Summer March 2016

A Study of Teacher-Buy-In and Grading Policy Reform in a Los Angeles Archdiocesan Catholic High School

Christian Martín De Larkin II
Loyola Marymount University

Summer July 2013

A Study of Teacher-Buy-In and Grading Policy Reform in a Los Angeles Archdiocesan Catholic High School

Christian Martín De Larkin II
Loyola Marymount University, mrdelarkin@gmail.com

LOYOLA MARYMOUNT UNIVERSITY

A Study of Teacher-Buy-In and Grading Policy Reform in a Los Angeles Archdiocesan Catholic High School

by

Christian Martín De Larkin II

A dissertation submitted to the Faculty of the School of Education,
Loyola Marymount University,
in partial satisfaction of the requirements for the degree
Doctor of Education

2013

A Study of Teacher Buy-In to a Grading Policy Reform
in a Los Angeles Archdiocesan Catholic High School

by

Christian Martín De Larkin II

Loyola Marymount University

School of Education
Los Angeles, CA 90045

This dissertation written by Christian De Larkin, under the direction of the Dissertation Committee, is approved and accepted by all committee members, in partial fulfillment of requirements for the degree of Doctor of Education.

July 26, 2013
Date

Dissertation Committee

[redacted]
Shane P. Martin, Ph.D., Committee Chair

[redacted]
Karen Huchting, Ph.D., Committee Member

[redacted]
Mary McCullough, Ph.D., Committee Member

ACKNOWLEDGEMENTS

This work is the product of several angels I have been blessed to know along my life journey. Each of you has impacted me in one way or another to bring me to this accomplishment. I pray that my presence in your life is as special for you as yours has been to me.

Thank you Mom, Dad, Easley, Carlos, Tami, Vanessa, Sensei Lu, Sensei Jeff, Chris, Wale, Mario, Justin, Mr. Klima, Ms. Hataishi, Mr. Kravitz, Mr. Blank, Robbie, Jon, Angelo, Michael, Max, Brian, Patrick, Malik, Tom, Father Tri, Monsignor Pilato, Ted, Diana, Betty, Jamie, Mike, Dr. Huchting, Dr. Binfet, Dr. McCullough, and Dr. Shane Martin. You are De Larkin's angels.

Iliana, you demonstrated the grace of faith by believing in me during those times that I didn't believe in myself. You supported my endeavors throughout the doctoral program and have suffered alongside with me as I battled to finish this marathon. I love you and will forever be appreciative of your support.

DEDICATION

This work is dedicated to my father, Easley Martin De Larkin Jr. He taught me how to be the man I am today. He is the reason I entered the field of education. He gave me my life and taught me what education means. He died leaving a legacy of love, intelligence, style, hope, and courage to all he met during his physical existence. Dad, I will continue to run life's marathons in your name. There is an official Dr. De Larkin in this world thanks to you!

TABLE OF CONTENTS

LIST OF TABLES

ABSTRACT

A Study of Teacher Buy-In to a Grading Policy Reform
in a Los Angeles Archdiocesan Catholic High School

by

Christian Martín De Larkin II

This study examined the construct of teacher buy-in (TBI) during a grading policy reform effort in a high school. The purpose of this study was to identify and describe teachers' perceived value to the grading reform. Additionally, the researcher studied teacher behavior by identifying the teachers' actual practice of the policy. The study finally compared the identified reported values of the participants with their actual grading practices to determine the convergence of values and practice.

The research provided empirical evidence for a new way to study TBI and its relationship to a reform implementation. This study addressed a school-site policy reform effort and described TBI contributing to, and perhaps challenging, current practices in school reform and teacher grading policies. This study described the extent to which teacher bought into the grading policies and provided a framework for studying TBI and grading policies in the context of Standards-Based Reform in the future. The findings and discussion highlight how grading policies are a critical element of the student evaluation process in the increasing movement towards national learning standards and testing.

CHAPTER ONE

INTRODUCTION

Educational challenges and opportunities face school leaders in the United States. Global competition, national learning standards, high stakes testing, and achievement disparities between the diverse demographics of the country are among some of the macro educational issues that school site leaders respond to on a daily basis. This study highlights one response to these issues by focusing on a high school's grading policy reform effort. This dissertation presents a description and analysis of a high school's effort to tackle a significant construct of curriculum and instruction. It examined reform at the site level and explored what can be generalized to the system level. Overall, this study addressed key challenges identified by teachers and administrators alike: student grading and teacher buy-in (TBI).

Over the past 10 years in the United States, and on the self-reported demographic information for the 2010 California SAT, a vast majority of students reported to be earning above average grades in high school (The College Board, 2001, 2002, 2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010). Out of the almost 200,000 students in California who reported their high school GPA on the 2010 SAT, 90% of them reported earning a 3.0 or greater high school GPA on a 4.0 scale (The College Board, 2011). Despite being self-reported, these GPA reports remained highly reliable sources of actual high school GPAs (Baird, 1976; Kuncel, Credé, & Thomas, 2005; Mattern & Shaw, 2009; Maxey & Ormsby, 1971; Sawyer, Laing, & Houston, 1988; Schiel & Noble, 1991).

The relationship between a student's high school grade point average and his or her score on the SAT is significant. Both items represent a measure that reports a similar construct: student achievement. The SAT measures students' academic skills learned in high school to construct meaning and problem solve in the areas of reading, writing, and math (Lawrence, Rigol, Essen, & Jackson, 2002). The SAT score communicates the summary of a student's level of skill in each area. A high school course grade communicates the summary of a student's achievement of the skills taught over the duration of a course (Haladyna, 1999).

Overall, a positive correlation exists between high school students' grade point averages (GPA) and their scores on the College Board SAT—a strong high school GPA projects a strong SAT score (Kobrin, Camara, & Milewski, 2002). However, this positive relationship does not exist for certain subgroups that are comprised of historically lower-achieving ethnic minority groups. We begin this study by portraying the educational achievement gap with evidence of how the overall positive correlation between GPA and SAT score failed to exist for these student subgroups within the whole student population of SAT testers.

The *2010 College-Bound Seniors State Profile Report*: *California* (The College Board, 2011a) revealed that 36% of students who took the SAT reported their race as Black, African American, Mexican, Mexican American, other Hispanic, Latino, or Latin American. In total, this group represents a population of historically low-achieving ethnic minorities. In 2010, they earned a mean aggregate SAT score of 1340, which was 301 points lower than the mean score of 1641 earned by students who indicated their race as White. Ethnic minorities scored 177 points lower than the overall mean SAT score of 1517 in the state of California, although average self-reported GPAs were over 3.0.

Table 1

Average Reported High School GPA and Average SAT Score by Race

Race	GPA	SAT
White	3.38	1641
Ethnic Minority	3.17	1333
Whole Population	3.31	1517

Note: Data taken from 2010 College-Bound Seniors State Profile Report for California (2011).

These scores support Kobrin et al. (2002) finding that ethnic minority and lower income students score at approximately one standard deviation, 34% on a normal bell curve, below what their high school GPA projects them to score in comparison to their White and higher income counterparts. High school GPA and SAT scores have an overall positive correlation only for students not considered historically low-achieving ethnic minorities. When researchers examined ethnic minority high school students' GPAs, they inadequately projected their level of performance on the SAT (Kobrin et al., 2002). This finding highlights a discrepancy, or gap, in achievement by race as predicted by high school GPA. Although students of all racial subgroups earned similar GPAs, their actual performance on the SAT is considerably lower.

