THE PREDICTIVE RELATION OF A HIGH SCHOOL MATHEMATIC GPA TO HIGH-STAKES ASSESSMENT ACHIEVEMENT SCORES IN MATHEMATICS

by

SUZANNE M. WEST

A DISSERTATION

Presented to the Department of Educational Methodology, Policy, and Leadership
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Education

June 2013

DISSERTATION APPROVAL PAGE

Student: Suzanne M. West

Title: The Predictive Relation of a High School Mathematic GPA to High-Stakes Assessment Achievement Scores in Mathematics

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Education degree in the Department of Educational Methodology, Policy, and Leadership. This dissertation has been approved and accepted by:

David Conley	Chairperson
Keith Zvoch	Core Member
Deborah Olson	Core Member
Christopher Murray	Institutional Representative

and

Kimberly Andrews Espy	Vice President for Research and Innovation, Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2013

DISSERTATION ABSTRACT

Suzanne M. West

Doctor of Education

Department of Educational Methodology, Policy, and Leadership

June 2013

Title: The Predictive Relation of a High School Mathematic GPA to High-Stakes Assessment Achievement Scores in Mathematics

Course grades, which often include non-achievement factors such as effort and behavior and are subject to individual teacher grading philosophies, suffer from issues of unreliability. Yet, course grades continue to be utilized as a primary tool for reporting academic achievement to students and parents and are used by most colleges and universities as an admissions measure. High-stakes assessment results are also used by schools to convey student achievement, and several states now require students to pass an exam to receive a diploma. What is less clear, however, is the relation between these two measures, GPA and high-stakes assessment results.

One purpose of this study was to examine the predictive relation of mathematics GPA to student performance on high-stakes assessments. Multiple regression models were used to analyze the predictive relation between mathematics GPA and performance on the ACT and the Oregon Assessment of Knowledge and Skills (OAKS), two high-stakes assessments. In addition, the regression analyses were used to examine the influence of other student-level variables such as talented and gifted status and math courses taken prior to testing on the relation between mathematics GPA and performance on the two high-stakes assessments.

In all, 299 high school students from a single grade-level enrolled in one Oregon suburban school district participated in the study. Results indicate that GPA is a significant variable in a high-stakes assessment outcome. Additionally, results of the multiple regression reveal significant student-level effects on assessment outcomes that reduce explained common variance in both the ACT and OAKS models. Implications for practice and suggestions for future research are discussed.

CURRICULM VITAE

NAME OF AUTHOR: Suzanne M. West

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene

George Fox University, Newberg, Oregon

San Francisco State University, San Francisco, California

DEGREES AWARDED:

Doctor of Education, Educational Leadership, 2013, University of Oregon

Master of Arts in Teaching, 2007, George Fox University

Bachelor of Arts, Business Management and Organizational Leadership, 1999, George Fox University

AREAS OF SPECIAL INTEREST:

District and regional leadership with a focus on policy and organizational change

Data-driven decision-making

Teacher development and performance management

Effective communication

Effective K-12 instruction

Instructional technology integration

Adult pedagogy

PROFESSIONAL EXPERIENCE:

Director of Technology and Innovation, Sherwood School District, Sherwood, Oregon, 2012-present

Associate Principal, Laurel Ridge Middle School, Sherwood, Oregon, 2011-2012

Race to the Top, Effective Teachers and Leaders Workgroup, State of Oregon, 2009

Chalkboard, CLASS Project, Differentiated Compensation Workgroup, Sherwood School District, Sherwood, Oregon, 2008-2011

Teacher, Sherwood School District, Sherwood, Oregon, 2007-2011

Store Manager, Home Depot, Oregon, 1998-2005

West Coast Training Manager, Thrifty-Payless, Rite Aid, Wilsonville, Oregon, 1994-1998

Assistant Manager, Thrifty-Payless, California, 1986-1994

GRANTS, AWARDS, AND HONORS

Richter Scholar, 2007

ACKOWLEDGEMENTS

I wish to express sincere appreciation to Dr. David Conley for providing the supports necessary to complete this endeavor and to Dr. Terri Ward for her valuable guidance. I also thank Dr. Keith Zvoch, Dr. Christopher Murray, and Dr. Deborah Olson for challenging my thinking. Additionally, I would like to express my sincere gratitude to Dr. Keith Hollenbeck, for his kindness, and to Dr. Nancy Heapes, who provided humor and well-timed encouragement throughout my University of Oregon career.

I would be remiss if I failed to express my sincere appreciation for my cohort colleagues, Dr. Jennifer Spencer-Iiams and Dr. Sarah DeBoy, who provided friendship, encouragement, and laughter during our doctoral program and the preparation of this manuscript. Finally, I want to thank Steve Emmert, my colleague, mentor, and friend, who vicariously shared this experience with me.

To God, for His strength, to DJ, for his support and patience during my many years of graduate school, and to my mother, who believed in me.

TABLE OF CONTENTS

Chapter Page

Chapter Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

The public scrutinizes schools more closely for their performance today than at any time in recent history. Mandates and initiatives such as the 2001 reauthorization of the Elementary and Secondary Education Act (ESEA), also known as No Child Left Behind (NCLB), and the ESEA Flexibility Waivers offered by the Obama Administration increase the expectation that schools will educate and graduate students prepared for college and careers and that educators will be held accountable for doing so (No Child Left Behind, 2002; Oregon Department of Education, 2012b). To this end, governments increasingly measure schools against myriad criteria including student attendance, graduation rates, college entrance rates, and both aggregated and disaggregated student performance on high-stakes assessments. This information is public and often used to identify schools as either successful or failing; however, among teachers, students, and parents, the primary means of communicating student performance remains course grades and individual assessment scores on state assessments.

The Need for Valid and Reliable Measures of Student Performance

Because of this increased governmental scrutiny, course grades and state assessments are becoming more important, and it is necessary to understand what each measures. Teachers use grades for a variety of purposes, including the improvement of instruction and the empowerment of students. As a method for improving instruction, both formative and summative grades inform teachers about student progress so that they may adjust their teaching in accordance with student need. Formative grades are those given during learning, reporting student progress on a specific standard or unit of study, while summative grades report student performance at the conclusion of a unit or learning period (Chappuis & Chappuis, 2008). Ideally, formative scores do not count in the overall student grade. Many educators, however, have individual discretion over which grades to report and use to calculate an overall course grade. Additionally, both formative and summative grades empower students to improve their understanding of and skill with a given topic. Grades can help inform students of their current progress in relation to a standard and highlight what gaps, if any, to address.

Course grades occasionally function as a proxy for program or curricular evaluation, as some educators equate grades with program or curricular assessment, reasoning that course grades measure program outcomes, so long as course assessments and assignments align to pre-specified outcomes. In other words, if most students pass a course, the program is assumed to work; however, many educators argue that grades are not granular enough to identify areas for improvement or success within a program (Carter, 2006; Rogers, 2006; Yoshino, 2012).

Similar reasoning guides the use of grades when making status decisions, such as graduation or university admission. That is, grades function as an indication of readiness or suitability for program entrance (National Honor Society, 2013; Oregon School Activities Association, 2013; University of Oregon, 2013). High school administrators use grades as a criteria to determine eligibility to participate in sports and some clubs, such as the National Honor Society, as well as for course advancement and graduation. Universities use grades as one of several measurements when making admissions decisions. At the university level, grades are used for some status decisions such as course advancement, graduation, entrance into honor societies, and Latin honor distinctions. With course grades serving so many purposes, the reliability and validity of their use for each of these purposes continues to draw scrutiny from educators and researchers.

The Complexity of Capturing Student Performance

Due to the importance assigned to course grades by our educational institutions, educators regularly wrestle with the complexities of measuring student success. Consequently, educators commonly employ several methods of capturing the range of variables that comprise student success. In 1956, Benjamin Bloom identified three educational domains—cognitive, affective, and psychomotor—and the prerequisite knowledge and skills upon which higher-order thinking in each level depended (Bloom, 1956). Educators primarily focus on Bloom's six cognitive categories to gauge student learning: knowledge, comprehension, application, analysis, synthesis, and evaluation. Known as Bloom's Taxonomy, these categories underwent a revision in 2000 that resulted in a new taxonomy better aligned with the skills and knowledge necessary for current generations of students to succeed. Similar to the original, Bloom's Revised Taxonomy added creation as the highest cognitive level, dropped synthesis, and changed terminology to more action-oriented words, resulting in the following taxonomy: remember, understand, apply, analyze, evaluate, and create (Krathwohl, 2002). Bloom's Revised Taxonomy remains in wide use; however, other models for capturing student success compete with it.

Gardner (2006), posited the existence of seven cognitive intelligences. They are logical-mathematical, spatial, linguistic, bodily-kinesthetic, musical, interpersonal, intrapersonal, naturalistic, and existential. Gardner's theory gained popularity among educators despite criticism from researchers that Gardner did not expand the definition of intelligence but rather applied the term intelligence to what others deem ability or talent (Klein, 1997; Sternberg, 1983).

More recent models of student success focus on both academic and career variables. Conley (2013) identified four keys to college and career readiness: key cognitive strategies (think), key content knowledge (know), key learning skills and techniques (act), and key transition knowledge and skills (go). Conley advocated the construction of school frameworks that both educate and assess students in these four key areas as a more holistic method of identifying student success than unidimensional frameworks such as Bloom's Revised Taxonomy. Similarly, with the release of the Common Core State Standards (CCSS) and their emphasis on college readiness, Marzano (2011) advocated the use of a standards-based grading system aligned with the CCSS as a means for identifying and assessing student success and college readiness. Yet, regardless of the model, educators continue to rely on course grades and assessments to gauge, quantify, and report student learning.

Strengths and Limitations of Grades and Assessment Scores

Grades typically purport to measure student performance against course-specific, aggregate criteria. For example, Algebra I includes concepts such as the structure of expressions, polynomials, rational expressions, and reasoning with equations and inequalities (Common Core State Standards Initiative, 2010). Yet, high schools typically report only an aggregate grade that includes student progress on all Algebra I criteria as opposed to progress on specific Algebra I concepts or standards. Thus, parents and students know overall Algebra I performance, but not Algebra I strengths or weaknesses.

Parents and students accept this reporting method, which is why it works today and has worked for more than a hundred years. Other benefits of the current grading system include content validity (grades generally reflect taught content), the potential to empower students to improve performance, and a means for teachers to inform instruction (Schwartz & Sharpe, 2011). Yet, as noted previously, aggregate grades fail to identify individual student strengths and weaknesses precisely. Additionally, grades often comingle behavior, attitudes, and learning skills with academic content knowledge, and the criteria teachers use for assigning a specific grade varies by teachers (Brookhart, 1993; Guskey, 2009). In sum, though accepted, course grades include inherent weaknesses that limit or restrict their value or appropriateness for certain decisions.

Similarly, large-scale assessments also have their own strengths and weaknesses. Different from course grades, large-scale assessments measure student skills and knowledge of a specific content sequence and may be required for graduation or college entrance. Strengths of large-scale assessments include standardized reporting, simplifying the comparison of student performance to other students or to specific criteria, low cost, reliability, and, in the case of computer adaptive tests, the ability to adjust the difficulty or ease of test questions based on student responses. Opponents of large-scale assessments argue that these tests adversely affect the education of K-12 students, especially minority students (Au, 2007; Jones & Egley, 2004; Kearns, 2011; National Center for Fair and Open Testing, 2012). Additionally, the pressure to perform well on large-scale state tests cause some educators to cheat by previewing the test prior to administering it to prepare students, using signals to indicate right or wrong answers, and erasing incorrect answers and inserting correct answers, among other actions (Schaeffer, 2013). Finally, large-scale assessment scores may not reflect what students actually know and can do. As a single data point, a large-scale assessment score is subject to many environmental and student variables such as student health, student attitude, room temperature, and fidelity to test procedures. Thus, as with course grades, large-scale assessments include inherent limitations that require acknowledgement.

The Need to Understand the Relationship between Course Grades and Large-Scale Assessments

When reporting student achievement, schools, states, and the federal government turn to course grades and large-scale assessment scores. When course grades and scores align, for example, when a student who excels at mathematics also scores well on large-scale mathematic assessments, educators, parents, and students expend little thought about the differing measures. When course grades and scores do not align, however, questions about potential inconsistencies arise. And while each measurement has value in and of itself, combined course grades and large-scale assessments explain more variance in first-year college success than either measure alone (American College Test, 2007; Oregon Department of Education, 2011c), suggesting that the differing measures may overlap to some degree while also contributing uniquely to predictions of first-year college success.

Because of the emphasis placed on course grades and assessment scores, the validity and reliability of these measures matter; however, the complexity of capturing student performance is such that no single grading system or assessment embodies all the variables that comprise student success. Additionally, while course grades and high-stakes assessments both encompass many strengths, such as low-cost, both also include inherent weaknesses that require acknowledgement and discussion. Finally, course grades and high-stakes assessments together explain the greatest amount of variance in first-year college success, suggesting that they share predictive power and while each contributing unique predictive power. Thus, an analysis of the relation between course grades and a high-stakes assessment score may prove useful in furthering our understanding of these two measures.

Statement of Problem

The purpose of this study is to examine the predictive relation between course GPA and scores on high-stakes assessments. More specifically, the guiding questions are two-fold: (a) To what degree are subject-specific grade point averages, specifically mathematics, predictive of mathematics achievement testing outcomes? (b) To what degrees are variables such as gender and economically-disadvantaged status more or less predictive of achievement test outcomes than grades?

CHAPTER II

LITERATURE REVIEW

Because academic success has long-term implications for a student's future, there is a growing movement in the United States to scrutinize grading and assessment practices at both the state and classroom level for congruity and effectiveness. Several relevant and sometimes nebulous factors affect the relation between a student's grade and her achievement on a state assessment. The criteria a teacher uses to assign a grade is the most unstable factor in this relation, sometimes to the frustration of the teachers themselves (Stiggins, Frisbie, & Griswold, 1989).

Teachers generally agree that grades should reflect content and skill achievement, but few agree on what impact effort, behavior, or ability should have on grades (Brookhart, 1993; Finkelstein, 1913; Guskey, 2009; Randall & Engelhard, 2009b, 2010; Stiggins, et al., 1989). Additionally, while standardized tests are both a statistically valid and reliable measure of a student's knowledge and skills (Hambleton, Swaminathan, Algina, & Couldon, 1978; E. J. Mason, 2007), factors such as testing environment, test facilitator, and student attitude do affect test outcomes. Despite these confounding factors, several states require that students demonstrate proficiency via a state assessment in content areas such as mathematics to receive a diploma, and teachers continue to assert that a student's grade is a measure of student academic performance (Guskey, 2009).

Reliability and Validity of Course Grades

Grading research is quite varied, situated in an historical context and within the current high-stakes accountability reality in which schools are under a high degree of scrutiny. Given this current emphasis on accountability, it is important to examine the reliability and validity of course grades.

Historical background on course grades. More than a hundred years ago, Cattell (1905) called for the scientific study of grading systems, noting that variations in grade assignment by teachers may have significant long-term consequences for individuals. Cattell argued that grades were used to "select individuals for specific purposes" (p. 367), reflecting a common notion of the time that aptitude was innate rather than developed. To support this argument, Cattell demonstrated a correlation between academic performance and standing in society, noting that valedictorians and Phi Beta Kappa men, all of whom excelled academically, were respectively five and two times more likely to be well known in the upper echelons of American society, as measured by entry in the *Who's Who in America* annual publication. Additionally, Cattell advocated that tuition be charged in proportion to merit, with more endowment funds being granted to those "whose education is the greater service to the community" as measured by their effort in academic pursuits (p. 378).

Eight years after Cattell's publication, Finkelstein (1913) conducted a scientific study of course grades, asking two important questions about grading criteria: (a) "What are the traits, qualities or capacities that we are actually trying to measure in our marking systems?" and (b) "What method ought we to follow in measuring these capacities?" (p. 1). Finkelstein justified his investigation of these questions by noting that student marks or grades were used to make decisions about induction into honorary societies, institution or academic awards, grade-level or class advancement, graduation, and, in some cases, job placement. It is of no small consequence that after one hundred years we are asking the same questions and utilizing grades for the same purposes. After his analysis of 20,348 individual marks and of the marking system used in 66 courses at Cornell University, Finkelstein (1913) concluded that marks should be based on accomplishment, but that marks were so affected by a student's personal characteristics that marks between students and teachers "show no similarity whatsoever" (p. 80).

Finkelstein (1913) collected student mark data for the nine years spanning from 1902-1911. He disaggregated the marks into 14 categories based on the 100-point scale used by Cornell University at the time and by 17 subject areas. Categories ranged from 0-39, then in five percentage point increments to 100. A 14^{th} category, exempt, was added as it was the practice at the time for some professors to excuse students from final assessments based on their level of performance during the term. His analysis of these data showed marks skewed to the right on a 100-point, left-to-right scale.

