students' actual performance.

Despite the history of concern expressed about invalid grading practices, the prevalence of sound research about what factors complicate the communicate of grading, and the documentation of the importance of grades on the lives of students, little change has occurred in the day-to-day grading practices of teachers. Allen (2005) posited that despite longstanding concerns about the role of validity and reliability in grading, they were assumed to meaningful communication about student academic achievement. Numerous attempts to help teachers understand the purposes and functions of grades within the overall evaluation system have resulted in little change in classroom practice. Likewise, it has been found that even providing teachers with measurement instruction made almost no difference in grading practice (Brookhart, 1993).

Knowing that multiple studies about the damage that invalid grading practices can do, and knowing that despite this research little has changed in the way many secondary teachers grade, the researcher desired to look at the relationship between two non-academic factors and grading in a different light. Brookhart (1994) wrote, "More training, by itself, will not cause grading practices to conform completely to (measurement specialist) recommendations" (p. 290). Thus, in order to encourage changed practices, both in practitioners giving grades and those institutions preparing future teachers, the extent to which inclusion of non-academic factors influence grading at the secondary level needed to be examined. By quantifying the extent of the

relationship between non-academic factors and final grades given by teachers, there lies potential for a reexamination of current practice.

Review of the Findings

Question 1: Are the differences in mean scores for teachers' perception of student effort statistically significant for students among grade distributions (As, Bs, Cs, Ds, and Fs)?

This research data affirmed that there was a difference in how teachers viewed the effort of students in this study who earned differing letter grades. The data also showed that this difference was statistically significant, and that this difference was highly unlikely to have occurred due to chance. Furthermore, a statistically significant difference was found between all grade groupings in how teachers perceived the effort of students. The results of this study confirm the results of previous studies into how teachers view the effort of students achieving at differing levels.

Before attempting to document the research suggesting effort is confounded with grading, it is important to understand what is meant by effort. Carbonaro (2005) defines student effort as "the amount of time and energy that students expend in meeting the formal academic requirements established by their teacher and/or school" (p. 28). He goes on to suggest three distinct types of effort. Carbonaro's (2005) definitions for procedural effort (when students try to meet the specific demands set forth by a teacher in a particular class, including completing assignments, participating in class, and turning work in on time) and intellectual effort (when students apply their cognitive facilities toward understanding the intellectual challenges posed by the curriculum) were most applicable to the effort reported within this study.

Combining effort and achievement in grading practices has been well documented (Brookhart, 1994). Cizek, Fitzgerald, & Rachor (1995) found that of the 143 teachers they surveyed, 41.9% stated they considered effort when assigning grades. Kelly (2008) found that 96.6% of the teachers interviewed in his study stated that effort played a role in how grades were determined. Manke and Loyd (1990) and Griswold (1993) utilized scenarios in which effort was used as a mitigating factor in grade decisions as a research design. Both discovered that effort was used pervasively by teachers when making grading decisions. Carbonaro (2005) concluded the higher the students' track (rigorous course of study), the more effort he or she exerts. Teachers reported they use a variety of factors, including effort, when assigning grades (Cizek et al., 1996), even when they knew it was contrary to recommendations, especially for lower level students (Stiggins, 1991).

This problem, more pervasive in the elementary setting, results from teachers' desires to offer feedback to students that they can use to change behavior to be more successful, often including the need to put forth more effort (Brookhart, 1994). Griswold (1993) found teachers expect student outcomes that are broader than achievement. Teachers desire to see students grow not only academically, but in their responsibility and citizenship. Thus, effort may be seen by teachers as both an outcome separate from academic achievement, as well as a mediator of academic achievement (Griswold, 1993). Brookhart (1994) suggested that finding valid and reliable ways to measure and report effort measures could offer multiple benefits, including offering feedback that can help change student behavior, enhance student learning, improve classroom management, and provide greater validity to the academic grades separated from effort marks. Ames

(1992) found when students are rewarded for exerting effort, they are more likely to continue to exert effort in the future, and a cycle of growth in achievement will unfold. Thus, it can be understood why so many teachers consider effort, and create connections between effort and grading, as both a formative process to provide feedback for growth, as well as a summative process for measuring achievement.

This study reconfirmed that effort is interwoven with achievement. The level of student effort perceived by the teachers within this study was statistically significantly different based upon the level of achievement of the student.

Question 2: Are the differences in mean scores for teachers' perception of student attitude toward subject matter statistically significant for students among grade distributions (As, Bs, Cs, Ds, and Fs)?