The theories explaining this achievement gap have been, are, and will be continually debated for as long as the gap exists (Nieto, 2005). "These theories have positioned students in various ways as genetically inferior, culturally deprived, culturally different, economically disadvantaged, victims of structural inequality, and more." (p. 45). As a school administrator, I reject theories that blame historically underachieving students for their own lack of achievement. Deficit theories such as genetic inferiority or cultural inferiority (Nieto, 2005) fail to provide school administrators with a framework for solutions in the current context of school reform efforts to increase academic achievement.

I have written this study for school site educational leaders such as principals, administrators, and teachers who have certain influence over school policy, not influence over students' genetics or cultural backgrounds. Readers must bear this context and purpose in mind. I subscribe to the ideologies that illuminate school-based factors that can largely affect student achievement outcomes.

Interpreters have attributed the discrepancy in GPA projections between ethnic minorities and White students to school-based factors related to student grading (Kobrin, Milewski, Everson, & Zhou, 2003; Koretz & Berends, 2001; U.S. Department of Education, 1994). Kobrin et al. (2002) found that grading outcomes, when compared between higher income White students and lower income minority students, were unequally aligned with the students' actual achievement as measured on the SAT. This finding suggested that lower income and minority students were graded differently than higher income and White students.

This study addressed teacher grading practices. It presents a reform effort to standardized grading practices, challenges teachers' autonomy, and confronts teachers' subjectivity in evaluation practices. Such subjectivity has become increasingly evident in standardized tests and college entrance exams. Overall, grading practices vary considerably between teachers. Consensus does not exist on the purpose of grading and the weights given to learning criteria that determine students' course grades (Guskey & Bailey, 2001).

According to Guskey and Bailey (2001), criteria to grade students fall under three main categories: product learning, process learning, and progress learning. Assignments to assess product learning are performance-based tasks that could include writing an essay, playing a musical instrument, or taking a final exam. Teachers can objectively evaluate these activities to

measure students' learning. Process learning activities evidence students' attempts to learn and could include completing homework, participating in class, or teacher-observed effort. Finally, progress-learning criteria measure improvement students have made over time. Portfolios or pre- and post-tests (Guskey & Bailey, 2001) fall under this type of assignment. This study will concentrate on the first two grading criteria: product learning and process learning.

Combined with teachers' expectations of students, their beliefs about grading remain subjective; shaping how they teach and grade (Barnes, Bull, Perry, & Campbell, 1998). Teachers' ideologies fuel their perception, instructional practice, and evaluation of students from different racial, socioeconomic, and gender groups. Teachers implicitly and explicitly act on preconceptions, often resulting in the unjust oppression of certain students (Day-Vines & Terriquez, 2008; Langhout & Mitchell, 2008; Nieto, 2005).

This study examined the construct of teacher-buy-in (TBI) during a grading policy reform effort in a Los Angeles Archdiocesan Catholic High School that I will refer to by the pseudonym of St. Miguel Jose High School (SMJ). This school serves a primarily lower income ethnic minority population in a small city of the southern area of Los Angeles County. SMJ's enrollment of 655 is predominantly African American and Latino. The incoming first year students range in standardized test achievement levels from fifth to tenth grade reading, writing, and mathematics skills. Students matriculate into the ninth grade from over 30 Catholic, public, charter, and private elementary schools located across South Los Angeles. The average incoming freshman at SMJ is two years below academic grade level performance. Over 60% of the students in the school receive financial assistance for tuition and fees.

In the 2009-2010 academic year, the administration implemented a school-wide curriculum and instruction reform effort. The effort aimed to strengthen the school's academic rigor by incorporating standards-based instructional planning and evaluation into the curriculum. In following academic year, administration introduced new grading policies as part of the reform. These policies addressed the discrepancy between student GPA and projected SAT scores by aligning grading practices with standards-based product grading criteria. The average graduating GPA at SMJ was 3.0, and the average score on the combined SAT was 1300 out of 2400, which was over 200 points lower than the state's average in 2009 (The College Board, 2010). This research focused on the TBI to the current grading reform element of the school's overall curriculum and instructional reform.

The previous grading policy at SMJ existed from 2000-2010. During that decade, a school-wide published grade scale described the relationship between the letter grades, percent earned in class, and grade points used for student GPAs. (See Table 2).

Table 2

SMJ Grade Scale 2000-2010

Grade	Percentage	Un-Weighted Grade Points	Weighted Grade Points
A	94-100	4.0	5.0
A-	93-95	3.7	4.7
B+	90-92	3.3	4.3
B	87-89	3.0	4.0
B-	85-86	2.7	3.7
C+	80-84	2.3	3.3
C	75-79	2.0	3.0
C-	70-74	1.7	2.7
D+	68-69	1.0	1.0
D	65-67	1.0	1.0
D-	62-64	1.0	1.0
F	00-61	0	0

Note: Data taken from SMJ Faculty Handbook 2009.

Under this grading policy, teachers determined the composition of student grades. For example, a mathematics instructor could decide that process grading criteria such as following classroom rules and participating in class would be weighted anywhere from 0% to 100% of the final grade. Teachers determined all details of grading until 2010-2011, when the school transitioned towards a standards-based and mastery learning instructional model. At that time, SMJ introduced a new grade scale, online grade book program, and new grading policies. The school eliminated the "D" grade, minus and plus marks, and adopted a four-mark grade reporting system, as illustrated in Table 3:

Table 3

SMJ Grade Scale 2010-2011

Grade	Percentage	Un-Weighted Grade Points	Weighted Grade Points
A (Advanced Proficiency)	90-100	4	5
B (Strong Proficiency)	80-89	3	4
C (Basic Proficiency)	70-79	2	3
F (Not Proficient)	69 and below	0	0

This transition aimed to standardize teachers' grading practices by writing new policy in the school-wide published faculty handbook and parent/student handbook. Course assessments fell into two major categories: practice of standards assessments and mastery of standards assessments. Practice of standards assessments were student experiences or assignments presented during coursework that facilitated understanding and practice of the learning objectives. For example, an in-class assignment of 10 math problems given to students to practice solving a system of equations or taking notes on a lecture would qualify for this category. Practice of standards assessments evaluated process-learning criteria such as effort, class preparedness, and completion of assignments. These assessments allowed students to process content, but did not directly measure a student's skill level of learning the content. The new policy stated that practices of standards assessments throughout the course could be included in final coursework grades at no more than 40% of the final grade calculation.

The handbooks described mastery of standards assessments as objective measurements of a student's level of mastery of the learning objectives. For example, an independent test that measured students' skill level in solving a system of equations qualified for this category. Mastery of standards assessments included product-learning criteria such as essay writing,

playing a musical instrument, or writing final exams. The new policy stated that student performance on mastery of standards assessments should be included in final coursework grades and be weighted at least 60% of the total grade.

The new policies de-emphasized subjectively measured process learning criteria while valuing objectively measured product-learning criteria based on student performance. Under the new grading policy, process learning grading criteria such as effort, behavior, class preparedness, attendance, and assignment completion fell within the practice of standards assessments category.

The administrative team of SMJ outlined parameters for the components of student evaluation and instituted teacher training to address the risks of teacher autonomy in student evaluation. The guiding framework for the grading policies was that grades served one main purpose: to communicate student achievement (Guskey & Bailey, 2001). The transition from the previous grading system to the new system aimed to identify and communicate accurate levels of student performance with a view to strengthening future instruction. In the spring before the 2010-2011 academic school year, SMJ introduced the new academic grading policies and implemented professional development workshops with the teachers to communicate the instructional reform efforts and establish TBI for the upcoming school year.