By conducting his analyses in this way, Finkelstein was able to measure factors such as the "personal equation" in marking distributions. Finkelstein defined the personal equation as the non-achievement factors that influence a professor's assignment of a mark. For example, Finkelstein examined two sets of marks given by the same professor in two different courses. The marking distribution was nearly identical. In another example, it was shown that a specific professor moved more than a third of his students to exemption status with only 1.5% failing his courses, while yet another professor assigned failing marks to 20.3% of his students. Finkelstein was also able to determine that some colleges within the university had distinct marking patterns. For example, during the period of Finkelstein's study, not one student in the College of Mechanical Engineering failed a course, nor did any student in that college earn a mark above 94.

In his conclusions, Finkelstein (1913) noted the following: (a) between courses, the percentage of students achieving exemption status ranged from 1.5% in the lowest example and 78% in the greatest example; (b) the marks given to the same student in the same subject varied based on the instructor; (c) ratings given by examiners to students of average academic achievement ranged more than 30 percentage points; and, (d) marking distributions of professors teaching the same class radically differed. Thus, Finkelstein argued that grades were based more on the personal equation than on student performance.

About mid-century, dialogue turned toward a nation-wide assessment of student achievement. As the United States and other countries began to investigate the achievement of their students, questions about the accuracy and predictability of teacher grading practices increased.

Two studies conducted in a British Columbia school district investigated the practice of assigning student grades based on a distribution of IQ scores (Mason, 1967; Mason & British Columbia Educational Research Council, 1965). During a two-year period, the researcher examined 103 classes of grade-six students, using four achievement exams. The resulting bell curve, based on student IQs, specified a specific percent of students would receive an A, B, C, D, or F letter grade; however, Mason's research indicated that student academic performance deviated from IQ scores to the extent that grading around an achievement average, or grading on the curve, was a poor method for assigning grades. Mason further analyzed the discrepancy between grading based on an average versus teachers' grading based on their own criteria and found no significant discrepancy. He concluded that a specified grading curve was unnecessary, as teachers automatically assigned grades around an average. Mason did not speculate as to why teachers may have automatically graded around an average.

The 1960s saw an increase in tracking programs, including accelerated or enrichment classes, which presented a challenge for traditional grading practices. Davidson (1964) noted two critical problems arising from the use of traditional grading practices in this new education model: (a) unwieldy grade reporting based on ability; and, (b) an unequal distribution of low-achievers and high-achievers, which perpetuated a lack of motivation at the low end and a lack of competition at the high end. Davidson proposed that the traditional grading curve transfer instead to tracking, with the top 10 percent of students assigned to an upper-level course, and so on. Though research has shown educational environments that do not track foster greater academic performance (Lleras & Rangel, 2009; Oakes & Wells, 1998), tracking exists in several forms in today's schools which may, as noted by Davidson (1964), impact course grade assignment.

Influence of teacher perceptions and student behavior on course grades. Today, teachers regularly use a criterion reference rather than a curve to determine student academic performance. Criterion-referenced assessments measure student performance against a standard or criterion rather than against their classmates, (Hambleton, et al., 1978), as was once done with a curve and as is still done with state or national normative assessments. However, teacher perceptions of students play an important role in student academic performance, as do both teacher and student behaviors, and these factors do influence course grades.

Rosenthal and Jacobson (Rosenthal & Jacobson, 1968) designed an experiment to test the effect of teacher perceptions on student achievement. For the purposes of their experiment, the researchers defined achievement as intellectual growth. The treatment group consisted of 65 randomly selected students in grades one through six in one elementary school. The researchers told the teachers of these 65 students that the students were “growth spurters” based on students’ results on the Harvard Test of Inflected Acquisition. In one school year, students in the treatment group gained 3.80 more IQ points than their peers in the control group, with the greatest gain-discrepancy in grade 1 (15.4 more IQ-point gain in the treatment group). Disaggregating the data for gender minority status revealed similar results; students in the treatment groups achieved greater IQ gain. The researchers concluded that teacher perception provided the impetus for the treatment group to experience greater IQ gain. Specifically, “one person’s expectations of another’s behavior may come to serve as a self-fulfilling prophecy” (Rosenthal & Jacobson, 1968, p. 20).

Randall and Engelhard (2009a, 2009b, 2010) researched the grading practices of teachers in a large, southeastern school district as they related to student variables such as behavior and effort. The study included 516 teachers in elementary, middle and high schools. Teachers completed a 54-question survey designed using Guttman’s Mapping Sentences. The survey combined the variables of ability, achievement, behavior, and effort with a specified level of mastery of course objectives. Teachers assigned these fictitious students a class grade ranging from A to F, with the option to assign a plus or minus.

The results showed that teachers in this study rewarded students who behaved well and worked hard with higher grades, regardless of academic achievement; these teachers also tended to bump borderline students up one grade level (e.g. C+/B- student receives the B-) (Randall & Engelhard, 2010). In fact, when viewing the data in graphic form, it was clear that although teachers assigned grades to low, average and high achieving students in roughly the bottom, middle, and upper thirds respectively, the actual grades themselves fluctuated significantly within those ranges. The data, when disaggregated by teaching levels (elementary or middle), also showed that elementary teachers consistently awarded higher grades for the same student scenario than their middle school peers (Randall & Engelhard, 2009a). In sum, current grading practices relied on teacher perceptions and weighed student characteristics as heavily as the grading practices of one hundred years ago.

An older and still relevant summary of teacher effectiveness research underscores the relation between teaching practices and student grades. Doyle (1977) examined teacher effectiveness through the lens of process-product research. Within this framework, the relation between teacher behaviors (process) and student outcomes (products) relates teacher variables directly to effectiveness. For example, Doyle posits that students with a teacher who utilizes higher-order thinking during class, but expects only recall on tests, will eventually differentiate their attention to focus only on recall information. In other words, students adjust their behavior to meet teacher expectations in order to earn an expected grade. Doyle characterizes this concept of performance-grade exchange as complicated, with much noise in the space between teacher behaviors and student outcomes.

Brookhart (1993) attempted to identify some of the noise in performance-grade exchange by examining teacher perceptions of score validity through a survey of student scenarios that included variables of effort/ability, missing work, and improvement. Teachers also explained their scoring decision. Results of this study indicated that among the 84 respondents, teachers were more likely to assign a higher grade to a low-achieving student, citing reasons such as support, work quality, and social consequences (Brookhart, 1993). Furthermore, the qualitative data collected in this study indicated that teachers perceived grades as “a form of payment to students…the coin of the realm” (Brookhart, 1993, pp. 131-132), supporting Doyle’s (1977) concept of performance-grade exchange. The results of this study imply that students do earn grades based partly on performance, but that teachers are acutely aware of the value and consequences of grades and make decisions about assigning grades accordingly.

The meaning and value of grades. Given that grades are influenced by teacher perceptions and student behaviors, a framework is needed to validate their use in the context of a high-stakes accountability environment. Messick’s (1989) framework considers both the meaning of a score and its value implications (e.g., how a grade is used) as critical matters when interpreting the validity of assessment results. Although Messick was referring specifically to assessments, his framework provides a structure for discussing the validity of grades as well, particularly because points earned on course assessments are regularly used as part of course grade computation. Validity may be considered in terms of a four-box matrix, with use and interpretation on the x-axis and evidential basis and consequential basis along the y-axis. As noted by Messick (1989) the implication of such a matrix is that all four components, use and interpretation and evidential and consequential basis, are intertwined and unified in the validation process. As applied to grading practices, a unified process of validity results in four considerations that go into every grade assigned to a student: (a) the meaning of the grade, (b) the meaning of the grade for a specific student or grade interpretation, (c) the value of the grade, and (d) the consequence of the grade (Brookhart, 1993).

Meaning and consequences are important components of validity (Messick, 1989). Guskey (2009) asked 556 teachers to identify the purpose of grade reporting, using a survey with 29 Likert-type scale items. Of the respondents, 36% indicated that they believed grades have a value as punishment (Guskey, 2009), while 74% indicated that student achievement was the primary purpose for reporting grades. Grades are often the only measure and communication between teachers and parents about student progress, so teacher perceptions of the purpose of grade reporting is a valid investigation. Questions embedded within these scales measured teacher perceptions of several grading policies and practices. Teachers were remarkably similar in agreement (98%) that assessments are informative, should align with teaching, and that students should know up front the criteria for grading (Guskey, 2009). On questions that asked for teacher perceptions of specific grading practices, teachers were evenly split. For example, about half the respondents agreed with averaging grades, considering only current evidence when assigning grades, the use of progress scales rather than letter grades or percentages, and including homework in overall grades (Guskey, 2009).

Grade inflation and between-school grade reporting. Yet another factor that affects grade validity is grade inflation. Grade inflation is the phenomenon associated with rising high school GPA without a corresponding rise in assessment achievement.

The American College Test (ACT) sponsored a study that showed that between the years 1991 and 2003, the average amount of grade inflation was about .25 (I. American College Test, 2005; Woodruff & Ziomek, 2004). In other words, grades rose about one-quarter of one grade-point (on a four-point scale) during a 13-year span without a corresponding increase in ACT performance. Grades rose across all subject areas tracked by the ACT, with the greatest increase occurring in English 12 (.30 grade points) and the smallest increase occurring in Calculus (.09 grade points). Such results lend credence to the argument that factors other than academic performance contribute to grade reporting.

A difference in grade reporting criteria also exists between schools. Using reading and math scores from the NELS:88 data collection, the Office of Educational Improvement (1994) conducted a study that showed that for students in high-poverty schools, schools in which 75% or more students qualified for free or reduced lunch, an "A" in English was the equivalent of a "C" or "D" in affluent schools. Similar results were seen in math; students in high-poverty schools who received "A" grades performed at the same level as students in affluent schools who received "D" grades. In such circumstances an "A" clearly does not carry the same meaning from school to school.

Grades as predictors of high school and first-year college success. Grade reporting is fraught with unreliability in terms of reporting criteria. In other words, course grades measure factors other than academic performance. Despite this lack of consistency in criteria for determining grades, grade reporting has been shown to be a reliable predictor of student outcomes such as dropping out, graduating, and first-year college success (Bowers, 2010b; Geiser & Studley, 2002; Lekholm & Cliffordson, 2008). As early as grade seven, GPA reliably predicts dropping out and graduation (Bowers, 2010b). In his study, Bowers (2010b) examined longitudinal data of 193 students who began grade 1 in 1994 and were on track to graduate from grade 12 in 2006. Bowers' results show that drop-out risk presents in GPA data in grade 7 and was highest for students with low grades.

In a study of 77,893 students in the University of California system, Geiser and Studley (2002) examined the predictive relation between high-school GPA and the SAT and first-year college success. High school GPA consistently exhibited the greatest predictive value in determining first-year university success. Interestingly, Geiser and Santelices (2007) conducted a follow-up study that examined the predictive value of high-school GPA in determining four-year college success. High school GPA retained the greatest predictive power across all disciplines.

The predictive value of high-school GPA crosses ethnic lines, as reported by Zwick and Sklar (2005), who determined that while both GPA and SAT scores were predictive of first-year college success among the 14,825 ethnic and language minority study participants, GPA was a stronger predictor. Additionally, the predictive relation between GPA and first-year college success holds true in at least one other nation, Sweden, as reported in a study of 164,106 Swedish students entering college during the years 1993-2001 (Cliffordson, 2008).

In sum, GPA is reliable predictor of first-year college success, as well as high school graduation and the likelihood of dropping out. GPA also measures more than just academic performance; GPA includes student behaviors and teacher perceptions that affect validity as well as the meaning and value of the measure.

The Nature of Mathematics in High School

Grades need to be considered, however, not only from a pure measurement issue but also in relation to their application in content subject areas. All classes in a high school curriculum report grades and often do so within a specific policy context.

Articulation of a mathematics curriculum. The teaching of mathematics is largely a matter of belief; that is, those things that a teacher believes to be true about mathematics will be taught to students (Beswick, 2012; Buckley, 2010). Beliefs and knowledge about math pedagogy are largely indistinguishable (Ball, Thames, & Phelps, 2008), so official articulation of a mathematics curriculum often becomes more a function of course sequence rather than the articulation of skills and concepts.

Many states have created a type of math articulation through the mandates of diploma requirements. Oregon, California, and Nevada, for example, require that a student complete three years of mathematics at the level of Algebra I or above to receive a diploma (California Department of Education, 2012; Nevada Department of Education, 2012; Oregon Department of Education, 2012a). A typical sequence beginning with Algebra I would progress to Geometry, and then to Algebra II or Trigonometry. Beyond these levels, students might take Analytical Geometry, Calculus, or Statistics. Post-geometry, the mathematics sequence a student follows is largely her own choice, depending on the courses offered by her school.

Yet, establishing a high-school math sequence is not the same as articulating an actual curriculum. To that end, states have written or adopted content standards articulated from grades K through high school with the goal of ensuring that teachers build on learned skills or concepts from previous levels. Some school districts emphasize specific state standards, commonly known as power or priority standards, those standards a school district commits to teach deeply.

Mathematics content standards. State standards have been the primary means of determining curriculum articulation for many decades, and states have largely based their standard articulation on traditional practices of individual school districts. For example the kindergarten through grade eight articulation proposed by the Scottsdale Public Schools in 1973 included many of the same content strands as the Common Core State Standards (CCSS) of 2010 (Common Core State Standards Initiative, 2010; Scottsdale Public Schools, 1973). The modern articulation that we see in the Scottsdale model and the CCSS is largely post-World War II. The late nineteenth and early twentieth century embodied the philosophy that advanced mathematics such as algebra were unnecessary for the majority of students (Bidwell & Clason, 1970).

In 1989, the National Council of Teachers of Mathematics (NCTM) published their first version of national math standards. The timing of the NCTM publication matched the movement among states and the federal government to hold schools to greater standards of accountability. The movement toward state standards was largely the result of the need to establish state funding rationale and compare student achievement across districts. Although the federal government's education focus has historically been on equity, states pushed schools to demonstrate quality as measured by more rigorous graduation and teacher licensing requirements, as well as more students enrolled in academic courses (Conley, 2003). States chose education standards as their method of accountability, and by 1999 all but one state had some form of state standards and all but two had tests to measure student learning against these criteria (Epstein, 2004). In 2001, President Bush presided over the reauthorization of the Elementary and Secondary Education Act, also known as No Child Left Behind (NCLB). With this legislation, accountability requirements tied to state standards increased significantly (No Child Left Behind, 2002). Although the effectiveness of NCLB's accountability criteria may be debated, President Bush's successor, President Obama, continues the push for accountability tied to standards (United States Department of Education, 2010).

The CCSS were developed by a consortium of governors and state education leaders from 48 states, two territories, and the District of Columbia. Over the course of a year, consortium participants wrote mathematics and English standards that incorporate research and rigor and, according to the Common Core State Standards Initiative (2010), will prepare students for college and careers. In 2011, Oregon adopted the CCSS for mathematics. By doing so, the State essentially dictated an articulation for Kindergarten through grade 12.

The CCSS currently drive the development and adoption of new large-scale tests such as the Smarter Balance Assessment adopted by the majority of states. Organized around mathematical practices and content similar to the practices and content advocated by the current version of the NCTM standards, the CCSS are written in such a way that students are progressively challenged to acquire more sophisticated mathematical skill and knowledge as they advance through the grades. For example, the CCSS state that a Kindergarten student must master counting, but students do not delve into statistics and probability until grade six (Common Core State Standards Initiative, 2010). Beginning at grade six, the organizing content changes to include more sophisticated mathematical skills and concepts such as probability and statistics (Table 2.1). Only one content frame, geometry, spans all grade levels. All other content is taught at developmentally appropriate and internationally comparable grade levels (Common Core State Standards Initiative, 2010). The CCSS are representative of a curriculum articulation (see Table 2.1), but do not represent the current practice in Oregon and many other states, as the transition to the CCSS will not be assessed using new assessments until the school year 2014-2015.

Table 2.1
Common Core State Standards Mathematics Articulation

	Grade Level									
	K	1	2	3	4	5	6	7	8	H
Counting and Cardinality	X									
Operations and Algebraic Thinking	X	X	X	X	X	X				
Number Operations in Base Ten	X	X	X	X	X	X				
Number and Operations - Fractions				X	X	X				
Measurement and Data	X	X	X	X	X	X				
Geometry	X	X	X	X	X	X	X	X	X	X
Ratios/Proportional Relationships							X	X		
The Number System							X	X	X	
Expressions and Equations							X	X	X	
Statistics and Probability							X	X	X	X
Functions									X	X
Number and Quantity										X
Algebra										X
Modeling										X

Table 2.1: Mathematics articulation based on the Common Core State Standards for Mathematics, (2010), published by the Common Core State Standards Initiative.

Thus, while individual high schools and school districts may determine a specific sequence of math courses for students to complete, state mandates, such as standards and graduation requirements, impose specific articulation requirements. Schools and districts that fail to respond to these impositions face an environment of increasing accountability in the form of new high-stakes assessments tied directly to state mandates.