While not nearly as extensively researched as the non-academic factor of effort, the data of this study also reconfirmed findings of previous studies. Teachers do consider student attitude when making grading decisions (Manke & Loyd, 1990; Griswold, 1993).

This study indicated that how teachers perceive student attitude toward their subject matter did differ significantly when considering the level of achievement of the student. In other words, students who earned an A were perceived to have a better attitude toward the subject matter as those who received a grade other than an A. Furthermore, a statistically significant difference was found between all grade groupings in how teachers perceived the attitude of students toward subject matter. This significant difference was unlikely to have occurred by chance.

Question 3: What is the relationship between the final semester grade a student earns in Algebra I and the non-academic factor of student effort as perceived by the teacher?

Measured through a correlation coefficient, the relationship between a teachers' perception of student effort and the final grade in a course was significant ($r = .6267$). The strength of this relationship suggested that what has been well documented as an invalidating factor in determining grades made up a significant portion of the overall grade given. The p value of the correlation indicated that the strength of this relationship was highly unlikely to have occurred by chance. This finding is similar to that found by Rich (2001) when he found a relationship between effort and grading that was slightly more significant ($r = .717$).

Question 4: What proportion of variance within the final semester grade of a student in Algebra I can be attributed to teacher perception of student effort?

The impact of teacher perception of effort on the final letter grade earned by students in this study made up a significant portion of that grade ($r^2 = .3927$). The proportion of variance that was shared between the variables was 39.3%. In other words, 39 percent of the grade students earned in Algebra I in this study can be attributed to the teacher's perception of those students' effort. This is a quantifiable affirmation of Carbonaro's (2005) assertion that effort is an important predictor of achievement.

Question 5: What is the relationship between the final semester grade a student earns in Algebra I and the non-academic factor of student attitude towards the subject matter as perceived by the teacher?

The relationship between teachers' perceptions of students' attitudes toward the subject matter and the final letter grade earned in Algebra I was also significant ($r = .5743$). This suggested that attitude was much more than an important factor included in grading. Rather, how teachers perceived students' attitudes toward the subject matter

played a significant role in determining the overall grade of that student. The p value of the correlation indicated the strength of this relationship was highly unlikely to have occurred by chance.

Question 6: What proportion of variance within the final semester grade of a student in Algebra I can be attributed to teacher perception of student attitude toward subject matter?

The coefficient of determination for teacher perception of student attitude toward subject matter ($r^2 = .3298$) indicated this non-academic factor played a significant role in determining students' grades. In other words, 33 percent of the grade students in this study earned in Algebra I can be attributed to their teacher's perception of their attitude toward the subject matter.

Suggestions for Further Research

Due to the inherently complex task of assigning grades to students, the number of possible factors that influence the grade including academic and non-academic factors, and the variety of audiences for whom grades are intended, much research is yet needed to understand how grade determinations are made and what impact those decisions have on students. It is the belief of the researcher that more quantitative research in the grading practices of teachers is needed. Noting this researcher's call for more quantitative research, as well as the limitations of this study, the following is suggested as further research.

In the procedures of this study, the researcher met with administrators in schools to explain the project and seek their permission to conduct research. It is possible that because teacher participants were selected based upon administrator recommendation,

that administrators biased the study by choosing teachers whom they assumed would positively reflect their beliefs about the appropriate relationship between grades and the independent variables studied. Likewise, the researcher met personally with each teacher participant to explain the study (prior to their issuance of final semester grades) and sought their voluntary participation. It is possible (although unlikely due to the nature of the results) that teachers altered their grading decisions based upon their knowledge of what was being researched. Further research in this area should attempt to seek to randomize the sample to ascertain if results can be verified.

While 851 student data points were generated and analyzed for this study, only 15 participant teachers made up the subjects of the study. Furthermore, all of the teachers taught Algebra I in secondary schools in or around the Des Moines, Iowa metropolitan area. Future research should attempt to include a broader base of teacher participants from other a variety of other areas of the country. Further research should be conducted to determine if the results of this study are generalizable to other content areas. Would English teachers or art teachers yield different results? Would a combination of teachers from different content areas yield different results? Would teachers from other areas of the country yield different results? Would elementary teachers demonstrate differing levels of inclusion of non-academic factors? Further research should be done to determine if the results of this study apply to schools other than secondary schools. Finally, further research should be done to determine if the results of this study apply to other non-academic factors identified in the research as impacting grading of students, but not studied here, such as behavior, attendance, and organization or completion of work on time.