Problem Statement

The grading policy reform effort at SMJ depended on teacher buy-in (TBI) for successful implementation. Teachers controlled and enforced grading practices. Turnbull (2002) defined TBI as teachers' level of value or perception of a school rule, policy, or change. Previous research on TBI to grading practices focused on measuring teachers' reported perceptions and

value. It suggested that reform efforts to obtain TBI and change teachers' grading practices were effective in changing their perceptions and values of grading practices (Roorda, 2008). A new facet included in this research was the inclusion of teachers' behavior. Teachers' actual practice of grading students after implementation of policy change in the school contributed to a stronger definition of the TBI construct. Studying teachers' actual practice and its relationship to teachers' perceptions of grading practices clarified the full dimensions of TBI. This study expanded the TBI definition to include the actual teacher practice of a policy combined with the level of perceived value of the policy. In other words, it identified whether teachers actually practiced what they reported to believe.

Purpose of the Study

The purpose of this study was to identify and describe teachers' perceived value of the grading reform at SMJ high school. Additionally, the researcher studied teacher behavior by identifying the teachers' actual practice of the policy to determine the convergence of values and practice. The study provides empirical evidence for a new way to study TBI and its relationship to a reform implementation. It addresses a school-site policy reform effort and describes TBI's contribution and challenge to current school reforms and its grading policies. This study describes the extent to which teachers bought-in to the grading policies and provides a framework for studying TBI in the future.

Significance of the Study

This study is important for several reasons. It suggests a more robust understanding of the elements of TBI for school reform efforts. It highlights the importance of TBI when implementing and evaluating school reform efforts that involve policies effecting classroom

instruction and evaluation. The study illuminates one of the impacts of the national standards movement on high school communities that serve historically disadvantaged and lower performing students.

This research measures and describes the components of TBI in the context of a school reform effort. Educational leaders benefit from understanding what TBI actually is and how it relates to teachers' practice of student grading for this specific study and the practice of school reform efforts in general. Aligning the beliefs, assumptions, and values of an organization with its members is imperative for successful school change (Evans, 1996). This study argues that school leaders who introduce school change involving teachers must consider TBI to adequately evaluate the effectiveness of the reform. If teachers chose not to implement the change, its effectiveness would have been minimized.

Once a successful reform is identified, accountability measures can then be implemented across the school, district, or on a greater scale. Accountability measures for school change are an important topic in today's educational climate. Public and non-public schools have implemented various accountability measures for their teachers in order to evaluate performance (Brill, 2010). This study highlighted a methodology for examining actual teacher practices of student grading. As standardized curriculum and instruction approaches increase, evaluation of these approaches becomes increasingly important to understand. Teacher unions, school leaders, policy makers, and the College Board can reference this school's reform effort and the study's content.

Grading practices and policies are key aspects of an instructor's professional responsibilities and a students' academic career. Assessment should influence instruction (Earl,

2003). The national standard debate and movement concerns all schools. National standards have been at the core of standardized assessment and evaluation. Educators see these in the forms of the ACT, SAT, state tests, and new common core assessments. This study describes SMJ's acceptance of the standards movement by standardizing its instructional practices measuring learning via the student grading policy. The study of TBI to new grading policies helps educational reformers and SMJ high school evaluate its reform effort implementation and create awareness of how its teachers respond to academic evaluation standards.

Teacher professional development, curriculum and instructional reform efforts, teacher education programs, and future studies that explore the construct of teacher grading reform in schools should utilize this work as a reference for establishing buy-in to school wide decisions and evaluating overall reform efforts. This study argues the critical reality that in order to research the effects of school change; it must have taken place and be evident.

The Loyola Marymount University School of Education conceptual framework detailed the importance of incorporating TBI and theory-based reform: "The integration of theory and practice is a dynamic and reciprocal process involving reflection and dialogue" (2009, p. 3). The premises of standards-based reform provided the framework that the administration at SMJ used to integrate theory and practice.

Contextual Framework & Guiding Premises

The national standards movement called standards-based reform (SBR) provides the context and assumptions for the reform effort at SMJ and this research project. Standards-based reform dates back to the 1980s and has been at the core of many efforts to establish national standardization throughout the United States education system. The National Commission on

Excellence in Education (1983) report titled *A Nation at Risk* highlighted the SBR premises on a national stage. The report established the framework for groups such as the National Educational Goals Panel (Vinovskis, 1999), National Council on Educational Standards and Testing (Congress of the U.S., 1991), and writers of critical national educational policy. The *No Child Left Behind Act* (2001), *Common Core State Standards Initiative* (National Governors Association Center for Best Practices, 2010), and the *Race to the Top Federal Funding Program* (U.S. Department of Education, 2009) are all rooted in the assumption that national standards are the answer to raising the achievement of all students.

The major premises of SBR are: (a) high academic expectations for all students; (b) alignment of the key elements of the educational system; and (c) assessing student achievement to measure outcomes (Hamilton, Stecher, & Yuan, 2008). The grading policy reform at SMJ and the following research for this dissertation uses SBR as its framework and rationale.

Research Questions

The goal of this dissertation is to identify and describe the components of TBI in relation to each other and the grading policy reform at SMJ high school. Three research questions guided my methodology and analysis of survey responses and grading artifacts that evidenced teachers' actual grading practices:

1. To what extent did teachers report buy-in to the school grading policy?
2. To what extent did teachers actually practice the school grading policy?
3. To what extent was there convergence between what teachers reported to buy-in to the new grading policy and teachers' actual practice of the grading policy?

Research Design and Methodology

To address all research questions, this cross-sectional study used both descriptive and correlational elements. It employed quantitative methodological approaches to collect and analyze data. The construct of TBI was comprised of two dependent variables: reported value (RV) and actual practice (AP). Reported value was the teachers' self-reported responses to survey items about the grading policies. The paper and pen survey instrument was an expanded version of the *Teacher Survey on Grading Practices* (Rich, 2001). The study calculated the result of RV using descriptive statistics of the data set to answer the first part of the first research question.

The study measured the AP of the grading policy through an analysis of two teacher artifacts: grade setup values and score sheets. The analysis utilized the school's written grading policy as the rubric to code the artifacts. It served as an objective lens and guided the researcher in his evaluation of the artifacts that revealed teacher alignment with grading policy. Each artifact was coded for its alignment with the key components of the grading policy and entered into SPSS. A further analysis of the data identified RV and its alignment with the school grading policy.

Limitations

The study noted the sample size for all results from the SPSS data analyses. A larger sample size would have yielded stronger results because outliers in the data set less influence, however the sample was limited to the number of teaching staff at SMJ which was 35.

The study is cross-sectional, limiting all survey and interview data to represent one specific moment in time. The grading reform at SMJ began in August 2010 and the study

collected data at the end of the academic year, June 2011. Teachers' RV, AP, perceptions, attitudes, and practices associated with the new grading policy could be significantly different from other times during the year. Data collection before the end of the school year was not possible due to the variability of teacher grading during that time. The end of the school year is the only opportunity to capture a complete representation of the total grading practices of the teachers.

The grade setup reports and score sheets represented a snapshot of grading practices at the end of the term and therefore only offered a limited perspective on actual grading practices throughout the academic term. Measuring each individual act of teacher grading for each individual student throughout the school year was an unrealistic task. The score sheets gave only limited insight to the grading of each individual student throughout the year because assignment dates could have been approximate rather than actual.