Large-Scale Testing Programs and High-Stakes Accountability

Large-scale testing programs involve high-stakes. Students who meet benchmark scores or score in the top percentiles on large-scale tests may be accepted into advanced courses, be awarded high school diplomas, and gain entry into prestigious universities. Because of the high-stakes involved in large-scale tests, ensuring reliability and minimizing testing problems are critical.

Purpose of large-scale testing programs. In 1969, Congress authorized the National Assessment of Educational Progress (NAEP), thereby federalizing the large-scale testing era in the United States. With school districts spread over 50 states, the federal government needed a mechanism to provide data about the educational progress of American school children. Originally designed to report only aggregate data about student academic performance, today NAEP reports not only aggregate, national data, but also disaggregated data that include ethnic and gender subgroups, as well as state-level data and student characteristics such as studying habits (National Center for Education Statistics, 2005).

The purpose of large-scale testing is three-fold: (a) report student progress; (b) report student achievement; and (c) hold states, districts, and schools accountable for student progress. It was this purpose which drove Congress to endorse the NCLB requirement that states must test all students in reading, writing, and mathematics at specific grade levels, and to penalize schools that failed to demonstrate adequate annual student progress (No Child Left Behind, 2002). It is this purpose that is driving many states to adopt high-stake central assessments as the ultimate criteria for high school graduation.

Research suggests, that students in countries and states that require high-stakes central assessments perform academically better than their peers in countries and states without this requirement. A study conducted by Bishop (2001) using 1995 data from the Third International Mathematics and Science Study (TIMMS) showed that countries that required a high-stakes assessment for graduation outperformed the United States at least one grade level in mathematics and science. Bishop (2001) notes that data from the International Assessment of Educational Progress and the International Association for the Evaluation of Educational Achievement support this same or larger margin of academic achievement between countries that employ rigorous central assessments and those that do not. Bishop did not find that minimum-competency assessments yielded similar results.

The key difference noted is that rigorous central assessments, such as the New York Regents Exam, require students to demonstrate a broad and deep range of knowledge across multiple subject areas, while minimum-competency assessments are not as difficult and generally assess only a few key areas, such as mathematics, reading, and writing. The study also determined that those nations that employ high-stakes central assessments generally have higher standards and salaries for teachers, and that teachers generally have a degree in the subject area they teach. Such differences may actually reveal a symbiotic relation between central assessments and greater teacher quality.

Woessmann (2002) supports Bishop's assertion that high-stakes central assessments improve student achievement. He found that in countries that administer central assessments, academic performance, regardless of family background, is statistically greater than that of students in countries without centralized assessments. Noting that test quality and a purposeful focus on test content, or "teaching towards the test" (Woessmann, 2002, p. 40), may influence test outcomes, Woessmann adds that centralized testing may be particularly beneficial in a decentralized educational system such as that in the United States, as it provides uniform data about student performance across states.

Local data also support the adoption of high-stakes central assessments, though there are inconsistencies. A comparison of stringency measures implemented by states in 2003 as a result of NCLB, such as a limited number of test retakes and increased progress timelines, against 2003 and 2005 mathematics and reading NAEP results revealed that these measures correlated to improved grade eight reading and mathematics achievement for White and Hispanic students, with the greatest gains achieved in mathematics (Wei, 2012). These same results, however, were not seen in Black grade eight students. Wei concluded that state accountability policies tied to central assessments vary in effectiveness, depending on the group and subject.

Validity issues in high-stakes assessments. While centralized assessments at a national or state level may lead to improvements in student achievement, such testing programs are not exempt from issues. For the 2011-2012 school year, 25 states required an exit assessment in order to receive a high school diploma (McIntosh, 2012), making exit assessments very high-stakes. While the idea of minimum competencies in subject areas such as mathematics may set high expectations for students, several issues arise as a result. One oft-overlooked problem is the “shame, humiliation, and embarrassment” that results from failing to pass a high-stakes test (Kearns, 2011, p. 126). A qualitative study of 16 students who had failed a central literacy assessment conducted by Kearns (2011), showed that marginalized students, those who performed poorly on high-stakes assessments, tended to fail more frequently than their peers who experienced regular success with assessments. Compounded over several efforts, students may cease caring or trying which skews test results, causes additional failures, and perpetuates the cycle.

When examining the cycle of failure indicated by Kearns, it is important to note that seven out of ten students attend schools in states with exit exams, and that within this group, a disproportionate number of students are ethnic or racial minorities or are economically disadvantaged. Specifically, of the 69% of students attending school in states requiring exit exams, 71% of these students were African American, 85% were Hispanic, 83% were English language learners, and 71% were considered economically disadvantaged by their eligibility for free or reduced meals (McIntosh, 2012). In other words, the comorbidity of student-level variables such as first language, ethnicity, and poverty may affect exam outcomes. Additionally, 19 of the 25 states that require exit exams also use these exams for NCLB reporting (McIntosh, 2012). Oregon is one such state. The Oregon Assessment of Knowledge and Skills (OAKS) functions as both an exit exam and an NCLB reporting measure, which raises the stakes on test outcomes for students, who must pass to graduate, and administrators and teachers, who are accountable for student outcomes.

Research suggests that general intelligence may also influence outcomes on high-stake assessments (Frey & Detterman, 2004; Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012). Kaufman, et al. (2012) examined 4,969 test results from two assessments, the Woodcock-Johnson III and the Kaufman Test of Educational Achievement- 2nd Edition. Results showed a mean correlation coefficient of .83 between the two assessments, suggesting a strong relation between achievement and intelligence. Frey and Detterman (2004) conducted two similar studies in which Scholastic Aptitude Test (SAT) results were correlated with IQ test results. In the first study of 917 subjects, the SAT strongly correlated ($r = .82$) with the Armed Services Vocational Aptitude Battery. In the second study of 104 subjects, the SAT was moderately correlated ($r = .483$) with the Raven's Advanced Progressive Matrices. The researchers conclude that such correlations indicate that the SAT is largely a test of IQ.

Results from high-stakes mathematics assessments may also be impacted by language interference experienced by many English language learners. Wright and Li (2008), in a qualitative study conducted in Texas with English language learners, determined that the language demands of the state assessment of mathematics far exceeded the functional math demands. In other words, because of the language difficulty, the Texas state test was more of a language test than a math test. Although in many states such as Oregon, math tests may be read aloud to students, the academic vocabulary used in those tests may exceed the academic vocabulary developed by English language learners (Wright & Li, 2008). An inference may be made from these findings that any student who is not reading at grade level or has not acquired grade-level vocabulary knowledge may experience similar difficulties.

Those in favor of high-stake testing argue that the accountability of testing motivates teachers to improve student outcomes (Wang, Beckett, & Brown, 2006). However, Marchant, Paulson, and Shunk (2006) examined National Assessment of Education Progress (NAEP) scores in three testing periods (1992, 1996, and 2000), searching for a link between large-scale assessments and increased achievement. The researchers controlled for demographic variables such as family income, education level of parents, and ethnicity/race, and then completed a regression analysis to determine the effect each demographic variable had on test results. The results revealed that, specific to math, achievement was not reliant on the use of large-scale testing. Rather, socio-economic and ethnicity/racial status was a greater determinant of achievement. In other words, only slight differences in demographics between states caused a significant difference in NAEP achievement at the aggregate level.

Analyses consistently report a stronger correlation between GPA and first year-college success than between SAT or ACT scores and first-year college success. In a study conducted by Mattern et al. (2011), data on 150,377 students from 110 colleges and universities were analyzed to identify whether GPA or SAT scores were the stronger predictor of first-year college success. The results indicated that, for students whose high school GPAs identified them as higher-performing than their SAT results indicated, using high school GPAs alone overpredicted first-year college success. In contrast using SAT results alone under-predicted first-year college success for those students whose SAT results indicated higher performance than their high school GPAs suggested. Minorities, females, non-native English speakers, and students from low socioeconomic backgrounds were more likely to perform better when measured by high school GPAs as compared to

SAT results. These findings support the idea posited by many researchers (Camara & Echternacht, 2000; Geiser & Santelices, 2007; Geiser & Studley, 2002) that high-stakes assessments may not be the best determinant of college-performance when used in isolation of other variables, such as GPA.

In sum, high-stakes and large-scale assessments provide state and federal governments with data to identify successful and failing schools, as well as successful and failing population subgroups. These data, however, may not be valid for subgroups when factors such as disaffectedness, first language, poverty, and ethnicity and race are considered. Additionally, high-stake assessments may function as proxies for IQ tests, which raise questions about their purpose and the consequences of outcomes. Finally, high-stakes assessment scores consistently have less predictive value than GPA for determining first-year college success, at least for some sub-groups of students.

Problem Proposition in Using Mathematics Grades as Predictors of Achievement

With grades positioned in a school curriculum and policy environment and being applied in the context of high-stakes accountability systems, it is important to understand both their meaning and value. GPA is still an important part of determining class valedictorians and college entrance, yet GPAs may vary significantly in what they represent. For example, a 3.2 cumulative mathematics GPA of a student who has taken geometry, trigonometry, and discrete mathematics represents higher mathematics achievement than a 3.2 cumulative mathematics GPA of a student who has taken Introduction to Mathematics, Foundations for Algebra, and Algebra I. Weighted GPAs may also skew cumulative mathematics GPAs. Specifically, some schools assign a greater point value to advanced classes. Consequently, an A grade in AP Calculus may be weighted at five points rather than four, thus elevating a student's cumulative GPA. Lastly, GPA is still largely unstable, as teacher grading may be based on student behaviors and teacher perceptions as well as on student performance.

Despite the problems inherent in using GPA as an analysis variable, schools and teachers continue to use grades and GPA as an indicator of student success, the assumption being that students with an "A" mathematics grade, or a 4.0 mathematics GPA, have mastered mathematics content. Thus, a prediction may be made that students who demonstrate classroom content mastery should also demonstrate achievement testing mastery.

This study aims to contribute to existing literature by examining two questions: (a) Are subject-specific grade point averages, specifically mathematics, predictive of mathematics achievement testing outcomes? And (b) Are variables such as gender and economically disadvantaged status more or less predictive of achievement test outcomes than grades? Because educators continue to tell students and parents that both grades and assessment scores are indicators of student learning, it is important to understand the relation between the two.

CHAPTER III

METHODS

The purpose of this study was to examine the value of GPA for predicting performance on a high-stakes assessment. To conduct this investigation, I collected and analyzed assessment data from a 2012 graduating grade-level and corresponding cumulative GPA data from a school district in a northwestern state. The school district was demographically and geographically representative of many rural and suburban northwestern school districts in that it was largely ethnically and racially homogenous. More specifically, this study was guided by the following two objectives: (a) to test the hypothesis that GPA is not a significant predictor of an outcome on high-stakes assessments; and (b) to examine the degree to which student factors other than GPA predict student performance on high-stakes assessments.

Participants and District Context

The sample originally included 334 students from Maple School District (pseudonym) within a single grade level who graduated in June 2012. The 334 students included in this sample were selected because they were enrolled at the time the data set was generated. Consequently, students who may have dropped out prior to the 2011-2012 school year are not identified, nor are those students who transferred to Maple High School late in their freshman, sophomore, or junior year.

This district was chosen for its willingness to share student data. During the 2010-2011 school year, the Maple School District met annual yearly progress (AYP) in every category except student mathematics achievement for the subgroup Students with Disabilities. Despite this, all seven schools in the Maple School District were rated

Outstanding by the Oregon Department of Education, and students at Maple High School outperformed their state peers in every measurement, including mathematics (Oregon Department of Education, 2011a).

To ensure the viability of the data set, a thorough examination the data set contents was completed. Students with reported assessment scores but no reported mathematics courses were excluded from analyses, as were students who completed the ACT exam prior to their OAKS assessment, or students who were missing both ACT and OAKS scores, resulting in a final sample size of 299. Mathematics GPA was computed using only those courses completed prior to the reported Oregon Assessment of Knowledge and Skills (OAKS) testing date. Performance on the American College Test (ACT) administered either concurrently or the year following the administration of the OAKS was also reported for these students. Both males and females were included in the study, and additional demographic information such as mathematics trajectory (track), learning disability status, and minority status were included as control variables.

Maple School District is similar to many of the rural and suburban districts in Oregon, and to the state itself, in that it is largely homogenous, but differs in its socio-economic status (SES) and in the small number of limited English proficient students (Oregon Department of Education, 2011b; U.S. Census, 2011). Located in a small but growing suburb, the city of Maple is 4.5 miles in circumference, but draws students from surrounding rural areas that double the school boundary.

The percentage of students who receive special education services is lower in the Maple School District (7.4%) than in the State of Oregon as a whole (13.2%) (Oregon Department of Education, 2011b). Maple also supported 5% of the sample population through a 504 Plan, which is a plan that delineates general education accommodations due to medical conditions such as attention deficit disorder. Finally, within the sample population, 5.7% of the students were identified as talented and gifted compared to 7.3% for the state (Oregon Department of Education, 2011b). Both females (49%) and males (51%) were included in this study.

Academic variables are at the core of this study, and an important factor in overall math performance is the type of mathematics courses taken prior to achievement testing. Thus, students were classified as being on one of three different tracks, based on the math courses in which they were enrolled: *remedial, standard,* or *accelerated.* As previously noted, several states require students to complete at least three mathematics courses beginning at Algebra I or higher. Using this framework as a guide, students whose first mathematics course was a course preceding Algebra I were classified as being on a *remedial* track; those whose first course was Algebra I were classified as being on a *standard* track, and those whose first course was more advanced than Algebra I were classified as being on an *accelerated* track. As might be expected, the majority of participants in this study began high school on a standard track (52.5%).

The district also supplied attendance and discipline data for this sample of students. Attendance was poorest while the students were freshmen, with students attending only 89% of scheduled class days. These numbers improved as the sample moved into their sophomore and junior years, with attendance climbing to 94% in both of these years. Discipline problems were relatively minor, with fewer than 10 major discipline issues among the entire sample throughout their freshman, sophomore, and junior years.

Study Measures: GPA, ACT, and OAKS

I analyzed three performance measures in this study: high school mathematics GPA, individual student scores on the high school mathematics OAKS, and individual student scores on the mathematics portion of the ACT. Information about GPA calculations and the technical adequacy of OAKS and ACT follow. In addition, I provide information about the operationalization of high school mathematics GPA in the district where the study was set, which has an indirect impact on both GPA and an outcome on the high-stakes assessment.

Grade point average calculations. As noted previously, GPA is an average of grades assigned for a term, a content area, a grade-level, or cumulated over an entire high school career. Grades are generally assigned at the completion of assignments, quizzes, and tests, which are all forms of classroom-based assessments. These assessment grades are typically averaged to determine a term grade assigned at the conclusion of each semester, trimester, or quarter. Occasionally, courses may be assigned a Pass or Fail grade rather than an A-F grade. Pass or Fail grades are not included in a GPA calculation. Factors that influence grades include student achievement, as well as ability, behavior, and effort (McMillan, 2001; Randall & Engelhard, 2010). Including factors other than student achievement has resulted in a challenging variable to interpret; however, grades are still the widest-used tool for reporting student academic achievement. Additionally, grades are predictive of life outcomes such as dropout rate and college success (Bowers, 2010a, 2010b; Cliffordson, 2008).

The manner in which a school or school district operationalizes mathematics may impact a student's overall mathematics GPA. Specifically, the exact nature of the math curriculum, the timing and duration of mathematics classes, and the ending and starting point are left to the discretion of non-state entities, so long as those entities adhere to the state standards and graduation requirements.

To illustrate this point, a comparison between two Oregon school districts may be made, Maple, included in this study, and Spruce (pseudonym) a large metropolitan district in Oregon. In the Maple School District, the high school organizes its classes into trimesters of approximately 58 days, and each class is approximately 70 minutes in length. A core class, such as Algebra I, requires two trimesters to complete. In other words, a Maple student may complete Algebra I in 8,120 minutes (7200 minutes is the minimum required for one credit hour). In contrast, Spruce High School, one of eight Spruce School District high schools, organizes its classes into semesters of approximately 95 days, and each class is approximately 50 minutes in length. A core class, such as Algebra I, requires two semesters to complete. Thus, a Spruce High School student may complete Algebra I in 9,500 minutes. In this example, Spruce High School students spend 1,380 more minutes in Algebra I than their counterparts in Maple, but both classes satisfy the state graduation requirements.

The sequence in which math courses are taken at the high school level is another decision left largely to individual districts. At Spruce High School, for example, students must follow the sequence of Algebra I, Geometry, Algebra II, then College Algebra before taking Trigonometry. The exception to this flow is that a student admitted to Honors Algebra II may skip directly to Trigonometry. At Maple High School, students must also take Algebra I before Geometry; however, after successfully completing Geometry, students may take Algebra II, Discrete Math, or Probability and Statistics.