Because the data collected for this study were rank-ordered data for all of the variables researched, the researcher was not able to complete a regression analysis of the data. Therefore predictability of the results may in question. Further research is suggested to determine the predictability of the results of this study. Research should be done to gather perception of effort and attitude ranks (or ranks of other non-academic factors mentioned within the research base) from teachers prior to the issuance of final grades. The researcher could then use those marks to predict final letter grades for students based upon the percentage of variance shared between the variables and the different means for each variable associated with each grade as determined by this study. The researcher would then collect final letter grades for those students, comparing the predicted grade with the actual final grade given to determine validity of predictability of the relationship between effort and grading and attitude and grading.

Summary and Conclusions

The purpose of this study was to examine the relationship between two non-academic factors, teachers' perception of student effort and teachers' perceptions of student attitude toward subject matter, and the final grade students earned in Algebra I. The researcher hypothesized there would be a significant relationship between these variables. The researcher also hypothesized that how teachers perceived these variables would differ significantly for students who earned As as compared to those who earned Fs. The results of the study verified the researcher's hypotheses. The researcher also found a statistically significant difference for all other grade comparisons for both variables.

That being said, the size of the impact of these variables on grading was surprising to the researcher. While numerous studies have concluded that non-academic factors do influence, and invalidate, grades, the extent to which these factors influence grades has been researched little. Knowing the grades students in this study earned in Algebra I were impacted as significantly as they were should be cause for concern and reason for reexamination of grading practices. The fact that how teachers perceived students' effort influenced the final grade students earned in Algebra up to 39 percent of the grade, and that how teachers perceived students' attitude toward the subject matter influenced the final grade students earned up to 33 percent has serious implications, both for the students and schools systems participating in the study and for educational systems on a broad scale.

On a practical level, the fact that students' grades were influenced by teachers' perceptions of effort and attitude by 39 percent and 33 percent respectively means the grades reported for these students were invalidated by the inclusion of non-academic factors and therefore lacked the clarity of meaning needed to be purposeful to students, parents, or future teachers. The results of this study call for teachers and administrators within these districts to revisit the grading policies, professional development around how to grade with validity, and grade reporting structures used. On a broader scale, the results of this study confirm statistically the findings of numerous researchers that non-academic factors do influence grading. The strength of the relationship between these factors and grades should serve as impetus for change in how grades are determined and reported.

Because of a lack of training provided to pre service teachers (Goodwin, 2001), they are too often left to themselves to determine how they will grade. Without proper

training on measurement principles, teachers often turn to others in the profession to help determine how they will grade, or grade their students in the same manner in which they were graded (Allen, 2005). The results of this study, particularly in the strength of the relationship, suggest that leaving teachers to determine grading practices in isolation, without guidance in measurement principles to avoid invalidating grades through the inclusion of non-academic factors, is not in the best interest of students or any other person relying on grades to communicate achievement results. This suggests responsibility for teacher preparatory institutions to ensure that training of sound measurement principles is mandatory for pre service teachers. The results also call for school systems to provide professional development to practicing teachers on similar principles of measurement to grade with validity.

Much of the miscommunication emanating from grades invalidated through the inclusion of non-academic factors is complicated by the reporting tool available to those issuing the grade. Because information about student achievement, effort, attitude, attendance, behavior, compliance, and other factors that are viewed as important to the development of students is forced to be represented by a single letter, what is interpreted by those viewing those grades is subjective. Letter grades allow for a concise, efficient, well recognized way to store and communicate information about achievement. Because of their efficiency and convenience, grades will be a permanent fixture in our educational system. However, how those grades are reported can, and should, be changed. The results of this study lend credence to measurement expert suggestions that a dual reporting system be used. By adopting reporting tools that allow for achievement measures to be separated from non achievement measures, schools have the ability to

improve the clarity of their communication. By grading only student achievement measures, and reporting on grading factors such as effort, behavior, attitude, etc. separately, schools have the ability to clearly communicate, free of invalidating biases, what students know and are able to do within a given academic content area.