Delimitations

Although there are various assumptions pertaining to the achievement gap between White and ethnic minority students, the selected scope of the problem addresses in this study focuses on teacher grading policies and assumes that the*y* are a factor for low performance of minority students compared to their White counterparts. I take the position that school structures, policies, personnel, and approaches with historically low achieving demographics are the most critical influence for addressing and resolving the achievement gap through SBR. I am aware that the achievement gap conversation does include different factors including deficit thinking and anti-standards ideologies. I choose to reject these assumptions as they enable schools to accept low-achievement as the status quo and stifle reform efforts.

Organization of the Study

As the reader has discovered, Chapter one presents an introduction to the study, statement of the problem, purpose of the study, significance of the study, research questions, research method and design, limitations, delimitations, and organization of the dissertation. Chapter two presents the relevant literature that develops the rationale for this study's methodology. This includes an overview of the literature on SBR; student grading; and TBI. Chapter three includes a description of the research design, methodology, procedures for data collection, and data analysis procedures. Chapter four includes the description of the results, conclusions, and analyses used to determine them. Chapter five includes the discussion of the study's results, limitations, contributions to the field of education, and implications for further research.

CHAPTER TWO

REVIEW OF LITERATURE

This study examines the teacher buy-in (TBI) to a grading reform effort at a Los Angeles area Catholic High School. The school served a historically low-performing and primarily low-income minority demographic. The grading reform effort was a part of the school's implementation of a standardized instructional model aiming to increase the measured academic achievements of its population. The guiding premises of SBR influence these grading policies. This chapter offers a review of: (1) theoretical guiding premises of SBR; (2) student grading; and (3) TBI. This review provides a context and rationale for the study's conceptual focus and research methodology.

Theories of Student Achievement: Standards-Based Reform

For decades, educational specialists have debated competing theories to explain the underachievement of minority and lower class students as compared to their White counterparts (Nieto, 2005, Skibia, Knesting, & Bush, 2002). These theories have provided insight leading reform efforts to raise student achievement in schools and be viewed through three lenses: the student as the source; the student as the victim; and the student as the actor.

Antiquated theories of underachievement have argued that minority students' innate genetic and cultural makeup account for their school performance; these students are the natural source of their own underachievement. According to these theories, minority students are genetically predisposed and culturally programmed from birth with deficits in the academic skill sets needed to perform at the level of their White counterparts (Jenson, 1969). Critics have

argued that deficit theories solely source student underachievement in the student and fail to recognize human agency and the contextual contributors, such as the sociopolitical and socioeconomic environment (Nieto, 2005).

These critics explain that these theories position students as victims or products of their societal experiences in schools, neighborhoods, and as members of a certain social class. Context is important: historically underachieving students find themselves disproportionately overrepresented in the single parent households; poverty line income levels; high crime neighborhoods; and under resourced schools (Skibia et al, 2002). These socioeconomic disadvantages serve to reproduce the status quo, limiting the upward mobility. According to this theoretical ideology, schools are systemic tools used to reproduce the social classes for society. Explicit and implicit policies, curriculum, and other schooling-based factors control student outcomes (Spring, 1972; Bowles & Gintis, 1976). These factors include culturally biased tests and curriculum; low teacher expectations; lack of classroom facilities and classroom resources; racially charged discipline; and tracking; all of which create an unjust schooling experience among historically lower achieving students (Skibia et al, 2002).

Social reproduction and economic theories limit their focus on human agency. Resistance theory, by contrast, takes human agency into account and incorporates it into reproduction theories. These theories view students as actors or facilitators of their own academic achievement. Students recognize the systemic institutions and sociopolitical environments of which they are a part and can consciously decide to resist the unjust educational experience as a political protest. This leads to drop out and low academic achievement records. Resistance theorists argue that the agenda of the schooling system in the country does not

empower lower income and minority students to access the same opportunities as White and upper class students, resulting in their political protest (Giroux, 1983; Cummings, 1996; Kohl, 1994).

These theories provide a broader context for the specific guiding assumptions of this study. Standards-Based Reform distances itself from explaining underachievement and aims to provide a framework for providing equitable academic rigor and achievement measurement for all students. It emphasizes the schooling system's need to provide equal expectations, standards, and measures of achievement for all students regardless of race and socioeconomic atmosphere. Standards-Based Reform provides practical guiding premises that school site leaders can adopt to guide reform efforts that aim to raise student achievement.

Standards-Based Reform

Standards-based reform has been at the core of the national conversation about improving the educational system. The 1983 report *A Nation at Risk* report exposed the need for national standards and reform as a response to its presentation of evidence that American schools were failing (U.S. Department of Education, 1983). It brought the educational conversation from a local and state matter to a national concern. Since then, major national events involving education have ensued throughout each presidential administration. President George H. W. Bush organized the first national summit on education in 1989, which led to an agreement among state governors that there was a need for national education goals and a National Educational Goals Panel (Vinoviskis, 1999). That conversation transformed into legislation in 1994 when President Clinton introduced the Improving America's Schools Act (U.S. Department of Education, 1993) that required standards and aligned assessments. President George W. Bush's

No Child Left Behind Act of 2001 (NCLB) and finally President Barack Obama's Race to the Top initiative of 2009 (U.S. Department of Education, 2009) have all contributed to the current atmosphere of an increasingly national educational system that promotes the guiding premises of SBR.

Components of standards-based reform. The major guiding premises that define SBR are academic expectations for all students, alignment of the key elements of the educational system, and assessing student achievement to measure outcomes (Hamilton et al., 2008). Former Secretary of Education Diane Ravitch (1995) defined standards in her book *National Standards in American Education* as "both a goal (what should be done) and a measure of progress toward that goal (how well it was done)" (p. 7). Her idea of standards promotes objectivity that frames learning as definable and measurable. The first and most foundational aspect of SBR argues that all students should be able to meet learning expectations at a certain level of proficiency in the educational process (Codding & Tucker, 1998). Academic and performance standards are commonly used jargon of SBR (Hamilton et al., 2008) and identify with the first part of Ravitch's definition. They clearly define what should be done or learned (i.e., a student must memorize the English alphabet).

A performance standard identifies with Ravitch's second part of the definition (Ravich, 1995). It defines the level at which the content standard was performed (i.e., advanced, proficient, or basic). For example, a proficient performance for alphabet content could be defined as having 90% of the English alphabet memorized and an advanced performance could be 100%. SBR applies these content and performance standards to all students across the country, which has caused some criticism.

Multicultural education advocates explain that the content in the curriculum has historically favored student groups who have acquired certain cultural, social, and economic capital (Banks, 2004; Nieto, 2005). Opportunity and access to this capital varies between the diverse demographics of the country, adversely affecting student learning. Take the previous example and compare a student who primarily speaks Spanish at home with their family to a student who speaks primarily English. Although the same academic expectations exist, the students' lifestyles, backgrounds, and culture vary and need to be taken into account when evaluating their learning.

A second criticism of academic expectations for all students exists with performance standards. Hamilton et al. (2008) explain:

> In essence, there are no standards for developing good standards. Most advocates of SBR argue that the standards should be uniform and apply to all students, i.e., the system should adopt common expectations for everyone it serves rather than expecting higher or lower levels of attainment from some students. Most advocates of SBR also emphasize that standards should be challenging; they should stretch educators' beliefs about what students can learn. (p. 12)

The remaining SBR components are all based on the critical first component that academic expectations are set for all students. The second component insists that schools should work to align the elements and policies of the educational process to facilitate the learning of the standards. Textbooks, lesson plans, assessments and any other tools used in the curriculum and instruction should be greater aligned to the standards across teachers, classrooms, schools, districts, and states (Clune, 2001). Third, a data component emerges for SBR: the assessment of

the student achievement of the standards is included to identify and communicate student performance to inform instruction and create historical records.