Algebra II is required before trigonometry. Both sequences fulfill the Oregon diploma requirements which stipulate that a student must take three or more credits of high school math starting with Algebra I (Oregon Department of Education, 2012a); however, varied course sequence may result in significantly different high school mathematics experiences which may affect outcomes on high-stakes assessments.

Individual mathematics departments or districts determine mathematics curriculum adoption based on approved state math curricula. For example, in Oregon, mathematics teachers in grades 9-12 may select from eight different math curricula. Additionally, districts may choose to design their own curricula or choose a source other than that approved by the state. In other words, while students at both Maple and Spruce High Schools may take Algebra I, the exact nature of the curricula may vary significantly.

Reliability and validity of large-scale tests. Large-scale and high-stakes tests require a certain level of reliability and validity to ensure fair and accurate reporting and accountability. An assessment is unreliable if it cannot be generalized. An assessment is invalid if it either does not actually measure that which it purports to measure, or the results of the assessment are used in a manner inconsistent with the design and purpose of the assessment.

Reliability measures estimate the consistency of results. In other words, a reliable test produces similar results when administered multiple times to the same student or student population, or when administered by a different testers. Generally, reliability coefficients are reported as *adequate* (.70 – .79), *good* (.80 – .89), or *excellent* (>.90) (George & Mallory, 2003). Reliability coefficients may be reported for alternate forms, split-half, concurrent, inter-rater, and test-retest. For example, concurrent validity coefficients for the OAKS for grade 11 mathematics have been reported as both .78 (California Achievement Test) and .82 (Northwest Evaluation Association), which places the grade 11 mathematics OAKS in the adequate to good range (Oregon Department of Education, 2007). Similarly, the American College Test (ACT) in mathematics, which is widely used as a determinant in college admission, reports a reliability coefficient of .82 (good) (American College Test, 2007).

Oregon recently increased the mathematics cut scores, or benchmarks, students are expected to achieve to better align with the CCSS. As a result, the percentage of students who met the math benchmark dropped across the state compared to previous years (Oregon Department of Education, 2011b). Although the increase in cut scores may better align with CCSS expectations, the increase also means that students who may have met the cut score two years ago, in 2009-2010, may not have met in 2010-2011.

Validity, in its simplest definition, refers to the degree to which a test measures what it purports to measure. Common aspects of validity include content, or knowledge and skills, and criterion, the extent to which a test predicts specific constructs. OAKS claims to assess Oregon mathematics standards through Algebra II. If it does not, if OAKS assesses Trigonometry or Calculus or does not assess Algebra II, then it is lacking in content validity. To ensure content validity, Oregon took the following steps in designing OAKS mathematic questions: (a) Adoption of clearly defined content standards, (b) test specifications that delineated each score reporting category; (c) test development that included the use of content experts to write and evaluate test questions; and, (d) alignment of test questions to standards (Oregon Department of Education, 2007). In conjunction, these measures increase the content validity of the mathematics OAKS.

In contrast, the ACT does not claim to measure student achievement against specific content standards, but rather focuses on general education achievement. The rationale for developers of the ACT taking this approach was that at the time the test was developed and current validity evidence gathered, all 50 states established their own content standards. A normed national test simply could not measure such a diaspora of standards. Still, the ACT took measures to ensure the content validity of its assessments by first identifying college readiness standards and then utilizing content experts who evaluated these standards for their relation to expected skills and knowledge and for "increasingly sophisticated skills and understanding across score ranges" (American College Test, 2007, p. 20). According to the ACT technical manual, for mathematics, raters agreed on the above content criteria on 95% of test items.

The high school mathematics OAKS spans three core standards: Algebra, Geometry, and Statistics (Oregon Department of Education, 2011c). Because Oregon now requires students to complete three credits of high school mathematics at the Level of Algebra I and higher (Oregon Department of Education, 2012a), most students have completed Algebra I and Geometry before taking the OAKS mathematics assessment in grade 11. Indeed, the school district in this study requires students to take Algebra I and Geometry before branching into differing strands of mathematics. Thus, a student who remains mathematically on track for the first two years of high school will have had an opportunity to learn two of the three strands assessed by the OAKS in mathematics. Similarly, the ACT measures students' knowledge of mathematics through trigonometry, but the majority of test items, 54 out of 60, focus on pre-algebra through geometry (American College Test, 2007).

Criterion validity refers to the extent to which an assessment predicts outcomes for a specific real-life situation. Oregon uses three criteria to measure content validity: (a) OAKS scores in relation to first year college performance, (b) OAKS scores in relation to pre-employment and pre-apprenticeship tests, and (c) that the assessment measures what it purports to measure. In regard to the first criterion, outcomes on the OAKS mathematics test have been reported to relate to first year college performance (Oregon University System, 2011a, 2011b). Specifically, OAKS mathematics test outcomes have shown to be a reliable predictor of Oregon University System mathematics success. Specifically, data for Maple School District show that students who *exceed* their mathematics OAKS benchmark earn higher college freshman mathematics GPAs than those students who *meet* or *do not meet*. Similarly, students who *meet* their mathematics OAKS benchmark earn higher freshman mathematics GPAs than students who *do not meet* (Oregon University System, 2011a).

At the state-level, similar results were seen when student OAKS scores were compared to pre-employment and pre-apprenticeship assessments. Students who met the OAKS benchmark scored significantly higher on these assessments than those students who did not meet (Oregon Department of Education, 2011c). Finally, when the mathematics OAKS was compared to the mathematics portion of the California Achievement Test, it scored in the *adequate* to *good* range, an indication that the OAKS measures those constructs it purports to measure (Oregon Department of Education, 2007).

The ACT also reports on studies of criterion validity against high school GPA and first-year college success. Data analysis showed a high correlation between ACT mathematics score means and high school GPA. Students with the highest GPAs scored highest, on average, on the mathematics ACT, while students with the lowest GPAs scored lowest on the mathematics ACT. A weighted regression analysis conducted by the ACT also showed that the GPAs of students who had taken advanced math, such as trigonometry or calculus, prior to testing were significant predictors in ACT mathematic achievement. Specifically, students who completed trigonometry and calculus prior to testing showed an average increase in mathematic ACT scores of 1.97 and 3.48 respectively (American College Test, 2007). The ACT may also have predictive validity in determining college success. Specifically, distributions across institutions of ACT composite scores, high school GPA and first-year college GPA showed significant relations, with the highest composite scores and high school GPAs related to the highest first-year college GPAs and the lowest composite scores and high school GPAs related to the lowest first-year college GPAs. In sum, the research established that ACT outcomes were related to high school GPA and that ACT outcomes were a valid predictor of first-year college GPAs.

The State of Oregon now requires students to pass the OAKS in reading, writing, and mathematics to receive an Oregon diploma (students may demonstrate proficiency using other measures, as well). Based on the aforementioned reliability and validity data, OAKS is both reliable and valid for its intended purpose, but OAKS was not designed to be a graduation assessment. Significant social consequences are associated with failing to graduate from high school, from reduced earning potential to increased unemployment.

Thus far, Oregon has not addressed the social consequences associated with the use of OAKS as a graduation assessment.

Procedures

This study was designed as a quantitative, crossed study of extant data; however, demographics allow for natural groupings. Correlation and regression analyses were conducted to look for possible multicollinearity issues and as a preliminary examination of the relation between high school mathematics GPA and mathematics OAKS and ACT outcomes. Extant data sets used for the analyses were collected with the permission and aid of Maple School District's assessment coordinator. Student names were coded for confidentiality.

Before the data were analyzed, they were cleaned by removing participants with null values in demographic data and course grades and then calculating GPA using the reported test date and course completion date. Specifically, courses concurrent or following the reported test were excluded from the GPA calculation. Excel was used for this calculation, importing the results into SPSS for analysis.

Dummy coding was used for all categorical variables. Dummy coding is a method of using zeros and ones to transform categorical variables into quantitative values so they may be used in regression analyses. When dummy coding, a one indicates a member of the category and a zero indicates all non-members of the category. For example, for the category gender used in this study, females were coded as zero and males were coded as one. When a categorical variable is not dichotomous, dummy coding requires *n-1* new variables, resulting in three coding combinations, (0,1), (0,0) and (1,0), with the coding (0,0) being the referent category.

Students were classified as being in a specific mathematics track based on their first reported math course. Students who began with a remedial course such as Foundations, or an extended course, such as Year-Long Algebra (rather than the standard two-trimester Algebra I course), were classified as being on a *remedial* trajectory. Students who began in Algebra 1 were classified as being on a *standard* trajectory, and students who began at Geometry or above were classified as being on an *accelerated* trajectory. These classifications were based on the requirement of several states that all students complete three mathematic credits at the Level of Algebra 1 or higher to receive a diploma (California Department of Education, 2012; Nevada Department of Education, 2012; Oregon Department of Education, 2012a). The mathematics trajectory classification was used to control for the varying levels and assortment of mathematics courses taken by participants prior to the reported test date, and *standard trajectory* functioned as the referent variable. In total, participants had completed 25 different mathematics courses.

Students were also categorized by race/ethnicity, and this variable, similar to mathematics trajectory, was not dichotomous. For purposes of this study, White was the referent category to which Hispanic and non-Hispanic, other minority were compared. All other categorical variables were dichotomous and included, (a) economically disadvantaged, (b) talented and gifted, (c) gender, (d) special education, and (e) 504 Plan. The remaining variables were continuous or quantitative and did not require dummy coding. These included, (a) math GPA, (b) major discipline issues, and (c) annual attendance rate.

Specific procedures are required for the administration of both the OAKS and the ACT. The Oregon Department of Education (ODE) requires that all OAKS test administrators receive annual training in testing procedures to ensure the validity and reliability of results. Accordingly, the Maple School District reported that it trained all test administrators to abide by ODE requirements for test preparation, test security and student confidentiality, test administration, including accommodations and modifications, and test completion. The ACT training for test administrators is similar; however, because the ACT is a paper/pencil test (the OAKS is administered via computers), the test administrator requirements also include security of test materials.

Analyses

Both correlation and regression analyses were conducted. A correlation analysis examines the strength of association between two variables. A regression analysis is used to test a predictive model for a data set of dependent and independent variables. In other words, regression determines the degree to which a given independent variable, or group of independent variables, predicts performance on a specific dependent variable.

In this study, dependent variables included scores on the OAKS and ACT mathematics assessments. Predictor variables included: (a) gender; (b) Hispanic ethnicity; (c) non-Hispanic, other minority ethnicity; (d) economically disadvantaged status; (e) talented and gifted status; (f) special education status; (g) accelerated mathematics trajectory; (h) remedial mathematics trajectory; (i) 504 Plan status; (j) attendance rate; and (k) major discipline issues. Limited English proficient status was reported, but not included in analyses as only one student was so classified.

CHAPTER IV

RESULTS

Relationships between mathematics GPA, eleven other student-level variables, and high-stakes assessment outcomes, are presented in this chapter. In the first section, descriptive statistics are presented for each student variable. In the second section, bivariate correlations are presented. In the third section, the results of two multiple regression models examining the predictive relationship of GPA and other student variables such as ethnicity and mathematics trajectory on two dependent variables, ACT and OAKS are presented.

Descriptive Statistics

Tables 4.1 through Table 4.4 present descriptive statistics. The statistics in Table 4.1 reveal that the analytic sample was more advantaged than the general population of Oregon students. Specifically, only one student (0.3%) was identified as Limited English Proficient (LEP). While that student is accounted for in Table 4.1, the student's scores are not included in Tables 4.2 through 4.4. For the remaining students, just 8.4% of the analytic sample were economically disadvantaged, 7.4% qualified for special education services and 11.1% were of Hispanic or non-Hispanic, other minority ethnicity. These percentages are below Oregon averages for the 2011-2012 school year. Specifically, Oregon reported that 13.2% of the student population qualified for special education and 33.7% of the student population identified as an ethnic minority (Oregon Department of Education, 2011b). Additionally, the majority of students (83.3%) began high school mathematics on a standard or accelerated mathematic track, 5.7% qualified for talented and gifted status, and major discipline issues were confined to only 3.0% of the sample (see Table 4.2). For the 2011-2012 school year, Maple High School never exceeded 3 referrals per day, compared with 15.08 median referrals per day nationally for similar-sized schools (School Wide Information System, 2013).

Table 4.1 also includes a comparison of those students included in the sample of the original 334 and those excluded. Students may have been excluded from the sample for several reasons: (a) a lack of both ACT and OAKS score, (b) a lack of reported math courses, and (c) an ACT test date that preceded the OAKS test date. This comparison shows that of the 35 students excluded, 31 were excluded due to a lack of course information; thus, it was not possible to accurately determine the representation for each math trajectory within the excluded population. More males (54.3%) were excluded from the study than females (45.7%), and as a percent of population, Hispanics (8.6%) and non-Hispanic, other minority (8.6%) were overrepresented in the excluded population when compared to the study population. Talented and gifted students (8.6%) were also overrepresented in the excluded population and no students with a 504 Plan were excluded from the study. Furthermore, students identified for special education (17.1%) and students with major discipline issues (40.0%) were disproportionately represented in the excluded population when compared to the sample population.

Table 4.1

Student-Level Variable Statistics Excluding Attendance Rates

Variables	Included in Sample, *N*=299		Excluded from Sample, *N*=35	
	N	Percent	*N*	Percent
Accelerated Math Trajectory	92	30.8	0*	0
Standard Math Trajectory	157	52.5	4*	11.4
Remedial Math Trajectory	50	16.7	0*	0
Female Gender	146	48.8	16	45.7
Male Gender	153	51.2	19	54.3
Talented and Gifted	17	5.7	3	8.6
504 Plan	15	5.0	0	0
Special Education	22	7.4	6	17.1
White	266	89.0	29	82.8
Hispanic	19	6.4	3	8.6
Other Minority Non-Hispanic	14	4.7	3	8.6
Limited English Proficient	1	0.3	1	0.3
Major Discipline Issues	9	3.0	14	40.0
Economically Disadvantaged	25	8.4	3	8.6

*The remaining 31 students of the 35 excluded were excluded due to a lack of course information.

Table 4.2 shows that the mean annual attendance rate was 89.77% (see Table 4.2) which fell below the state average of 91.7% for the 2011-2012 school year (Oregon Department of Education, 2011a). High schools calculate attendance rate by dividing the number of class periods available for a student to attend by the number of class periods a student did attend. Additionally, as displayed in Table 4.2, the mean score for the mathematics portion of the ACT was 20.82, placing the sample at the 51st percentile nationally (The American College Test, 2013) and the mean mathematics OAKS score was 241.52, a score that fits well within the average or *meets* range (a *meets* score for 11th-grade students on the OAKS is a score between 236 and 250). Moreover, it should also be noted that mean GPA was 2.77, the equivalent of a high "C" mark (see Table 4.2).

The sample population for the ACT score (254) and the OAKS score (297) were less than 299 as not all students completed both measures. Maple High School administers the ACT to all students during one day of the Winter term. Consequently, those students for whom no ACT score was reported were absent on the day of ACT administration. Of the 45 students who did not take the ACT, 43 were White, one was Hispanic, and one was non-Hispanic, other minority. Additionally, 14 students on a standard math trajectory did not take the ACT, and neither did 26 students on a remedial math trajectory and five students on an accelerated math trajectory. Of the 15 students on a 504 Plan included in the sample, eight did not take the ACT, and of the 22 special education students, 18 did not take the ACT. Among the population of 25 economically disadvantaged students in the sample, eight did not take the ACT. All 17 talented and gifted students completed the ACT.

The two students for whom no OAKS score was reported were both White, one was male and one was female. The male student was on a remedial math trajectory and the female student was on a standard math trajectory.

Table 4.2

Descriptive Statistics of Academic Variables, Including Attendance Rates

Variables	*N*	*M*	*SD*	Minimum	Maximum
Attendance Rate	299	89.77	19.39	53	100
ACT Score	254	20.82	4.20	12	33
OAKS Score	297	241.52	8.31	217	276
GPA	299	2.77	0.99	0	4

Table 4.3 displays an analysis of ACT scores, disaggregated by all independent variables except attendance rates and major discipline issues, revealing mean scores that ranged from a low score of 15.83 for the Remedial Math Trajectory students to a high of 26.59 for talented and gifted students (see Table 4.3). Students on an accelerated math trajectory (*M* = 24.21) outperformed those on a standard math trajectory (*M* = 19.59) and those on a remedial math trajectory (*M* = 15.83). Male students (*M* = 21.29) performed slightly better than female students (*M* = 20.35) and White students (*M* = 21.12) outperformed both Hispanic students (*M* = 18.33) and non-Hispanic, other minority students (*M* = 19.15). The mean score for students with special education status was 20.25, while students with a 504 Plan earned a mean score of 18.00 and economically

disadvantaged students earned a mean score of 18.71. Attendance rates and major discipline issues were not included in Table 4.3 as they are quantitative variables.