BIBLIOGRAPHY

- Airasian, P. W., & Jones, A. M. (1993). The teacher as applied measurer: Realities of classroom measurement and assessment. *Applied Measurement in Education, 6*, 241-254.
- Allen, J. D. (2005). Grades as valid measures of academic achievement of classroom learning. *The Clearing House, 78*(5), 218-223.
- Allen, J. D., & Lambating, J. (2001, April). Validity and reliability in assessment and grading: Perspectives of pre service and in-service teachers and teacher education professors. Paper presented at the annual meeting of the American Educational Research Association. Seattle, WA.
- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology, 84*, 261-271.
- Baily, J., & McTighe, J. (1996). Reporting achievement at the secondary level: What and how. *Educational Assessment, 4*(1), 29-42.
- Barnes, S. (1985). A study of classroom pupil evaluation: The missing link in teacher education. *Journal of Teacher Education, 36*(4), 46-49.
- Black, P. J., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy, and Practice, 5*(1), 7-73.
- Brookhart, S. M. (1991). Grading practices and validity. *Educational Measurement: Issues and Practice, 101*(1), 35-36.
- Brookhart, S. M. (1993). Teacher grading practices: Meaning and values. *Journal of Educational Measurement, 30*(2), 123-142.

- Brookhart, S. (1994). Teacher's grading: Practice and theory. *Applied Measurement in Education*, 7(4), 279-301.
- Brookhart, S. (2004). *Grading*. Upper Saddle River, NJ: Pearson/Merrill/Prentice Hall.
- Buzzelli, C. A., & Johnson, B. (2002). *The moral dimensions of teaching: Language, power, and culture in classroom interaction*. New York: Routledge/Falmer Press.
- Carbonaro, W. (2005). Tracking students' effort, and academic achievement. *Sociology of Education* 78, 27-49.
- Csikszentmihályi, M., Rathunde, K., & Whalen, S. (1997). *Talented teenagers: The roots of success and failure*. Edinburgh, UK: Cambridge University Press.
- Cizek, G., Fitzgerald, S., & Rachor, R. (1995). Teachers' assessment practices: Preparation, isolation, and the kitchen sink. *Educational Assessment*, 3(2), 159.
- Clymer, J. B., & Wiliam, D. (2007). Improving the way we grade science. *Educational Leadership*, 36-42.
- Cross, L. H., & Frary, R. B. (1996, April). Hodgpodge grading: Endorsed by students and teachers alike. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.
- Cuerton, L. W. (1971) The history of grading practices. *Measurement in Education*, 2(4), 1-8.
- Dockery, E. R. (1995). Better grading practices. *Education Digest*, 60(5), 34-37.
- Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of educational measurement*. (4th ed.). Englewood Cliffs, NJ: Prentice Hall.

- Entwisle, D. R., & Alexander, K. L. (1988). Factors affecting achievement test scores and marks of black and white first graders. *Elementary School Journal* 88, 449-471.
- Finkelstein, I. E. (1913). The marking system in theory and practice. *Educational Psychology Monographs* (10), 1-26.
- Fraenkel, J. R., & Wallen, N. E. (2006). *How to design and evaluate research in education*, (6th ed.). Boston: McGraw Hill.
- Frary, R. B., Cross, L. H., & Weber, L. J. (1993). Testing and grading practices and opinions of secondary teachers of academic subjects: Implications for instruction and measurement. *Educational Measurement: Issues and Practice*, 12(3), 23-30.
- Friedman, S. J., & Frisbie, D. A. (1995). The influence of report cards on the validity of grades reported to parents. *Educational and Psychological Measurement*, 55(1), 5-26.
- Gallagher, J. D. (1998). *Classroom assessment for teachers*. Upper Saddle River, NJ: Merrill / Prentice Hall.
- Glasser, W. (1971). *The effect of failure on the life of a child*. Washington DC: National Association of Elementary School Principals.
- Goodwin, A. L. (2001, April). The case of one child: Making the shift from personal knowledge to informed practice. Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA.
- Gredler, M. E. (1999). *Classroom assessment and learning*. New York: Longman.
- Griswold, P. A. (1993). Beliefs and inferences about grading elicited from student performance sketches. *Educational Assessment*, 1(4), 311-328.