St. Miguel Jose High School used SBR's guiding premises as rationale to develop its new grading policy. The policy attempted to define the students' overall level of performance in course content learning with the grades: A, B, C, and F. The levels for student grading were standardized across the entire school to align the construct of student evaluation and accurately communicate student learning.

Student Grading

Grading is an essential part of the educational process. According to Haladyna (1999), "grades are simply summaries of school achievement, typically assigned for subject matter or course of study and covering a specific time, such as a semester or other grading period" (p. 3). Grades have been communicated in various ways: school unique marks, percentages, letters, levels of proficiency, check marks, plus and minus signs (Guskey & Bailey, 2001). Grading practices across the United States have differed by district, school, and teacher. The systems of grading in many schools have operated within teachers' autonomy, which has provided an opportunity for inconsistency. The use and value of grading criteria when determining a students' final grade such as tests, quizzes, homework, class work, final exams, behavior, attendance, and participation have greatly differed from district to district in the United States (Polloway & Epstein, 1994). Grading trends have been so varied and complex that researchers have questioned their validity in the larger context of student evaluation (Marzano, 2000). This section provides an in depth overview of the major issues of grading that the administration of

SMJ encountered when operating with the premises of SBR as their foundation for its grading policy reform.

History and Reform

The complexity of grading has existed since formal schooling began in the 19th century (Guskey & Bailey, 2001; Trumbull & Farr, 2000). Scholars have heavily debated and reformed the practice of grading students over the past century and a half (Guskey & Bailey, 2001; Trumbull & Farr, 2000). During the middle of the 19th century, teachers wrote narrative summaries specific for each student, communicating their observation and understanding of a students' level of performance of skills over the course of the grading period (Kirschenbaum, Napier, & Simon, 1971). After the passing of compulsory schooling laws, class sizes increased and public schools were built across the country, exponentially increasing teachers' evaluative responsibilities (Trumbull & Farr, 2000). By the early 20th century, high school teachers abandoned the narrative grade reporting method and adopted percentage scales (Trumbull & Farr, 2000). They condensed the once detailed account of students' progress to a percent, limiting the evaluative depth of the previous narrative grading system. Although all teachers commonly used both systems across schools, the judgment of performance was always determined on an individual subjective basis from the instructor.

The concern with subjectivity of these judgments arose with Starch and Elliot's (1912) criticism of percentage grading, legitimately questioning its validity in grading essays and mathematical content. Percentage grades were too arbitrary and subjective for accurate and valid measurement. They pushed to narrow the focus of teacher grading to standard categories such as proficient, below proficient, and average. Debate on how many categories were adequate led to

years of discussion between educators on best grading practices. In 1918, a categorical letter grading system was created to rank levels of student achievement adopting the: A (excellent), B (good), C (average), D (poor), F (failing) grades. The majority of today's high schools currently used these letter grades (Trumbull & Farr, 2000).

Purposes of Grading

Differing beliefs of the purpose of grades have existed (Brookhart, 2009; Guskey & Bailey, 2001). Six common categories have been identified that summarize what teachers and administrators believe the purpose of grading and reporting in schools should be (Feldmesser, 1971; Frisbie & Waltman, 1992; Guskey & Bailey, 2001; Linn, 1983).

First, grading should serve to communicate the achievement status of students to educational stakeholders. A grade should communicate a specific meaning about a student's level of achievement. Benjamin Bloom's (1956) taxonomy of educational objectives created a common classification of a student's level of achievement and created a way for teachers to evaluate how well a student could master a learning standard. Mastery grading was based on ranking students according to their level of mastery of a subject matter during and at the end of a course (Brookhart, 2009). The student's final grade reflected failure in a course. Using grades, a teacher could formally communicate that the student needed further instruction in that subject matter to adequately meet an acceptable level of achievement.

Second, grades should serve to communicate during the actual course of study to prevent failure (Guskey & Bailey, 2001). Grades should be a formative tool. They should provide information students use for self-evaluation. Students should be able to view the grade they earned on their first test as an indicator of how they might need to improve their study habits or

reaffirm what they have achieved. Students should be able to use their grades to gauge if they are on track to pass their class, doing well in school, or qualifying for certain colleges.

Third, grades should serve to track students into groups. State, district, community, school, and classroom programs have used grades to determine which students they should enroll. Teachers often recommend the students with the highest grades to enroll in gifted or advanced placement programs whereas students with the lowest grades can be recommended for special education programs (Guskey & Bailey, 2001). College admissions decisions, financial aid decisions, athletic eligibility procedures, and achieving honor roll status are all examples of the grouping systems determined by student grades.

Fourth, grades should be extrinsically motivating. They should serve as rewards in a token economy, giving students a goal for which they can strive. They should be the carrot. Parents often can enforce this purpose of grades by incentivizing higher grades with rewards so students put in more effort into school. Students may not be intrinsically motivated to learn about the subject matter, but achieve to get the "A".. In contrast, grades can provide proof for students' lack of effort (Guskey & Bailey, 2001). The low grade serves to motivate through punishment. Teachers can threaten to give low grades if students do not show effort and acceptable behaviors.

Fifth, grades should serve to evaluate the effectiveness of instruction. Teachers who give all A grades or F grades might be investigated by their administration. Is the bell curve of grades the best representation of good teaching? Instructors can use a grade or trend in grades in their class to inform their instructional methods. A teacher could try out a new exercise or

instructional technique and use the students' grades to determine its effectiveness compared to their previous practice (Guskey & Bailey, 2001).

Sixth, and finally, grades should serve to provide evidence of lack of student effort or responsibility. One of the major aims of the school's grading reform is to obtain TBI to the first purpose of grading. Grading should be used to communicate student achievement. With SBR, grading needs to be an effective and accurate communicator of student achievement and nothing else. The other purposes of grading move the instructional model further from SBR.

Learning Criteria in Grading

Teachers can assign many types of assessments and exercises in the classroom. Assignments such as homework, class work, quizzes, tests, and portfolios are all assessment practices that can fall under three distinct learning criteria for grading practices: product, process, and progress. Learning criteria connect grading practices to the purposes of grading (Kovas, 1993).

Product learning criteria show what the student can objectively produce to show what they have learned. This learning criterion is used by educators who believe the purposes of grades are to communicate student achievement against measurable learning standards or track students (Guskey & Bailey, 2001). This is the most objective way to grade students (Ornstein, 1994). The most common uses of product learning criteria are exams, final compositions, final projects, exhibitions of work, final presentations, final portfolios, and any type of student produced measurable performance, behavior, or action that evidences learning of a content standard (Guskey & Bailey, 2001).

Process learning criteria demonstrate how the students arrive at their product learning criteria. Teachers who value this criterion believe that the learning process has value. This learning criterion is used by educators who support the belief that the purpose of grades is to extrinsically motivate students or to give proof of a students' lack of effort. The most common uses of progress learning criteria are class quizzes, student journals, student notebooks, classroom observation, homework completion, homework quality, class participation, work habits, neatness, effort, punctuality, attendance, behavior, attitude and any type of subjective student action that evidences student effort towards learning (Guskey & Bailey, 2001).

The final type of learning criterion refers to progress. This type is unique as it incorporates a diagnostic element. It inquires where a student began their learning journey in the classroom and measures the final grade on how far they progressed. Also known as the value-added approach (Wiggins, 1996), progress criteria individualizes assessment for the student, incorporating evidence from the product criteria that is designed to first diagnose, then assess how much improvement the student has made over time. Adequately giving a grade for progress is individualized. Students are compared to themselves in relation to content standards in this type of criteria. Progress criteria grading supports the belief that the purpose of grading is to give information for student self-evaluation and communicate student achievement (Guskey & Bailey, 2001).