Table 4.3

Descriptive Statistics of Disaggregated Academic Variables for ACT Scores, N = 254

Variables	*N*	*M*	*SD*	Minimum	Maximum
Accelerated Math Trajectory	87	24.21	3.43	15	33
Standard Math Trajectory	143	19.59	3.32	12	28
Remedial Math Trajectory	24	15.83	1.81	13	21
Female Gender	127	20.35	4.02	13	29
Male Gender	127	21.29	4.34	12	33
Talented and Gifted	17	26.59	3.022	22	33
504 Plan	7	18.00	4.203	15	25
Special Education	4	20.25	8.54	15	33
White	223	21.12	4.19	12	33
Hispanic	18	18.33	3.58	13	27
Other Minority Non-Hispanic	13	19.15	3.91	14	25
Economically Disadvantaged	17	18.71	3.75	12	25

Table 4.4 presents an analysis of OAKS scores, again disaggregated by all independent variables except attendance rates and major discipline issues, revealing scores that ranged from a mean low of 231.98 for the Remedial math Trajectory students to a mean high of 255.41 for talented and gifted students (see Table 4.4). Students on an accelerated math trajectory (*M* = 247.27) outperformed those on a standard math trajectory (*M* = 241.13) and those on a remedial math trajectory (*M* = 231.98). Male students (*M* = 242.34) performed better than female students (*M* = 240.67) and White students (*M* = 242.02) outperformed both Hispanic students (*M* = 238.11) and non-Hispanic, other minority students (*M* = 236.71). The mean score for students with special education status was 233.36, while students with a 504 Plan earned a mean score of 235.33 and economically disadvantaged students earned a mean score of 239.44. Again, attendance rates and major discipline issues were not included in this analysis for the reasons stated previously.

Table 4.4

Descriptive Statistics of Disaggregated Academic Variables for OAKS Scores, N = 297

Variables	*N*	*M*	*SD*	Minimum	Maximum
Accelerated Math Trajectory	92	247.27	7.54	224	276
Standard Math Trajectory	156	241.13	6.09	220	273
Remedial Math Trajectory	49	231.98	6.40	217	243
Female Gender	145	240.67	7.80	217	260
Male Gender	152	242.34	8.71	218	276
Talented and Gifted	17	255.41	9.27	245	276
504 Plan	15	235.33	8.20	217	251
Special Education	22	233.36	11.13	218	276
White	264	242.02	8.32	217	276
Hispanic	19	238.11	7.08	224	250
Other Minority Non-Hispanic	14	236.71	7.30	224	248
Economically Disadvantaged	25	239.44	7.06	217	250

Correlation of Student-Level Variables

Paired correlations between all variables are presented in Table 4.5. Limited English proficient status was not included in the analysis as only one student was so identified. Positive and statistically significant correlations existed between ACT score and OAKS score ($r = .783$, $p <. 001$), between math GPA and ACT score ($r = .647$, $p <.001$), and between math GPA and OAKS score ($r = .607$, $p <. 001$). Correlations of this size suggest that ACT score, OAKS scores, and math GPA are strongly aligned with one another. In other words, when a student performs well on the ACT, she is likely to perform well on the OAKS, or when a student has a high math GPA, she is likely to perform well on both the ACT and OAKS.

Statistically significant correlations also existed between accelerated mathematics trajectory and ACT score ($r = .584$, $p <. 001$), accelerated mathematics trajectory and OAKS score ($r = .465$, $p <. 001$), accelerated mathematics trajectory and mathematics GPA ($r = .428$, $p <. 001$), and between OAKS score and talented and gifted status ($r = .413$, $p <. 001$). Overall, correlations ranged from a high of .783 (between OAKS score and ACT score) and a low of -.002 (between special education and non-Hispanic, other minority. None of the correlations showed the degree of redundancy necessary to present an initial multicollinearity concern (Abrams, 2007) so all were included in the multiple regression analyses.

Table 4.5

Bivariate Pearson Correlations, Student-Level Variables, n = 299

	Hispanic	Non-Hispanic Other Minority	Gender	504 Plan	SpEd Status	TAG Status	Economic-ally Disad-vantaged	Attendance Rate	Discipline Issues	ACT Score	OAKS Score	Math GPA	Remedial Math Trajectory
Non-Hispanic Other Minority	-.058												
Gender	.035	-.037											
504 Plan	.003	-.051	.071										
SpEd Status	-.073	-.002	.122*	-.006									
TAG Status	-.064	-.054	.009	-.056	-.014								
Economically Disadvantaged	.218**	-.010	-.019	-.014	.054	-.074							
Attendance Rate	-.015	-.191**	.078	.009	-.150**	.079	-.023						
Discipline Issues	-.021	.014	.173**	.036	.350**	-.016	.036	-.040*					
ACT Score	-.164**	-.092	.113	-.113	-.017	.369**	-.135*	.092	-.085				
OAKS Score	-.108	-.129*	.100	-.172**	-.278**	.413**	-.076	.210**	-.176**	.783**			
Math GPA	.042	-.022	-.155**	-.146*	-.261**	.254**	-.164**	.255**	-.238**	.647**	.607**		
Remedial Math Trajectory	.067	.028	.025	.184**	.560**	-.110	.156**	-.174**	.185**	-.384**	-.512**	-.296**	
Accelerated Math Trajectory	-.085	-.011	-.030	-.120*	-.160**	.274**	-.123*	-.039	-.094	.584**	.465**	.428**	-.299**

*$p < 0.05$. **$p < 0.001$.

Regression Analyses Controlling for Student Variables

Results for each of the two regression models are presented in three steps. The first step presents the results associated with a test of model assumptions and an examination of model-based multicollinearity using Variance Inflation Factors (VIF). The second step presents the overall test of model significance. Finally, individual coefficients for each of regression model are presented and discussed.

ACT regression analyses. For the first regression model, ACT score was the dependent variable or outcome and independent or predictor variables included math GPA, gender, ethnicity, 504 Plan, special education, talented and gifted, economically disadvantaged, attendance rate, major discipline issues, and math trajectory.

Analyzing model assumptions and multicollinearity. When conducting an analysis using a linear regression, first steps include verification that the data may actually be analyzed using a linear regression. Several assumptions must be met to satisfy this process. First, a linear relationship must exist. As seen in Appendix A, a linear relation between variables does exist. The second assumption is that within the variables, there are no influential cases, and the third assumption is that homoscedasticity violations do not exit. These assumptions were both met, as seen in the residual scatter plot (see Appendix B). Neither outliers nor residual patterns exist. Finally, it is assumed that the model is approximately normal. Again, this assumption was met (see Appendix C).

In addition to examining assumptions, tolerance and Variance Inflation Factor (VIF) are examined to rule out multicollinearity. Tolerance explains the proportion of variability not explained by linear relationships with other independent variables in the model. A value close to zero indicates a near-linear relationship between two independent variables while a value close to one indicates that little of a variable's variance is explained by other independent variables (Norusis, 2002). Multicollinearity occurs when two variables exist in a near-perfect linear relationship. The reciprocal of tolerance, large VIF values, typically those exceeding 10 indicate excessive collinearity (Mansfield & Helms, 1982).

The tolerance statistics in Table 4.6 indicate that multicollinearity was not a problem for any of the variables. Variables in the ACT model showed tolerances ranging from .687 (math GPA) to .937 (504 Plan) which indicate that these eleven variables are not collinear (Tomkins, 1992).

VIF statistics for all ACT model independent variables were below 10 (see Table 4.4), indicating that multicollinearity was not likely a problem (Belsey, Kuh, & Welsch, 1980; Gammie, Jones, & Robertson-Miller, 2003). VIF statistics ranged from a low of 1.067 (504 Plan) to a high of 1.455 (math GPA).

Table 4.6

Tolerance and Variance Inflation Factor Matrix for ACT Model

	Tolerance	VIF
Math GPA	.687	1.455
Gender	.933	1.072
Hispanic	.890	1.123
Non-Hispanic Other Minority	.929	1.077
504 Plan	.937	1.067
Special Education	.927	1.078
Talented and Gifted	.871	1.148
Economically Disadvantaged	.884	1.131
Attendance Rate	.870	1.150
Discipline Issues	.929	1.077
Remedial Math Trajectory	.874	1.144
Accelerated Math Trajectory	.717	1.395

Connection among measurement variables and research questions. The two research questions guiding this study concern the unique predictive relationship between mathematics GPA and a high-stakes exam outcome and the unique predictive relation of student-level variables such as Hispanic ethnicity and talented and gifted status to a high-stakes exam outcome. Regression analyses provide statistics to answer these questions.

Overall model results revealed that the ACT regression model was statistically significant, $p < .0001$ and explained 66.7% ($R^2 = .667$) of ACT score variance (see table

4.7 for complete ANOVA statistics). The overall model results indicate that one or more predictor variables was statistically related to the outcome.

Table 4.7

ANOVA Statistics for ACT Model

Model	Sum of Squares	*df*	Mean Square	*F*	Sig.
Regression	2853.842	12	237.820	38.832	.000
Residual	1426.976	233	6.124		
Total	4280.817	245			

Table 4.8 provides results for the multiple regression using ACT as the outcome or dependent variable. As shown in Table 4.6, the following variables were statistically significant predictors of ACT scores: (a) math GPA, (b) gender, (c) talented and gifted status, (d) remedial mathematics trajectory, and (e) accelerated mathematics trajectory. The following variables were non-significant: (a) Hispanic ethnicity, (b) non-Hispanic, other minority (c) 504 Plan, (d) special education, (e) economically disadvantaged, (f) attendance Rate, and (g) discipline issues.

Unstandardized coefficients allow for a mean comparison of outcome variables in terms of predictor variables. For example, specific to the ACT model in this study, for every one unit increase in math ACT score for female students, math ACT scores for male students would rise 1.740 points (b = 1.740) and for every one unit increase in White scores, Hispanic scores would fall 1.142 points (b = -1.142). For context, students may achieve a score from one to 36 on the math portion of the ACT. Additional results include: (a) non-Hispanic, other minority (b = -1.444) compared to White students, (b)

504 Plan (*b* = -.061) compared to non-504 Plan, (c) special education (*b* = 1.477) compared to non-special education, (d) talented and gifted (*b* = 2.207) compared to non-talented and gifted, (e) economically disadvantaged (*b* = .138) compared to economically advantaged, (f) remedial math trajectory (*b* = -3.375) compared to a standard math trajectory, and (g) accelerated math trajectory (*b* = 2.335) compared to a standard math trajectory.

Because math GPA, attendance rate, and major discipline issues do not have referent groups, it is appropriate to interpret the standardized coefficients for these variables. Standardized coefficients are standardized to a mean of zero and a standard deviation of one and function as the *x*- and *y*-axis coordinates on a graph. Thus, for every one standard deviation for ACT score (*x*-axis), math GPA increases by .515 standard deviations (*y*-axis). Additional results include: (a) attendance rate (β = -.052), and (b) major discipline issues (β = .002).

Table 4.8

Regression of Mathematics ACT

Model	Unstandardized Coefficients		Standardized Coefficients	*t*	Sig.
	b	Std. Error	Beta		
Math GPA	2.419	.214	.515	11.280	.000
Gender	1.740	.327	.209	5.326	.000
Hispanic	-1.142	.678	-.068	-1.823	.093
Non-Hispanic Other Minority	-1.444	.792	-.072	-1.823	.070
504 Plan	-.061	.980	-.002	-.063	.950
Special Education	1.477	1.493	.039	.990	.323
Talented and Gifted	2.207	.666	.134	3.312	.001
Economically Disadvantaged	.138	.680	.008	.202	.840
Attendance Rate	-.017	.014	-.052	-1.286	.200
Major Discipline Issues	.035	.754	.002	.047	.963
Remedial Math Trajectory	-3.375	.591	-.231	-5.707	.000
Accelerated Math Trajectory	2.335	.394	.265	5.923	.000

Table 4.9 provides more in-depth information about the regression analysis, specifically zero-order, partial, and semi-partial correlations for the ACT model. Of the variables which were statistically significant (see Table 4.6), the semi-partial correlation for math GPA (pr = .427) explained the greatest single amount of variance. The square of the coefficient showed that 18.23% of the variance was uniquely explained by

mathematics GPA. The variable that explained the next greatest amount of variance was accelerated mathematics trajectory (*pr* = .224), which uniquely explained 5.02% of the variance. The remaining three significant variables uniquely explained variance in this order: (a) remedial mathematics trajectory (*pr* = -.216), 4.67%, (b) gender (*pr* = .201), 34.04%, and, (c) talented and gifted status (*pr* = .125) 1.56%.

Table 4.9

Part and Partial Correlations: Mathematics ACT Model

Model	Correlations		
	Zero-order	Partial	Semi-Partial
Math GPA	.649	.594	.427
Gender	.145	.329	.201
Hispanic	-.124	-.110	-.064
Non-Hispanic Other Minority	-.096	-.119	-.069
504 Plan	-.117	-.004	-.002
Special Education	.031	.065	.037
Talented and Gifted	.375	.212	.125
Economically Disadvantaged	-.156	.013	.008
Attendance Rate	.078	-.084	-.049
Major Discipline Issues	-.065	.003	.002
Remedial Math Trajectory	-.364	-.350	-.216
Accelerated Math Trajectory	.592	.362	.224

OAKS regression analyses. For the second regression model, OAKS score was the dependent variable or outcome. Independent variables included math GPA, gender, Hispanic and non-Hispanic ethnicity, 504 Plan, special education, talented and gifted, economically disadvantaged, attendance rate major discipline issues, remedial math trajectory, and accelerated math trajectory.

Analyzing model assumptions and multicollinearity. As with the ACT model, three assumptions were first tested for the OAKS model: (a) that a linear relation between variables does exist, (b) that within the variables, there are no influential cases, and (c) that homoscedasticity violations do not exit. As seen in Appendices D-F, these assumptions were met for the OAKS model.

Tolerance and variance inflation factor were also examined for the OAKS model. The tolerance statistics in Table 4.10 indicate that tolerances ranged from .609 (remedial math trajectory) to .927 (504 Plan).

VIF statistics for all variables were below ten (see Table 4.10), thus indicating that multicollinearity was not a problem among variables, VIF statistics ranged from a low of 1.079 (504 Plan) to a high of 1.641 (remedial math trajectory).

Table 4.10

Tolerance and Variance Inflation Factor Matrix for OAKS Model

	Tolerance	VIF
Math GPA	.640	1.563
Gender	.925	1.081
Hispanic	.903	1.107
Non-Hispanic Other Minority	.922	1.085
504 Plan	.927	1.079
Special Education	.612	1.634
Talented and Gifted	.885	1.130
Economically Disadvantaged	.876	1.141
Attendance Rate	.852	1.174
Discipline Issues	.868	1.152
Remedial Math Trajectory	.609	1.641
Accelerated Math Trajectory	.728	1.373

Connection among measurement variables and research questions. ANOVA results showed the OAKS model as significant, $p < .0001$ and explained 60.7% (R^2 = .607) of OAKS score variance (see table 4.11 for complete ANOVA statistics). A significant model means that at least one of the independent variables in the model was a significant predictor of an OAKS score.

Table 4.11

ANOVA Statistics for OAKS Model

Model	Sum of Squares	*df*	Mean Square	*F*	Sig.
Regression	11416.783	12	951.399	34.766	.000
Residual	7388.821	270	27.366		
Total	18805.604	282			

Table 4.12 provides results for the multiple regression using OAKS as the outcome or dependent variable. The regression showed the following variables were statistically-significant predictors of ACT scores: (a) math GPA, (b) gender, (c) Hispanic, (d) non-Hispanic, other minority, (e) talented and gifted status, (e) economically disadvantaged, (f) remedial mathematics trajectory, and (g) accelerated mathematics trajectory. The following variables were non-significant: (a) 504 Plan, (b) special education, (c) attendance Rate, and (d) discipline issues.

Being an adaptive test, OAKS scale scores range from zero to infinity, with most students scoring in the 150 to 300 range. Regression results showed the following unstandardized coefficients: (a) male scores (b = 3.055) compared to female scores, (b) Hispanic (b = -3.106) compared to White, (c) non-Hispanic, other minority (b = -3.267) compared to White, (d) 504 Plan (b = -.628) compared to non-504 Plan, (e) special education (b = .917) compared to non-special education, (f) talented and gifted (b = 7.783) compared to non-talented and gifted, (g) economically disadvantaged (b = 3.081) compared to economically advantaged, (h) remedial math trajectory (b = -6.954) compared to a standard math trajectory, and (i) accelerated math trajectory (b = 2.550) compared to a standard math trajectory. Standardized coefficients include: (a) math GPA (β = -.450), (b) attendance rate (β = -.003), and (c) major discipline issues (β = -.031).