- Guskey, T. R. (1996). *Communicating student learning*. ASCD Yearbook. Alexandria: VA: Association for Supervision and Curriculum Development.
- Hargis, C. H. (1990). Grades and grading practices: Obstacles to improving education and helping at-risk students. Springfield, IL: Charles C. Thomas.
- Hopkins, K. D., Stanley, J. C., & Hopkins, B. R. (1990). *Educational and psychological measurement and evaluation*, (7th ed.). Englewood Cliffs, NJ: Prentice Hall.
- Jasmine, T. (1999). Grade distributions, grading procedures, and students' evaluations of instructors: A justice perspective. *Journal of Psychology*, 133(3), 263-271.
- Kain, D. L. (1996). Looking beneath the surface. Teacher collaboration the lens of grading practices. *Teachers College Record*, 97(4), 569-587.
- Karmel, L. J. (1970). *Measurement and evaluation in the schools*. Toronto, ONT: Collier-MacMillan.
- Kelly, S. (2008). What types of students' effort are rewarded with high marks? *Sociology of Education*, 81(1), 32-52.
- Kohn, A. (1999). Grading is degrading. *Education Digest*, 65(1), 59-64.
- Linn, R. L., & Gronlund, N. E. (2000). *Measurement and assessment in teaching*. (8th ed.). Englewood Cliffs, NJ: Merrill/Prentice Hall.
- Lomax, R. G. (1996). On becoming assessment literate: An initial look at pre service teachers' beliefs and practices. *Teacher Educator*, 31(4), 292-303.
- Manke, M. P., & Loyd, B. H. (1990). *An investigation of non achievement related factors influencing teachers grading practices*. Boston: National Council on Measurement of Education.

- Marzano, R. J. (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.
- McDonald, J. H. (2009). *Handbook of biological statistics*, (2nd ed.). Baltimore, Maryland: Sparky House Publishing.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher*, 18(2), 5-11.
- Nitko, A. J. (2001). *Educational assessment of students*, (3rd ed.). Upper Saddle River, NJ. Merrill/Prentice Hall.
- Noddings, N. (2002). *Educating moral people: A caring alternative to character education*. New York: Teachers College Press.
- Olson, G. H. (1989, March). On validity of performance grades: The relationship between teacher assigned grades and standard measures of subject matter acquisition. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.
- Oosterhof, A. (2001). *Classroom application of educational measurement*. Upper Saddle River, NJ: Prentice Hall.
- O'Connor, K. (2009). *How to grade for learning, K-12*, (3rd ed.). Thousand Oaks, CA: Corwin
- O'Sullivan, R. G., & Chalnack, M. K. (1991). Measurement related course work requirements for teacher certification and recertification. *Educational Measurement: Issues and Practice*, 10(1), 17-19.

- Pintrich, P.R., & Schunk, D. H. (2002). *Motivation in education*, (2nd ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Rich, R. H. (2001). *Hidden factors in teachers secondary grading practices*. Unpublished doctoral dissertation: Seton Hall University.
- Rosenbaum, J. (2001). *Beyond college for all: Career paths for the forgotten half*. New York: Russell Sage Foundation.
- Schafer, W. D. (1991). Essential assessment skills in professional education of teachers. *Educational Measurement: Issues and Practice*, 10(1), 3-6.
- Sinclair, R. L., & Ghory, W. I. (1987). *Reaching marginal students: A primary concern for school renewal*. Berkeley, CA: McCutchan.
- Smallwood, L. M. (1935). *Examinations and grading systems in early American universities*: Cambridge: Harvard University Press.
- Stiggins, R. J. (1991). Relevant classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 10(1), 7-12.
- Stiggins, R. J. (1999). Evaluating classroom assessment training in teacher education programs. *Educational Measurement: Issues and Practice*, 18(1), 23-27.
- Stiggins, R. J. (2001). *Student-involved classroom assessment*. (3rd ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Stumpo, V. M. (1997). Three-tier grading sharpens student assessment. *Education Digest*, 63(4), 51-54.
- Thorndike, R. M. (1997). *Measurement and evaluation*, (6th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.

Tombari, M. & Borich, G. (1999). *Authentic assessment in the classroom*. Upper Saddle River, NJ: Merrill/Prentice Hall.

Tomlinson, A. T. (1994). The easy lie in educating kids. *Education Digest*, 60(2), 29-33.

Troug, A. L., & Friedman, S. J. (1996, April). Evaluating high school teachers' written grading policies. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.

Waltman, K. K., & Frisbie, D. A. (1994). Parent's understanding of their children's report card grades. *Applied Measurement in Education*, 7(3), 223-240.