Teachers usually incorporate all three learning criteria to some extent in their final grade marks for students. Weighting of each calculation varies significantly from teacher to teacher within and between subject matter (Guskey & Bailey, 2001). Assumptions that relate to teacher grading practices and student evaluation connect to a teachers' psyche (Morris, 2005).

Teacher Bias in Classroom Instruction

Deficit thinking is the theory that teachers' use to explain the history of widespread failure among children of color. Valencia (1997) explained that this theory speculates that students who fail in school have intellectual deficiencies and limitations that are characteristic of their race and low-income level. He also argued that this pervasive theory is at the core of the policies and practices of most educational systems that serve low-income and students of color. Teachers automatically enforce academic expectations of their students, promoting process-learning criteria due to the students' race and class.

Jean Anion's (1997) research on five schools suggested how deficit thinking exists and operates in the classroom. She observed 30 hours of instruction in a 5th grade classroom across five schools. The first and second schools served a primarily working-class population, the third served a middle-class population, the fourth school served an affluent-class population, and the fifth served an elite-class population. In the first school, the nature of student work was mechanical and routine, devaluing the need for creativity and planning. It prepared students for low-wage type of work with low academic expectations. The middle-class school rewarded students for knowing the right answers to problems, procedures, and resources. Teachers did not encourage student creativity and criticism. The affluent-class school pushed students to develop skills of negotiation, reason, and creativity that enabled them to be more autonomous learners. The final school gave students the control to manipulate the system around them by using necessary skills of inquiry, analysis, language, and reason. This study suggested that socioeconomic status of the schools and teachers' academic expectations of students positively correlate.

Mickelson (1980) observed two Los Angeles High Schools that served different socioeconomic populations. Beverly Hills High School served a majority upper-class population and Morningside High School served a working-class population. Similar to Anyon (1997), Mickelson documented students in at Beverly High School in an environment that promoted high academic expectations. They were given more autonomy when selecting classes, had schedules similar to those in colleges, were free from a dress code, and were disciplined far less than those on the Morningside campus. Administrators on the Morningside campus explicitly defended their belief that structure and discipline were preparing their students for work. Beverly Hills High School's curriculum emphasized more of a collegiate push and was more supportive of high-achieving students. The vocational programs at the two schools were vastly different in the types of vocational training offered. Morningside's vocational training programs focused on lower level blue-collar labor occupations.

Mickelson (1980) and Anyon (1997) both suggested a contrast in teacher expectations of students. Langhout and Mitchell (2008) observed the student teacher dynamics of a second grade classroom in a Title I distinguished school. Located in a mid-sized Northeastern town of the United States, Bridges Elementary School primarily served a minority and low-income population. The second grade teacher, Ms. Merlin, was a young White female who had been teaching at the school for two years. Her class demographic was 50% White and 50% Black or Latino. Eleven of her students were male. The researchers were participant-observers for a total of 96 hours over three months. They took extensive field notes with a focus on describing the student and teacher behaviors. They also analyzed a daily behavior chart kept by the teacher.

Finally, they interviewed Ms. Merlin at the end of the observation period to collect her input on the researcher's observations and data (Langhout & Mitchell, 2008).

All three sources of data suggested that deficit thinking exists in the classroom reinforcing low academic expectations for low income and ethnic minority students. Ms. Merlin explained that the behavior chart was a tool to motivate and reinforce positive behavior. In practice, the chart did not function as the teacher intended. Misbehaving students received incremental punishments. Instead of functioning as a deterrent to undesired behavior in the class, it served as a predictor and reminder for the minority students of their unnatural behavior. Minority students were more likely to have their name moved during the 96-hour period of observation. Having one's name moved on the chart did not affect an increase or decrease in having it moved again. Male minority students had the highest mean number of moves on the behavior chart over the four-month period (Langhout & Mitchell, 2008).

In addition to its failure as a behavior influence, the most negative effect of the behavior chart was evident in the academic disengagement it caused with students. The researchers observed that when students were moved on the behavior chart and punished for behavior, they became less academically engaged. Specific behaviors were noted as punishable, such as calling out of turn. One Latino male student called out of turn when he was excited by the picture of an elephant he saw in his reading book. Ms. Merlin's reprimand and punishment quickly choked his excitement. She removed the student from the reading circle for that outburst, ultimately disengaging him from the content. Researchers noted other examples of teacher-generated disengagement of minority male students.

Although the student showed signs of being on task and engaged, an optimal state for learning, his unintentional behavior to call out or exhibit anything that Ms. Merlin constituted as misbehavior became the focus of the moment (Langhout & Mitchell, 2008). This example highlights the teacher's implicit actions and an unconscious ideology of maintaining an unjust status quo. Dewey's (1956) writing further illuminated this instance: "Upon the ethical side, the tragic side of the present school is that it endeavors to prepare future members of the social order in a medium in which the conditions of the social spirit are eminently wanting" (p. 10).

Ms. Merlin's relationship with the Black and Latino male students in the classroom contrasted her relationship with the female and White students. In one instance, a young White female student pointed and called out of turn in class, stating that she liked a pin Ms. Merlin was wearing. At that time, Ms. Merlin was going over problems on the board with the whole class, but instead of reprimanding the student, she smiled, thanked the student, and then moved on with the lesson. In comparing the Latino male with the White female, it is evident that the White female was more off task with her disruption of the class. The Latino male was at least engaged with his reading material (Langhout & Mitchell, 2008).

This lack of consistency led to a reinforcement of privilege for certain groups of favored students. They knew that Ms. Merlin would let them get away with behaviors that the Black and Latino boys could not. Students in the classroom began to internalize that they were the problem and were obstacles to their learning instead realizing that the focus should really be on Ms. Merlin. Some minority students started to show resistance by continuing their "misbehavior" and yelling out the right answer. They wanted to prove to the class that they were smart, but they also wanted to resist the unfair environment in which they were trapped.

The interview and interactions between the researchers and Ms. Merlin shed light on her actual beliefs and ideology. Contrary to her practices in the classroom, Ms. Merlin did not like to discipline students and believed that all students could succeed. Her reputation in the school suffered due to her lack of acceptable classroom management. Other teachers had come by, yelled at her students for being too noisy, and questioned her authority in front of her students. She was under pressure to gain control of her class and internalized that she needed to implement tools of the hidden curriculum such as the behavior chart.

Langhout and Mitchell (2008) reaffirmed the idea of the existence of deficit thinking and its ramifications in the classroom. Hatt-Echeverria and Jo (2005) studied the racial dynamics of a newly opened charter school in an increasingly racial minority neighborhood. A majority White board of community members stated frustrations with their public school's prevalence of drugs, large class sizes, and violence and developed the Eagles Landing Charter School. Hatt-Echeverria and Jo found that White students chose to enroll at the new middle school to get away from their current public school but that the Black students chose to attend because of its proximity to their homes. The study conducted 20 hours of observations and interviewed five parents and 15 teachers to gather descriptive data on the site.

The school based its mission on the educational philosophy of *paideia*, which relates to SBR. It subscribed to the main principles that all students could learn and should be afforded a quality education. The school focused on implementing a strong academic curriculum for all students to learn while aiding in their citizenship development. Each student earned an academic grade and a citizenship grade. Overall, school employees displayed an investment in the school's mission and exhibited ownership.