Table 4.12

Regression of Mathematics OAKS

Model	Unstandardized Coefficients		Standardized Coefficients	*t*	Sig.
	b	Std. Error	Beta		
Math GPA	3.811	.404	.450	9.444	.000
Gender	3.055	.647	.187	4.723	.000
Hispanic	-3.106	1.377	-.091	-2.255	.025
Non-Hispanic Other Minority	-3.267	1.607	-.081	-2.033	.043
504 Plan	-.628	1.490	-.017	-.421	.674
Special Education	.917	1.629	.027	.563	.574
Talented and Gifted	7.783	1.391	.227	5.594	.000
Economically Disadvantaged	3.081	1.192	.105	2.584	.010
Attendance Rate	.017	.028	-.003	.800	.424
Major Discipline Issues	-.786	1.035	-.031	-.759	.448
Remedial Math Trajectory	-6.954	1.121	-.303	-6.206	.000
Accelerated Math Trajectory	2.550	.790	.144	3.228	.001

Table 4.13 displays zero-order, partial, and semi-partial correlations for the OAKS model. Within the OAKS model, the same variables were statistically significant as in the ACT model, with the addition of several other variables. Of the variables which

were significant predictors of OAKS performance (see Table 4.12), the semi-partial correlations indicated that math GPA (*pr* = .360) explained the greatest single amount of variance (12.96%). Remedial mathematics trajectory (*pr* = -.237) explained the next greatest amount of variance (5.62%), followed by (a) talented and gifted (*pr* = .213), 4.54%, (b) gender (*pr* = .180) 3.24%, (c) accelerated math trajectory (*pr* = .123), 1.51%, (d) economically disadvantaged (*pr* = .099) .98%, (e) Hispanic (*pr* = -.086) .74%, and (f) non-Hispanic, other minority (*pr* = -.078) .61%.

Table 4.13

Part and Partial Correlations: Mathematics OAKS Model

Model	Correlations		
	Zero-order	Partial	Semi-Partial
Math GPA	.604	.498	.360
Gender	.113	.276	.180
Hispanic	-.091	-.136	-.086
Non-Hispanic Other Minority	-.118	-.123	-.078
504 Plan	-.159	-.026	-.016
Special Education	-.247	.034	.021
Talented and Gifted	.419	.322	.213
Economically Disadvantaged	-.102	.155	.099
Attendance Rate	.224	.048	.030
Major Discipline Issues	-.160	-.046	-.029
Remedial Math Trajectory	-.469	-.353	-.237
Accelerated Math Trajectory	.477	.193	.123

Summary of Findings

In summary, the purpose of this study was two-fold: a) to examine the relation of GPA to an outcome on two high-stakes assessments, and, b) to examine the degree to which student factors other than GPA predict an outcome on two high-stakes assessments. Multiple regression analyses were performed to identify statistically significant predictors of mathematics ACT and OAKS scores. Findings are summarized below and are visually represented in residual scatter plots in Appendices A through F.

The first model, using ACT as the dependent variable, was statistically significant ($F_{12, 233} = 38.832$, $p < .0001$) and the complete model explained 66.7% of score variance ($R^2 = .667$). An examination of the semi-partials revealed that six variables were statistically-significant predictors of ACT score: (a) math GPA (*pr* = .427), (b) accelerated math trajectory (*pr* = .224), (c) remedial math trajectory (*pr* = -216), (d) gender (*pr* = .201), and, (e) talented and gifted (*pr* = .125).

Similar to the ACT model, the OAKS model was also statistically significant ($F_{12, 270} = 34.766$, $p < .0001$) and the complete model explained 60.7% of score variance ($R^2 = .607$); however, in addition to the six significant variables identified by the ACT model, the OAKS model included two more. Specifically, eight of twelve variables were statistically significant and the complete model explained : (a) math GPA (*pr* = .360), (b) remedial math trajectory (*pr* = -.237), (c) talented and gifted (*pr* = .213), (d) gender (*pr* = .180), (e) accelerated math trajectory (*pr* = .123), (f) economically disadvantaged (*pr* = -.099), (g) Hispanic (*pr* = -.086), and (h) non-Hispanic, other minority (*pr* = -.078).

In both the ACT and OAKS models, math GPA explained the greatest amount of variance. Accelerated math trajectory explained the second greatest amount of variance in the ACT model and remedial math trajectory explained the second greatest amount of variance in the OAKS model. The remaining predictor variables differed slightly between models in both order and percent of variance explained. These data suggest a relationship exists between the variables GPA and math trajectory, which will be discussed in the next chapter.

CHAPTER V

DISCUSSION

The first section in this chapter includes a summary and interpretation of the findings obtained from the study, looking specifically at GPA and other student-level predictors. The second section addresses the limitations of the study, including threats to both internal and external validity, and the third section includes a discussion of the implications of the findings, potential areas for future research, and conclusions.

Summary of Findings

The purpose of this study was two-fold: (a) to examine the relation between math GPA and performance on high-stakes assessments, and (b) to examine the degree to which student factors other than GPA predict performance on high-stakes assessments. Data from 299 students from one graduating high school sample were used to examine these research questions. Answering these research questions required the use of two statistical techniques: bivariate correlations to examine variable relations, and multiple regression modeling to conduct analyses examining the effects of student-level predictors on the relation between GPA and high-stakes assessment outcomes.

Both the ACT and OAKS regression models were statistically significant and, in both, math GPA explained the greatest amount of variance. Accelerated math trajectory explained the second greatest amount of score variance in the ACT model and remedial math trajectory explained the second greatest amount of score variance in the OAKS model. The remaining significant variables differed slightly between models in both order and percent of variance explained. An examination of the semi-partial correlations indicated five variables were significant predictors of ACT score: (a) math GPA, (b) accelerated math trajectory, (c) remedial math trajectory, (d) gender, and (e) talented and gifted. In addition to the five variables that were significant predictors of performance on the ACT, economically disadvantaged, Hispanic, and non-Hispanic, other minority were significant predictors of performance on the OAKS mathematics test.

Implications

As noted in Chapter IV, both ACT and OAKS regression models were statistically significant. Specific findings from this study indicate that mathematics GPA, remedial math trajectory, accelerated math trajectory, gender, and talented and gifted status contribute to the amount of variance observed in student performance on both the ACT and OAKS assessments. Of these variables, mathematics GPA explained the greatest amount of high-stakes assessment score variance: 18.23% of the variance for the ACT and 12.96% of the variance for the OAKS.

Findings related to the literature review. As noted in Chapter II, the criteria that comprises GPA is unreliable, varying from teacher to teacher. Factors such as behavior, effort, and teacher perceptions are often included in assigned marks (Brookhart, 1993; Guskey, 2009; Randall & Engelhard, 2010; Rosenthal & Jacobson, 1968), and some teachers believe that course marks serve a purpose as a punishment (Guskey, 2009). Yet, despite their weaknesses, student grades have been documented to be a reliable predictor of student outcomes such as dropping out, graduating from high school, and first-year college success (Bowers, 2010b; Geiser & Santelices, 2007; Geiser & Studley, 2002). As previous studies have reported, high school grades and high-stakes assessment outcomes, combined, explain more variance in first-year college success than either variable alone (American College Test, 2007; Oregon Department of Education, 2011c). Of particular interest to this study were the relations between GPA, other student-level variables, and high-stakes assessment outcomes.

Results from this study indicate that math GPA is a statistically significant predictor, and explains the greatest amount of variance, of an outcome on the math portion of the ACT and OAKS. These findings do not contradict research that indicates teachers use unreliable grade reporting criteria as grading criteria was not an element of this study; however, the findings do add to the relationship between grades and outcomes such as achievement. The question remains whether student performance or student characteristics, both of which research has shown to be included in teacher grade calculations, is more predictive of these outcomes.

Findings related to school-based practice. The findings of this study suggest a closer examination of GPA and grade assignment by teachers and school leaders, as well as an examination of the process for course assignment, may be warranted. A closer examination of what factors contribute to student grades and to course assignments may provide insights that will enable practitioners to more closely align course marks with achievement and course assignments with skill. Unstandardized coefficients indicate that being enrolled in a remedial math trajectory results in the greatest amount of mean score difference relative to its referent group (standard math trajectory) for the ACT model (b = -3.375) and the second greatest amount of mean score difference relative to its referent group for the OAKS model (b = -6.954). Because of these mean score differences, it is important to note that grades typically provide the starting point for placement into the different math trajectories. Without low grades in prior math classes, the odds of a student being placed onto a standard or an accelerated math trajectory are limited. Based on the results from this study, the comorbidity of these variables should not be ignored. These variables, math GPA and remedial math trajectory, are the two predictors of performance on high-stakes assessments over which a school has significant control.

The manner in which schools operationalize subjects and grading practices, such as math courses and math grades, bears enormous weight on student outcomes. Specific to operationalizing math, each individual high school generally determines which math courses to offer, the order in which those math courses can be taken, and which students are eligible to participate in the different math courses (Maple High School, 2012). Some states do require students to successfully complete specific courses prior to graduation (California Department of Education, 2012; Nevada Department of Education, 2012; Oregon Department of Education); however, decisions about how to implement such mandates reside largely with individual high schools. This flexibility means that high schools may offer support courses to help students catch up, offer advanced courses to all students regardless of previously-earned grades, provide flexible course sequencing, or possibly offer none of these options.

Schools also control grading practices. Administration at a district or school level may mandate specific grading practices, such as reporting student characteristics separately or by weighting student characteristics at a lesser value than student performance. When schools and districts do not make these decisions, teachers determine what criteria to include in grade reporting, and as demonstrated in the research, this criteria is varied and unreliable (Brookhart, 1993; Guskey, 2009; Randall & Engelhard, 2010; Rosenthal & Jacobson, 1968).

Findings related to the ACT, the OAKS, and school variables. This study examined GPA and 11 different student-level variables. Math GPA explained the greatest variance in both the ACT regression model (18.23%) and the OAKS regression model (12.96%); however, remedial math trajectory accounted for the greatest mean difference (b = -3.375) in math ACT scores when compared to students on a standard math trajectory and the second greatest mean difference in math OAKS scores (b = -6.954). For the OAKS model, talented and gifted accounted for the greatest mean score difference (b = 7.783) compared to its referent group (students not categorized at talented and gifted).

A possible explanation for the strength of the relation between mathematics trajectory and performance on the high-stakes assessments relates to students' exposure to particular math concepts and curriculum. Students on an accelerated mathematics trajectory are those whose first high-school mathematics course was Geometry or higher, and it may be argued these same students have thus been exposed to more advanced mathematics prior to testing and would therefore be expected to perform better on assessments that cover a wide range of mathematics content. Results from this study support this as seen in the unstandardized coefficients for accelerated math trajectory for ACT (b = 2.335) and OAKS (b = 2.550).

Similarly, students on a remedial mathematics trajectory, having started their high school mathematics coursework with classes less challenging than Algebra 1, would have been exposed to fewer concepts and skills by the time they participated in the high-stakes assessment as high school juniors. Their relatively poor performance on the assessments makes sense purely from an opportunity to learn perspective. Not having been taught some of the concepts on the high-stakes tests, students whose mathematics course-taking trajectory can be classified as remedial would not be expected to perform as well on the high-stakes assessments as their peers who entered high school already having successfully completed Algebra 1 or who began high school with Algebra 1.

As noted previously, schools largely control math trajectory via the supports, prerequisites, and course sequence offered. Results from this study indicate that GPA explained the greatest amount of variance for both the ACT (18.23%) and OAKS (12.96%) outcome; however, accelerated math trajectory explained the next greatest amount of variance in the ACT model, explaining an additional 5.02% and remedial math explained the second greatest amount of variance in the OAKS model, an additional 5.62%. Given the research on teachers' grading practices (Brookhart, 1993; Guskey, 2009; Randall & Engelhard, 2010) it is not surprising that adding mathematics trajectory to the model added predictive power. A student may well earn an A in very easy mathematics courses yet still not know how to correctly address math questions related to content knowledge she has not been taught.

Based on these results, students would benefit from schools re-examining their math scope-and-sequence from kindergarten through grade 12 to ensure students are provided the opportunity to complete a standard or accelerated math trajectory. In addition, working to ensure that grades accurately reflect student mastery of content knowledge may help increase the utility of math GPA for predicting performance on high-stakes assessments. It should be noted that to the extent that grades do capture student knowledge and reflect the same content as is used to create the large-scale assessments, mathematics GPA may become an even stronger predictor of student performance on high-stakes tests. Even with increased GPA criteria fidelity, however, it is likely that students' exposure to more challenging math curriculum, as is operationalized in this study by their classification into a *remedial, standard*, or *accelerated* math trajectory, will continue to contribute unique predictive power to the analyses.

Findings related to race / ethnicity and economically disadvantaged. An examination of *t*-scores in the ACT model revealed that that ethnicity and economically disadvantaged were *not* statistically significant predictors of performance on the ACT. Characteristics of the students who did not take the ACT likely had little effect on *t*-scores. Specifically, of the 45 students within the sample who did not take the ACT, only one was Hispanic and one was non-Hispanic, other minority. Of the 25 economically disadvantaged students in the sample, eight did not take the ACT.

Changing the model to substitute OAKS scores as the outcome produced different results that included economically disadvantaged status, Hispanic ethnicity, and non-Hispanic, other minority as statistically-significant predictor variables. Among these variables, the unstandardized coefficient for economically disadvantaged (b = 3.081) is worth noting. It was unexpected that economically disadvantaged students would outperform their economically advantaged peers. One plausible explanation is that teachers are not privy to economically disadvantaged information; thus, their perceptions may not be influenced by knowledge of student economic status.

Teacher perception may also explain the statistical significance of the Hispanic and non-Hispanic, other minority variables. Teacher behaviors and perceptions teachers have of students are associated with student outcomes (Doyle, 1977; Rosenthal & Jacobson, 1968). It is possible that teacher perceptions of minority students may have influenced student performance and, thus, OAKS outcomes. Because teachers administer the OAKS directly to their own students, in contrast to the ACT which is proctored by an adult who may or may not teach the students for whom she is proctoring, students may simply be meeting teacher expectations. It is also possible, given the low minority population in this analytic sample that teachers for this sample of students lacked the awareness or skills to differentiate instruction for minority subgroups. These students may also have entered school with an academic gap that remained static or grew prior to the 2011-2012 school year.

Schools cannot control teacher perceptions, but schools can take steps to minimize placement of students from ethnic and racial minority and economically disadvantaged backgrounds from being placed on a K-12 remedial math trajectory. Specifically, schools can adopt a holistic approach to preparing students that includes explicit instruction not only in content knowledge, but also in cognitive strategies, learning skills and techniques, and transition skills that include self-advocacy (Conley, 2012). Additionally, schools can adopt structures that minimize ability grouping, especially in the younger years, as studies have shown that ability tracking tends to perpetuate learning gaps, whereas “detracked” courses close gaps and raise student performance (Lleras & Rangel, 2009; Oakes & Wells, 1998).

Findings related to other factors. Major discipline issues, attendance rate, special education status, and 504 Plan status were not statistically significant predictors of ACT or OAKS performance. Both the ACT and OAKS assessment allow for several accommodations such as additional time, a smaller test environment, and the use of calculators. It is important to note that while many accommodations, like those listed previously, are available to all students regardless of classification, such accommodations may be necessary for students with 504 Plans or special education and may be one reason special education and 504 Plan were not statistically significant predictor variables. Another reason may be the small sample size. Only seven students with a 504 Plan and four students who were receiving special education services completed the ACT, while 15 students with a 504 Plan and 22 students receiving special education services completed the OAKS. Small sample, sizes such as seen for the variables 504 Plan and special education, may impact power, which is the probability of detecting a true effect when one does exist (Grimm, 1995). It is worth noting, however, despite the statistical insignificance of special education in this study, special education was moderately correlated with remedial math (r = .560) which was a statistically significant variable in both models.

That major discipline issues and attendance rate were not statistically significant predictors may be attributable to the demographics of this specific population. As noted in Chapter IV, the students included in this study generally achieved a 90% attendance rate and major discipline issues were confined to just 3% of the sample population. Thus, there was very little variance in these specific variables, effectively limiting their potential contribution to the regression models.

Study Limitations

Several limitations to this study pose threats to both internal and external validity and thus to interpretations of the findings. One notable limitation was the limited number and relative lack of diversity of the students participating in the study. The small sample size and the homogeneity of the sample not only affect statistical power, but also generalizability of findings, particularly in relation to ACT outcomes as only 254 students, 45 less than the sample, completed the ACT. A small sample size makes it more difficult to detect true effects and the homogeneity of the sample makes it difficult to generalize study results to more ethnically and economically diverse populations.

Another limitation was the inability to control for all possible variables. For example, research shows that teacher grading philosophies impact grading practices (Brookhart, 1993; Guskey, 2009; Randall & Engelhard, 2009b, 2010), but the scope of this study did not include a survey of the grading philosophies of Maple High School's mathematics teachers. Thus, I am limited to expressing caution in interpretation and describing the potential variability in grading practices rather than including specific, measured, grading practices as part of the analytic model. Future research might include information about grading practice in the model and thus provide more precise understanding of this variable and its ability to predict performance on high-stakes assessments.