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

Appendix A Survey Directions

Dear Teacher:

Your assistance is requested in the gathering of data for the purposes of completing a doctoral study. The information obtained will remain confidential and anonymous. The data collected will be used, aggregated with data from other central Iowa Algebra I classrooms, in the analysis of factors that influence grading. Please read, sign, and return the attached consent form along with the completed chart below. Your time in completing this chart is greatly appreciated. To show that appreciation, the flash drive that is being given to you containing the survey directions, survey instrument, and consent forms may be kept as a gift for your participation in the study.

Directions

Please do not indicate your name, nor the name of your school anywhere on the survey instrument. For two of the sections of Algebra I that you teach, please fill out completely the survey instrument attached. Once completed, please email the survey instrument to my research assistant, Bridget Arrasmith at bridget.arrasmith@drake.edu

Student ID – please do not identify students by name on this survey. Rather, please use the coding system provided by the researcher. These ID codes and processes will ensure student information is anonymous. You will need to remember which student corresponds to each code, so as to ensure accurate reporting of data for each variable.

Final Semester Grade – please indicate the final 2nd semester Algebra I grade for the identified student.

Teacher Perception of Effort – please use the Likert-type scale below to rank each student on your perception of the effort they put into the 2nd semester of Algebra I.

Very little / no effort	Below average (poor) effort	Average effort	Above average (good) effort	Excellent Effort
1	2	3	4	5

Teacher Perception of Attitude Toward Subject – please use the Likert-type scale below to rank each student on your perception of the student's attitude toward Algebra I.

Extreme dislike for Algebra	Below average (poor) attitude	Average attitude	Above average (good) attitude	Excellent attitude
1	2	3	4	5

Appendix B
Survey Instrument

Student ID	Final sem. Grade in Algebra I	Teacher perception of effort	Teacher perception of student's attitude towards subject
S001			
S002			
S003			
S004			
S005			
S006			
S007			
S008			
S009			
S010			
S011			
S012			
S013			
S014			
S015			
S016			
S017			
S018			
S019			
S020			
S021			
S022			
S023			
S024			
S025			
S026			
S027			
S028			
S029			
S030			
S031			
S032			
S033			
S034			
S035			

Appendix C Informed Consent Form

Research Subject Informed Consent Form

This study is a research based project designed to gain information on the grading practices of local secondary teachers. Information will be collected on the relationship between two dependent variables and the final semester grade students earn in Algebra I. The researcher is completing this study as a doctoral candidate at Drake University.

You have been selected as a potential participant of this study based upon a recommendation given to the researcher from your school administrator. You have been chosen in part due to you having taught in a high school setting for a minimum of three years and due to your teaching of Algebra I.

In order to collect data for this research project, a survey instrument will be completed. Data will be collected on students' final semester grades in Algebra I, teachers' perceptions of student effort, and teachers' perception of student attitude toward Algebra. The completed survey data will be sent to a graduate assistant at Drake University where data will be aggregated with other data from teachers and schools. Data will be kept secure at Drake University. Once aggregated to ensure anonymity, the data will be stored on the researcher's password protected laptop computer. The project will be completed by approximately June 1st, 2012 at which time both electronic and hard copies of the data will be disposed of. Copies of these transcripts, and or the final project, will be made available to participating subjects and their school districts upon request.

Benefits of participation in the study include access to the results of the study. These results and the body of literature gathered for this project may be beneficial in helping teachers, schools, and districts in an examination of grading practices based upon researched best practices. The risk of the participant in both social and professional arenas is very low. Because all data will be aggregated, and measures will be taken to ensure no student, teacher, building, or district can be identified individually, there is virtually no risk to any subject of the study to be identified. Because no personal information is being collected, no risk to student or teacher reputation or job status exists. By using anonymous student codes, no student will be able to be identified. By asking participating teachers to email their results to program assistant Bridget Arrasmith, no individual teacher, building or district will be able to be identified by the researcher. Mrs. Arrasmith will aggregate all data before sending a new electronic file to the researcher, ensuring the researcher does not know which teacher, school, or district submitted which data. Questions about this research project can be directed to Mr. Jeff Hawkins, (515) 289-1894, or the Drake Institutional Review Board at 271-3472.

Your involvement in this research project is voluntary and you should not feel compelled in any manner to participate. You have the right to withdraw your participation from this

project at any time during the project. If you would choose to withdraw from participation, no penalty or sanction will apply and any data gathered will be discarded, deleted, or disposed of immediately.

Mr. Jeff Hawkins has discussed the above information with me and I understand the purpose of this research and my rights as a research subject. My signature below indicates my understanding of this information.

Participant signature

Date