On the surface, Eagles Landing Charter was an SBR model of schooling. Through further investigation, however, Hatt-Echeverria and Jo (2005) began to see dichotomous race relations. Black students explicitly called certain teachers racist; citing instances where teachers would comment on the lower standards required for Black students or would enforce dress code more stringently on Black students. Students who left the school for disciplinary or academic reasons were mostly Black. The sentiment felt from the Black students on campus was that of alienation and unequal treatment.

Hatt-Echeverria and Jo's (2005) research provided another example of the dominant ideology in practice at a newly formed school. It was evidenced in the policies of the school's student contract and implementation of the dominant instructional practices, behavior expectations, and meritocratic functions of schooling. This study highlighted a consensus theory use of the citizenship grade as a practice of socializing students to subscribe to valued normative behaviors. They add:

> The image of the orderly, conforming, hard working, and high achieving individual as a 'good citizen', was constantly reinforced in Eagles Landing School. Teachers often disciplined or reprimanded students for 'not being a good citizen' when the students were noisy, not doing their work, or not orderly. In most cases, the notion of 'good citizen' was reduced to terms of behavioral modification that students need to follow or often used as a measuring stick of discipline. (Hatt-Echeverria & Jo, 2005, p. 62)

Teachers commented that paideia only works for students who are motivated to learn and are actively engaged on their own accord. One teacher explained that if students do not accept the school's rules, policies, and culture then they are asked to withdraw as a part of their

contract. The teachers did understand that a race and class dynamic existed. One teacher commented that some parents only view the school as a daycare. Another mentioned that students needed to understand that in America, you could succeed if you put in effort. Coded language of the teacher interviews was especially revealing. Terms like "dealing with," "not motivated," and "underachievers" indirectly implied that the minority students were attending the school from the surrounding lower-income neighborhood. Expressing hopes of keeping the higher-achieving student population in the school, teachers talked of minority students leaving the area. Teachers believed that the students were the ones responsible for their failures at Eagles Landing (Hatt-Echeverria & Jo, 2005). These interviews revealed an overall alignment with deficit model theories of student achievement.

Langhout and Mitchell (2008) and Hatt-Echeverria and Jo (2005) reaffirmed the idea of the existence of deficit thinking and its ramifications in the classroom. Attempts to deconstruct and displace deficit thinking on a micro and macro level have been successful through creating accountability on low-performing schools (Skrla & Scheurich, 2001), teaching teachers about culturally responsive instructional strategies (Ladson-Billings, 2006), and staff development (García & Guerra, 2004).

Morris (2005) found that teachers viewed Latino and African American students as harder to manage and teach. Their practice corroborated this view. They wrote more referrals, issued more suspensions, and gave lower grades far more for these students as compared to their White counterparts. Adults in the school also viewed female African American students to exhibit behavior that is more masculine. Consequently, teachers imposed gender roles more heavily on them as compared to their White female counterparts. Teachers and administrators on

school grounds showed evidence of disciplining African American, Latino males and females significantly more than their Asian and White counterparts. Schoolteachers implicitly acted out their racial, gender, and class stereotypes (McLaren, 1989).

Although it is not as extensively documented as the previous research reviewed on teacher discipline and instruction, teachers' assumptions regarding student ability also influenced their grading practice. Teachers weighted process learning grading criteria such as effort differently depending on their assumption of s student's academic ability. Teachers who assume that a student is low achieving tend to give effort a stronger weight and inflate the final grade (Brookhart, 1993). This assumption of weak academic ability leads to low expectations and grade inflation. Consequently, academic expectations of their teachers shape students, who then perform at the set level of expectation. Teachers reward students with grades inflated by effort. They socially promote students to the next grade level and deceive students into believing they are academically achieving. Low academic expectations not only produce low academic achievement, but also reinforce an academic injustice of grade inflation (Thorndike, 2005). Personal philosophy of education, moral values, opinion of the student, and social aims of the school are all factors that lead teachers to mark their final grade (Guskey & Bailey, 2001).

The administration of SMJ high school developed its grading policy reform effort based on the previous overview of literature on teachers' bias and possible threats that teacher autonomy has on student grading. These findings have set the foundation for SMJ's move towards standardization of grading. The connection between teachers' assumptions about their students and grading practices is integral for understanding the future of grading policy reform. The recent emphasis on educational learning standards that can be objectively measured by

student performance has transitioned grading practices to be more aligned with a standards-based philosophy of instruction and assessment (Trumbull & Farr, 2000).

Standards-based and Mastery Grading

Product criteria such as tests and other performance-based tasks are increasingly encouraged in grading practices over process and progress learning criteria (Brookhart, 2009; Guskey & Bailey, 2001). Standards-based grading allows the communication between teachers and educational stakeholders to accurately portray the primary purpose of standards-based grading: to communicate student achievement of objective academic learning standards (Guskey & Bailey, 2001).

Creating standards-based grading models requires four steps. First, practitioners identify an established set of learning standards for each course of study. Second, they set product based learning criteria for the learning standards. Third, they classify quality of mastery for each product to rate a students' ability level. Fourth, they develop a reporting tool to communicate the teachers' evaluation of student learning of each standard (Andrade, 2000).

Although standards-based grading is the most detailed and accurate reporting of student achievement, it is not a flawless system. Parents and students can be confused by standards-based marks. In the beginning of a marking period, a student who is not proficient in a specific learning standard can be interpreted in many different ways. If they earn a grade of "not proficient" on their learning standards, it could mean that they are on track to becoming proficient by the end of the course or it could mean that the student is not on track (Guskey & Bailey, 2000). Grade marks can be misleading and fail to communicate the appropriateness of students' level of achievement. Standards-based grading can also be very tedious for teachers.

The report cards in this model break each course of study into the learning standards, requiring teachers to follow the steps of standards-based grading for each specified learning standard in the classroom (Guskey & Bailey, 2001).

Mastery level grading practices address the main concerns in a standards-based grading model. It breaks the course content into specific learning units and rates students' level of mastery of the main learning standards for each unit according to Bloom's taxonomy of learning (Bloom, Madaus, & Hastings, 1981). Formative assessments of student learning occur during and at the end of each unit to inform instructional methods and provide remediation for students in need so that they can continue to have learning experiences for mastery facilitated by the instructor (Bloom et al., 1981). Mastery grading gives the student another chance to perform. This model often works in many high schools that administer cumulative final exams that test the student on the same content standards on which they have been previously tested.

Mastery grading uses Bloom's taxonomy of learning to determine the levels of mastery. Teachers are given the autonomy to determine what level of learning in the taxonomy should be considered mastery for each learning standard in their units of instruction. This method leaves a standard of learning up for interpretation, which is one of the pitfalls of mastery grading. It limits the final grade mark to only two categories of student learning: mastery or non-mastery, clarifying student achievement to stakeholders. This method has led to positive impacts on student achievement and attitude at all educational levels (Kulik, Kulik, & Bangert-Drowns, 1990a, 1990b).

The current study evaluated the TBI to the new grading policies of SMJ, which the administration based on the premises for grading reviewed in previous sections of this study.

The goal of the grading reform is to support the school's effort to increase student academic achievement as measured by standardized testing.