Threats to internal validity. One potential internal validity threat is testing, specific to the dependent variable OAKS scores. Students have multiple opportunities to take an OAKS assessment. For the sample of students in this study, opportunity to take the OAKS actually began in grade 10. The OAKS scores used in this study represent the students' best score on the OAKS, which may have been achieved after one or several attempts. Not only does this provide students more time to acquire content knowledge, it also provides students several exposures to the assessment format, both of which may result in an improved outcome over time. No information about number of attempts or specific point in a student's high school career at which the OAKS score was recorded was provided in the data set used in this study. Thus, I was not able to include this information in my modeling. Furthermore, in the State of Oregon, ACT scores trump OAKS scores. In other words, if a student performs better on the mathematics ACT than the OAKS, the ACT may be used in place of the OAKS to fulfill graduation requirements. Additional factors which may impact testing validity include: (a) under-accommodating, (b) over-accommodating, (c) test environment, (d) student motivation, and (e) testing equipment (e.g. working computers). When any one of these occur, student performance may not accurately reflect what a student knows and can do.

Threats to external validity. The generalizability of this study may be limited. First, the sample population in this study was one of convenience. Maple School District agreed to provide data, and this was the primary reason they were selected for the study. Although the study population is representative of the overall demographics of Oregon, the state in which Maple High School is located, it is not necessarily representative of many of the school districts in the state, particularly those districts with large minority populations or many students from low socio-economic backgrounds. Furthermore, the outcome measures used in this study may not generalize to other mathematics measures, such as the Scholastic Aptitude Test (SAT) or the Trends in International Mathematics and Science Study (TIMSS). Finally, one outcome measure, the OAKS assessment, will be used in the State of Oregon for only one more year (during the 2013-2014 school year). Beginning in the 2014-2015 school year, Oregon will convert to use of the Smarter Balance assessment; thus, results specific to OAKS scores may have limited generalizability beyond the 2013-2014 school year.

Suggestions for Further Study

Although these findings indicate that GPA is a statistically significant predictor of performance on high-stakes assessments, it is clear that additional research is necessary to comprehensively address the research questions posed in this study. Specifically, research into the following areas is needed: (a) the grading philosophies of mathematics teachers, (b) consideration of other student-level variables that may be influencing high-stakes assessment outcomes, (c) the process by which students are assigned to a specific mathematics course, and (d) the generalizability of these results to other student populations.

Adding a qualitative component to the study in which mathematics teachers are queried to learn their views on the purpose and role of grade assignment would add another lens through which to view these data. Research has indicated, for example, that some teachers view grades as a form of punishment and that most teachers recognize the social implications of grade assignment (Brookhart, 1993; Guskey, 2009). Research has also shown that student characteristics factor heavily into grade assignment (Finkelstein, 1913; Randall & Engelhard, 2009b, 2010). Questions that may yield valuable insights include the role of assessment versus homework or classwork in grade assignment, the role of effort and behavior in grade assignment, and the purpose of grades.

Additionally, because many factors others than those included in this study affect both academic performance and student achievement, further examination of such student-level variables is warranted. It would be worthwhile to examine factors such as student motivation, extra-curricular involvement, one- or two-parent households, and parental involvement. Such an examination would expand current understanding about the relation between academic performance and achievement and the many qualitative factors, such as student motivation, that may influence outcomes, but are data seldom collected for analysis at a school or district-level.

Finally, this study warrants expansion into other districts with more diverse student populations. Although the ethnic and racial demographics of the study population mirror the population in the State of Oregon (U.S. Census, 2011), the state within which Maple High School resides, the sample population is not representative of many Oregon school districts, particularly those in urban areas. Furthermore, the socio-economic status of Maple High School students is significantly higher than Oregon school districts in general (Oregon Department of Education, 2011b).

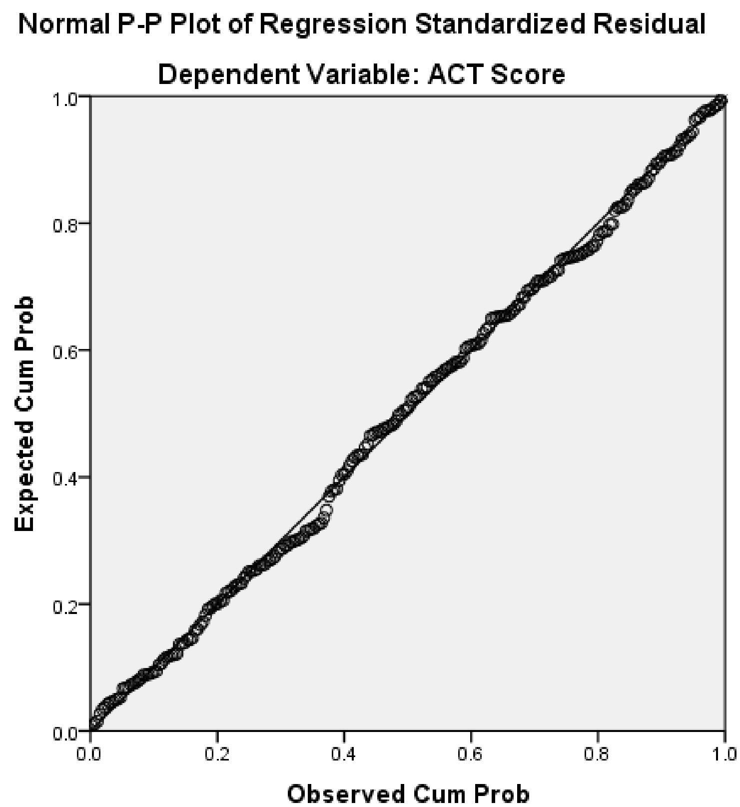
Conclusions

Two general conclusions emerge from this study. First, of the variables analyzed, mathematics GPA explained the greatest amount of variance for both the ACT and OAKS assessments in mathematics; thus, for this sample, GPA is a statistically significant predictor of an outcome on a high-stakes assessment. Second, of the remaining eleven student-level variables examined in both models, accelerated mathematics trajectory and remedial mathematics trajectory followed math GPA as the variables that explained the second greatest amount of variance in the ACT and OAKS models, respectively. This finding suggests that freshman-year math placement matters in predicting high-stakes assessment outcomes.

APPENDIX A

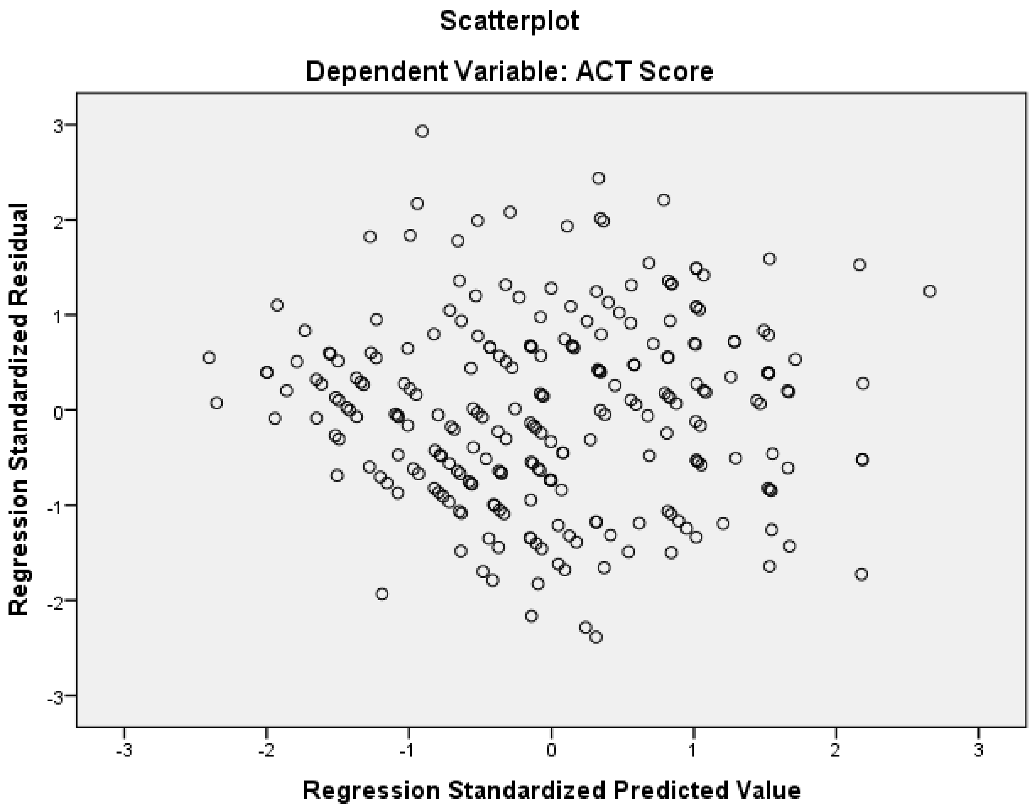
RESIDUAL NORMAL PROBABILITY PLOT FOR DEPENDENT VARIABLE

MATHEMATICS ACT SCORE



APPENDIX B

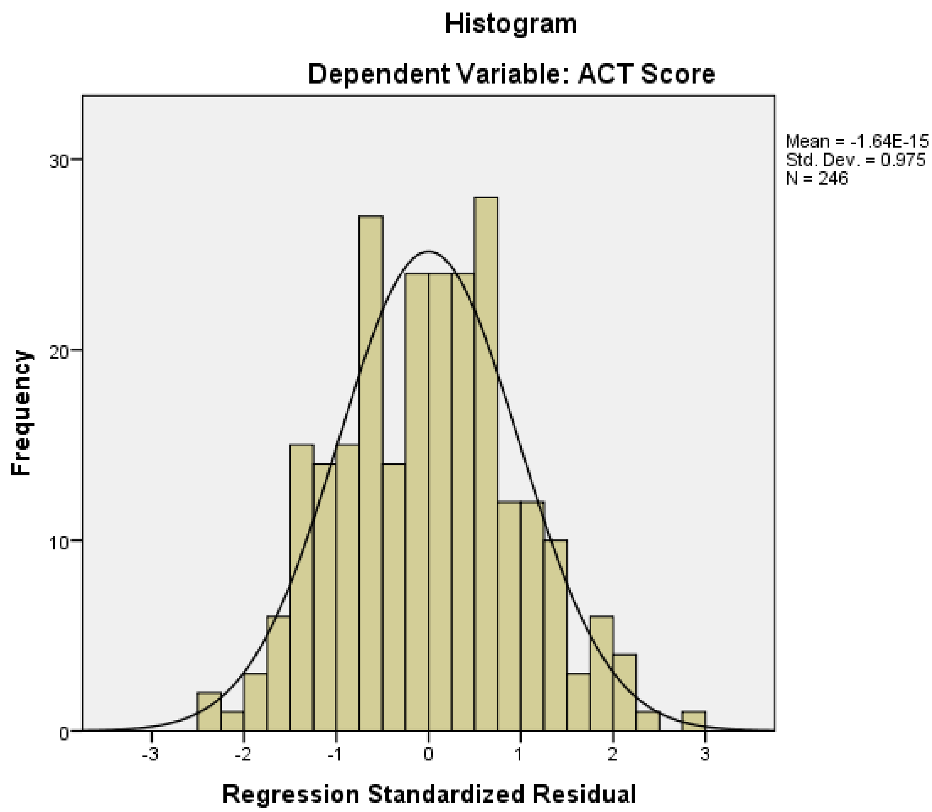
RESIDUAL SCATTER PLOT FOR DEPENDENT VARIABLE MATHEMATICS ACT SCORE



APPENDIX C

RESIDUAL HISTOGRAM FOR DEPENDENT VARIABLE MATHEMATICS ACT

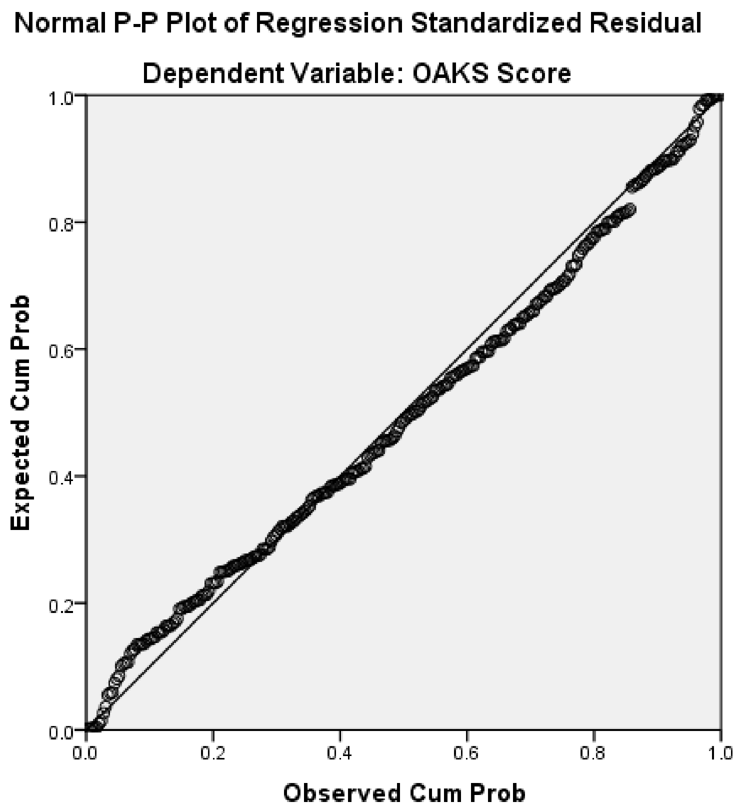
SCORE



APPENDIX D

RESIDUAL NORMAL PROBABILITY PLOT FOR DEPENDENT VARIABLE

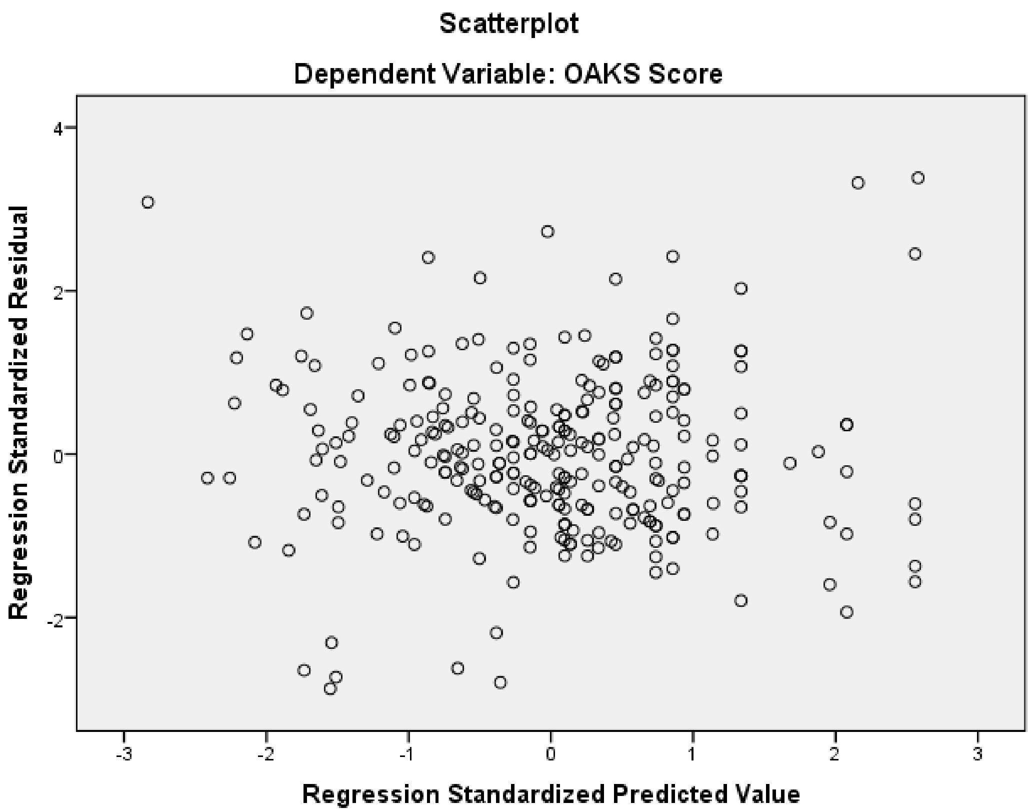
MATHEMATICS OAKS SCORE



APPENDIX E

RESIDUAL SCATTER PLOT FOR DEPENDENT VARIABLE MATHEMATICS

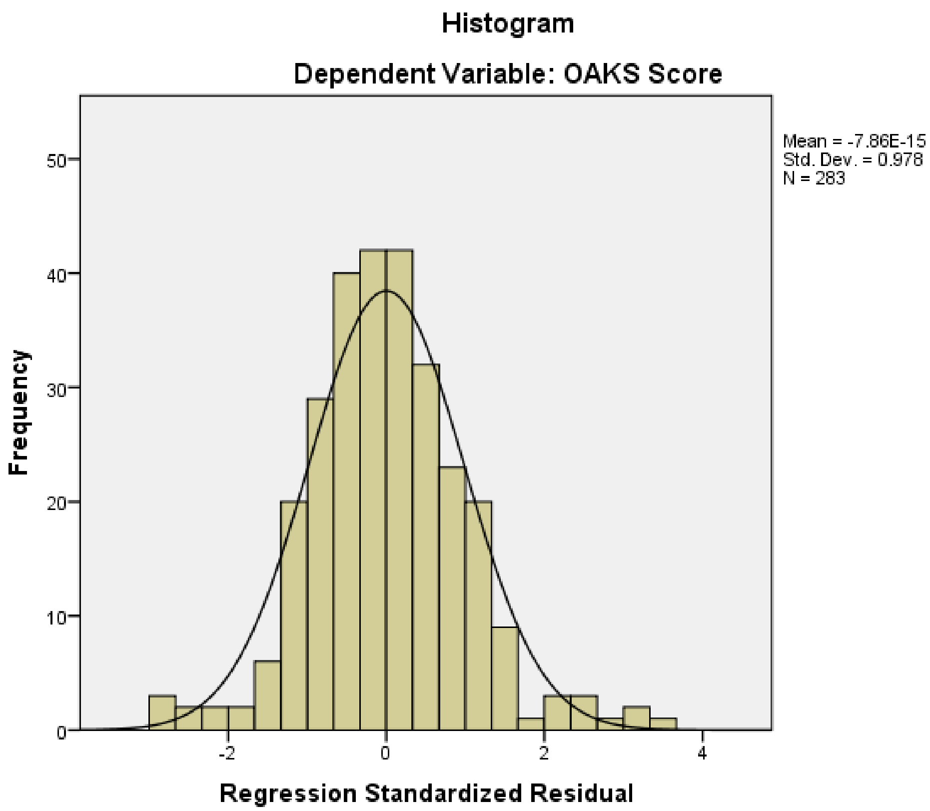
OAKS SCORE



APPENDIX F

RESIDUAL HISTOGRAM FOR DEPENDENT VARIABLE MATHEMATICS OAKS

SCORE



REFERENCES CITED

Abrams, D. R. (2007). Introduction to Regression Retrieved 17 April 2013, from http://dss.princeton.edu/online_help/analysis/regression_intro.htm

American College Test. (2007). The ACT Technical Manual.

American College Test, I. (2005). Are High School Grades Inflated? Issues in College Readiness: ACT, Inc.

Au, W. (2007). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher, 36*(5), 258-267.

Ball, D. L., Thames, M. H., & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education, 59*(5), 389-407.

Belsey, D. A., Kuh, E., & Welsch, R. E. (1980). *Regression Dianogstics: Identifying Data and Sources of Collinearity*. New York: John WIley and Sons.

Beswick, K. (2012). Teachers' Beliefs about School Mathematics and Mathematicians' Mathematics and Their Relationship to Practice. *Educational Studies in Mathematics, 79*(1), 127-147.

Bidwell, J. K., & Clason, R. G. (1970). *Readings in the history of mathematics education*. Washington D.C.: Nat'l Council of Teachers of Mathematics.

Bishop, J. H. (2001). A Steeper, Better Road to Graduation. *Education Next, 1*(4), 56-61.

Bloom, B. S. (1956). *Taxonomy of educational objectives; the classification of educational goals*. New York: Longmans, Green.

Bowers, A. J. (2010a). Analyzing the longitudinal K-12 grading histories of entire cohorts of students: Grades, data driven decision making, dropping out and hierachial cluster analysis. *Practical Assessment, Research & Evaluation, 15*(7), 1-18.

Bowers, A. J. (2010b). Grades and graduation: A longitudinal risk perspective to identify student dropouts. *Journal of Educational Research, 103*(3), 191-207.

Brookhart, S. M. (1993). Teachers' grading practices: Meaning and values. *Journal of Educational Measurement, 30*(2), 123-142.

Buckley, L. A. (2010). Unfulfilled hopes in education for equity: Redesigning the mathematics curriculum in a US high school. *Journal of Curriculum Studies, 42*(1), 51-78.

California Department of Education. (2012). Graduation requirements Retrieved 20 February 2012, from http://www.cde.ca.gov/ci/gs/hs/hsgrtable.asp

Camara, W. J., & Echternacht, G. (2000). The SAT I and High School Grades: Unity in Predicting Success in College *Research Notes* (Vol. 10, pp. 1-12): The College Board.

Carter, M. (2006). Why Aren't Grades Enough: North Carolina State University.

Cattell, J. M. (1905). Examinations, grades, and credits. *Popular Science Monthly, 66*, 367-378.

Chappuis, S., & Chappuis, J. (2008). The Best Value in Formative Assessment. *Educational Leadership, 65*(4), 14-19.

Cliffordson, C. (2008). Differential Prediction of Study Success across Academic Programs in the Swedish Context: The Validity of Grades and Tests as Selection Instruments for Higher Education. *Educational Assessment, 13*(1), 56-75.

Common Core State Standards Initiative. (2010). Common core state standards for mathematics.

Conley, D. T. (2003). *Who governs our schools?: Changing roles and responsibiliites*. New York: Teachers College Press.

Conley, D. T. (2012). A Complete Definition of College and Career Readiness. *Educastional Policy Improvement Center Publication*.

Conley, D. T. (2013). *Student Ownership of Learning: The Key to College and Career Readiness and Common Core Success*. Paper presented at the International Forum on Education and the Economy, Orlando, Florida. http://www.epiconline.org/publications/student-ownership-of-learning-the-key-to-college-and-career-readiness-and-common-core-success

Davison, H. M., & Pennsylvania State Univ, U. P. (1964). *The problem of marks*.

Doyle, W. (1977). Paradigms for Research on Teacher Effectiveness. *Review of Research in Education, 5*(ArticleType: research-article / Full publication date: 1977 / Copyright © 1977 American Educational Research Association), 163-198. doi: 10.2307/1167174

Epstein, N. (Ed.). (2004). *Who's in charge here?: The tangled web of school governance and policy*. Washington, D.C.: Brookings Intitution Press.

Finkelstein, I. E. (1913). *The marking system in theory and practice.* Warwick & York, inc., Baltimore. Available from http://worldcat.org /z-wcorg/ database.

Frey, M. C., & Detterman, D. K. (2004). Scholastic Assessment or g?: The Relationship Between the Scholastic Assessment Test and General Cognitive Ability. *Psychological Science, 15*(6), 373-378. doi: 10.1111/j.0956-7976.2004.00687.x

Gammie, E., Jones, P., & Robertson-Miller, C. (2003). Accountancy undergraduate performance: A statistical model. *Accounting Education, 12*(1), 63-78.

Gardner, H. (2006). *The development and education of the mind : the selected works of Howard Gardner*. London: Routledge.

Geiser, S., & Santelices, M. V. (2007). Validity of high-school grades in predicting student success beyond the freshmen year: High school record vs. standardized tests as indicators of four-year college outcomes (Vol. 9): Center for Studies in Higher Education.

Geiser, S., & Studley, R. (2002). UC and the SAT: Predictive Validity and Differential Impact of the SAT I and SAT II at the University of California. *Educational Assessment, 8*(1), 1-26.

George, D., & Mallory, P. (2003). *SPSS for Windows step by step: A simple guide and reference* (4th ed.). Boston: Allyn & Bacon.

Grimm, L. G. Y. P. R. (1995). *Reading and understanding multivariate statistics*. Washington, D.C.: American Psychological Association.

Guskey, T. R. (2009). Bound by tradition: Teachers' views of crucial grading and reporting issues: Online Submission.

Hambleton, R. K., Swaminathan, H., Algina, J., & Couldon, D. B. (1978). Criterion-referenced testing and measurement: A review of technical issues and developments. *Review of Educational Research, 48*(1), 1-47.

Jones, B. D., & Egley, R. J. (2004). Voices from the frontlines: Teachers' perceptions of high-stakes testing. *Education Policy Analysis Archives, 12*(39).

Kaufman, S. B., Reynolds, M. R., Liu, X., Kaufman, A. S., & McGrew, K. S. (2012). Are Cognitive "g" and Academic Achievement "g" One and the Same "g"? An Exploration on the Woodcock-Johnson and Kaufman Tests. *Intelligence, 40*(2), 123-138.

Kearns, L.-L. (2011). High-stakes standardized testing and marginalized youth: An examination of the impact on those who fail. *Canadian Journal of Education, 34*(2), 112-130.

Klein, P. D. (1997). Multiplying the Problems of Intelligence by Eight: A Critique of Gardner's Theory. *Canadian Journal of Education, 22*(4), 377-394.

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice, 41*(4), 212-218. doi: 10.1207/s15430421tip4104_2

Lekholm, A. K., & Cliffordson, C. (2008). Discrepancies between school grades and test scores at individual and school level: Effects of gender and family background. *Educational Research and Evaluation, 14*(2), 181-199.

Lleras, C., & Rangel, C. (2009). Ability Grouping Practices in Elementary School and African American/Hispanic Achievement. *American Journal of Education, 115*(2), 279-304. doi: 10.1086/595667

Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. *American Statistician, 36*(3), 158-160.

Maple High School. (2012). Maple guide: 2012-2013.

Marchant, G. J., Paulson, S. E., & Shunk, A. (2006). Relationships between high-stakes testing policies and student achievement after controlling for demographic factors in aggregated data. *Education Policy Analysis Archives, 14*(30), 1-34.

Marzano, R. J. H. T. (2011). Grades That Show What Students Know. [Article]. *Educational Leadership, 69*(3), 34-39.

Mason. (1967). *An investigation of achievement grading based on scholastic ability distribution*. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED014132&login.asp&site=ehost-live&scope=site

Mason, & British Columbia Educational Research Council. (1965). *Studies and reports: An empirical analysis of a system of achievement grading based on the distribution of scholastic aptitude in a class*.

Mason, E. J. (2007). Measurement issues in high stakes testing: Validity and reliability. *Journal of Applied School Psychology, 23*(2), 27-46.

Mattern, K. D., Shaw, E. J., & Kobrin, J. L. (2011). An Alternative Presentation of Incremental Validity: Discrepant SAT and HSGPA Performance. *Educational and Psychological Measurement, 71*(4), 638-662.

McIntosh, S. (2012). State high school exams: A policy in transition: Center on Education Policy, .

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

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

National Center for Education Statistics, W. D. C. (2005). The Nation's Report Card: An Introduction to the National Assessment of Educational Progress (NAEP). NCES 2005-454 Revised: National Center for Education Statistics.

National Center for Fair and Open Testing. (2012). Racial Justice and Standardized Educational Testing: FairTest.

National Honor Society. (2013). Student Membership: Scholarship Retrieved 16 April 2013, from http://www.nhs.us/tabid/4018/default.aspx?topic=Student_Membership

Nevada Department of Education. (2012). Resources: Graduation requirements Retrieved 20 February 2012, from http://nde.doe.nv.gov/Resources_GradRequirements.htm

No Child Left Behind (NCLB) Act of 2001, Pub. L. No. 107-110, § 115, 1425 Stat. (2002).

Norusis, M. J. (2002). *SPSS 11.0 Guide to Data Analysis*. Upper Saddle River, NJ: Prentice Hall.

Oakes, J., & Wells, A. S. (1998). Detracking for High Student Achievement. *Educational Leadership, 55*(6), 38-41.

Office of Educational Research and Improvement. (1994). What do student grade mean? Differences across schools. Washington D.C.: U.S. Department of Education.

Oregon Department of Education. (22 December 2011). Oregon diploma: Essential skills Retrieved 22 January 2012, from http://www.ode.state.or.us/search/page/?id=1670

Oregon Department of Education. (2007). *Reliability and validity*. Salem Oregon: Oregon Department of Education Retrieved from http://www.ode.state.or.us/teachlearn/testing/manuals/2007/asmttechmanualvol4_validity.pdf.

Oregon Department of Education. (2011a). 2010-2011 Maple School District Report Card.

Oregon Department of Education. (2011b). 2010-2011 Statewide Report Card: Oregon Department of Education.

Oregon Department of Education. (2011c). Mathematics test specifications and blueprints, from http://www.ode.state.or.us/search/page/?id=496

Oregon Department of Education. (2012a, 16 February 2012). Oregon diploma: Credit requirements Retrieved 20 Febaruary 2012, from http://www.ode.state.or.us/search/page/?id=1681

Oregon Department of Education. (2012b). *State of Oregon ESEA flexibility request*.

Oregon School Activities Association. (2013). Individual Eligibility: Academic Elegibility Retrieved 16 April 2013, from http://www.osaa.org/publications/handbook/1213/02RulesAthleticCheerDanceDrill.asp#_Toc330294435

Oregon University System. (2011a). High School Transitions: Entering Freshman Profile Maple SD Class of 2010.

Oregon University System. (2011b). High School Transitions: Entering Freshman Profile Spruce SD Class of 2010.

Randall, J., & Engelhard, G. (2009a). Differences between teachers' grading practices in elementary and middle schools. *Journal of Educational Research, 102*(3), 175-185.

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

Randall, J., & Engelhard, G. (2010). Examining the grading practices of teachers. *Teaching and Teacher Education, 26*(7), 1372-1380. doi: 10.1016/j.tate.2010.03.008

Rogers, G. (2006). Do Grades Make the Grade for Program Assessment: Accreditation Board for Engineering and Technology.

Rosenthal, R., & Jacobson, L. (1968). Pygmalion in the classroom. *The Urban Review, 3*(1), 16-20. doi: 10.1007/bf02322211

Schaeffer, B. (2013). Standardized Exam Cheating in 37 States and D.C.: New Report Shows Widespread Test Score Corruption Retrieved 14 April 2013, from http://www.fairtest.org/2013-Cheating-Report-PressRelease

School Wide Information System. (2013). Average Referrals per Day per Month. Retrieved 16 April 2013 from https://www.pbisapps.org/Pages/Default.aspx

Schwartz, B., & Sharpe, K. (2011). Do Grades as Incentives Really Work. Retrieved from http://www.psychologytoday.com/blog/practical-wisdom/201101/do-grades-incentives-work

Scottsdale Public Schools, P. A. Z. (1973). *Staff utilization for continuous progress education. Math: A K-8 scope and sequence*.

Sternberg, R. J. (1983). How much Gall is too much gall?{Review of Frames of Mind: The theory of multiple intelligences}. *Contemporary Education Review, 2*(3), 215-224.

Stiggins, R., Frisbie, D. A., & Griswold, P. A. (1989). Inside high school grading practices: Building a research agenda. *Educational Measurement: Issues and Practice*, *8*(2), 5-14.

The American College Test. (2013). National Ranks for Test Scores and Composite Score Retrieved 10 March 2013, 2013, from http://www.actstudent.org/scores/norms1.html

Tomkins, C. A. (1992). Using and interpreting linear regression and correlation analysis: Some cautions and considerations. *Clinical Aphasiology, 21*, 35-46.

U.S. Census. (2011, 12 April 2011). U.S. Census State and County Quick Facts Retrieved 14 January 2011, from http://quickfacts.census.gov/qfd/states/41000.html

United States Department of Education, O. o. P., Evaluation and Policy Development. (2010). *ESEA Blueprint for Reform*. Washington, D.C.

University of Oregon. (2013). How We Evaluate Your Application, from http://admissions.uoregon.edu/freshmen

Wang, L., Beckett, G. H., & Brown, L. (2006). Controversies of standardized assessment in school accountability reform: A critical synthesis of multidisciplinary research evidence. *Applied Measurement in Education*, *19*(4), 305-328.

Wei, X. (2012). Are More Stringent NCLB State Accountability Systems Associated with Better Student Outcomes? An Analysis of NAEP Results across States. *Educational Policy*, *26*(2), 268-308.

Woessmann, L. (2002). How Central Exams Affect Educational Achievement: International Evidence from TIMSS and TIMSS-Repeat.

Woodruff, D. J., & Ziomek, R. L. (2004). *High school grade inflation from 1991 to 2003*: ACT.

Wright, W. E., & Li, X. (2008). High-stakes math tests: How "No Child Left Behind" leaves newcomer English language learners behind. *Language Policy, 7*(3), 237-266.

Yoshino, K. (2012). How to Make Course Grades Work in Measuring Program Learning Outcomes. Retrieved from http://blog.blackboard.com/company/featured/how-to-make-course-grades-work-in-measuring-program-learning-outcomes/

Zwick, R., & Sklar, J. C. (2005). Predicting College Grades and Degree Completion Using High School Grades and SAT Scores: The Role of Student Ethnicity and First Language. *American Educational Research Journal, 42*(3), 439-464.