Grading and Standardized Tests

Researchers have criticized the SAT as being an incomplete assessment of intelligence and reasoning (Berry, 2008). As the most researched standardized test, it is still widely used to be a valid predictor of college academic success by college admissions committees. The College Board recognizes that this test is not a complete picture nor a predictor of a student's academic achievement, but they emphasize that this nationally standardized test is the most effective way to compare students in the country since there is a "there is great variation in grading standards and course rigor within and across high schools" (The College Board, 2011b, para 1). The relationship between high school GPA and the SAT is usually described in terms of predicting college academic GPA. The College Board asserts:

> Writing is the most predictive section of the SAT, slightly more predictive than either math or critical reading. In the California study, SAT scores were slightly more predictive than high school grade point average (HSGPA). In the College Board analysis of the more than 150,000 students included in all 110 ACES studies, HSGPA was slightly more predictive than SAT scores. (The College Board, 2011c, para. 4)

Schools today are increasingly interested in the relationship that high school GPA has with SAT scores. High school grades and the SAT measure similar constructs as evidenced in their moderate correlation coefficient of .47 (Kobrin, Milewski, Everson, & Zhou, 2003). Kobrin, Camara, and Milewski (2002) surveyed 48,410 college freshmen's records to identify the relationship of their high school GPA with their SAT scores. They created three groups to

categorize this relationship; non-discrepant scores (NDS), high school GPA discrepant scores (HSD), and SAT discrepant scores (SATD). The HSD group represented students who had a high school GPA that was at least one standard deviation, 34%, greater than their SAT score. The SATD group represented the opposite.

College success of the HSD group was also discrepant. Although their high school GPAs were higher than students in the SATD and NDS groups, their college GPAs were significantly lower than the students in those groups. Lower-income, Female, Asian, Black or African American, or Hispanic students were significantly represented more in the HSD group than they were in the other two groups (Kobrin et al., 2002). These discrepancies suggest that the students in the HSD group were misled about their grades and possible victims of grade inflation or grading practices based on process learning factors. The demographic of the students also suggest the environments in which this type of grading practice occurs. Students of ethnic minority and low socioeconomic backgrounds are more prevalent in urban settings (Hodgkinson, 1999).

Researchers attributed these types of discrepancies to school based factors such as grade inflation and other unjust grading practices for ethnic minorities and urban schools (Kobrin et al., 2003; Koretz & Berends, 2001; U.S. Department of Education, 1994). This discrepancy is further evidenced with middle school students on the *National Education Longitudinal Study*. "A" students in low income urban schools earn standardized test scores equivalent to the "C" and "D" students in high-income rural schools (U.S. Department of Education, 1994).

Higher standardized test scores among higher income students is attributed to their privileged and social advantages, which allow access to complementary educational

opportunities such as tutoring over the course of a student's academic career (Willingham, Pollack, & Lewis, 2002). However, since high school GPA is a product of the teachers' practice, this study focused on the parameters within a school and teachers' control to influence student achievement. As previously noted, grading practices connect to teacher assumptions. Teacher grading practices of ethnic minority and lower income students fail to communicate accurately student achievement as compared to their White and higher income counterparts. This study aimed to examine the relationship of teachers' assumptions, implicit theories, and deficit thinking with teacher practice and buy-in to a grading policy reform at SMJ.

Teacher Buy-In

Teacher buy-in (TBI) refers to teachers' level of value or perception of a school rule, policy or change (Turnbull, 2002). Previous research indicates that reform efforts to change teachers' grading practices were effective in changing teachers' perceptions of grading practices (Roorda, 2008). Roorda examined how the implementation of professional development functioned to influence TBI to grading practice that emphasized the value of achievement-based factors. The professional development consisted of a topical overview of current assessment and grading practices, obstacles in implementing the practices, and training on the best practices. Discussions between teachers and administration created opportunities for dialogue about teacher concerns. Roorda concluded that professional development was linked to successful buy-in.

Teacher buy-in to professional development is important for understanding rationale and vision for change. Thadani, Breland, and Dewar (2010) examined how college professors' interest in and choices of professional development related to their implicit theories of teaching

skills by administering Dweck, Chiu, & Hong's (1995) measure of implicit theories on intelligence and a survey of teachers' disposition on professional development.

Entity theorists reported less interest in professional development. They also rated professional development opportunities that involved high scrutiny of their own teaching practices significantly lower than the rest of the opportunities (Thadani et al., 2010). This finding creates a rationale for understanding why teachers may not buy-in to professional development. Entity theorists by definition are naturally resistant to this idea and display less buy-in since they subscribe to the theory that attributes are fixed.

Teachers of historically disadvantaged and low performing students have voiced concerns about SBR's effectiveness for their target populations (Loeb, Knapp, & Elfers, 2008). Their survey of 400 teachers across the state of Washington revealed three major findings. First, teachers were responding to the state's reform efforts to implement standards-based instructional models in all schools. The teachers reported to be implementing the state's standards, assessments, and accountability system elements. They reported to buy-in to most of the reform policies. However, most teachers predicted that the standardization of the curriculum and instruction would result in the increased low performance of the students of color and historically disadvantaged students in their classrooms. They overwhelmingly agreed that the reform did not adequately take into account the needs of every student (Loeb et al., 2008).

Previous research on TBI to grading practices is limited in its measurement because it has solely measured teacher perceptions through survey instruments. It gives only some insight on level of TBI. A new facet of TBI that will be included in this research is the teachers' behavior. Teachers' actual practice of grading students after a professional development or policy change

and its relationship to teachers' perceptions of grading practices more adequately portray TBI. This research expanded the TBI definition to include the actual teacher practice of a policy combined with the level of perceived value of the policy. In other words, this study operationalized the construct of teacher buy-in, measuring the correlation between reported teacher opinions and actual teacher behavior in relation to the new grading policies.

Conclusion

This chapter offered a review of literature on standards-based reform, student grading, and TBI, providing a framework and rationale for the focus and methodology of this dissertation. The literature reviewed on SBR explained the guiding premises of the curricular and instructional context not only of SMJ high school but also of all high schools in the United States. Standards-based reform reinforces the standardization of student achievement evaluation (Hamilton et al., 2008). St. Miguel Jose High School responded to this by incorporating the premises of SBR into the development of its instructional reform effort and new grading policies.

The literature on student grading reviewed its history, purposes, and types of criteria used when grading students. It also presented relationships between teacher bias and unequal expectation and unfair evaluations of certain student groups, especially students of ethnic minority backgrounds (Valencia, 1997; Anyon, 2006; García & Guerra, 2004; Hatt-Echeverria & Jo, 2005; Ladson-Billings, 2006; Langhout & Mitchell, 2008; McLaren, 1989; Mickelson,1980; and Morris, 2005). The research found that most of the teacher biases were evident in their actual observed behavior in and outside of the classroom. This provided rationale for examining teachers' actual practice of the grading policy reform.

The relationship between student GPA and SAT score presented the unjust reality that

ethnic minority students' GPAs misrepresent their actual level achievement as measured by the SAT when compared to White students (Kobrin et al., 2002). It demonstrated the need for SMJ, and perhaps all schools, to develop standardized grading policies and establish TBI to address the misalignment of student GPA and SAT scores.

The literature on TBI presented a framework for applying it to grading policy reform. teacher buy-in can be established in schools with the use of professional development (Roorda, 2008). Further analysis of the literature also created awareness about a limitation of previous research on TBI. Previous research described TBI primarily on self-reported data, not actual practice, prompting this research methodology to concentrate on both self-reported data and actual practice.

After reviewing the literature on standards-based reform, student grading and TBI the following research questions emerged in regards to the reform effort at SMJ high school.

1. To what extent did teachers report buy-in to the school grading policy?
2. To what extent did teachers actually practice the school grading policy?

To what extent was there convergence between what teachers reported to buy-in to the new grading policy and teachers' actual practice of the grading policy?

CHAPTER THREE

METHOD

The administration at Saint Miguel Jose (SMJ) High School decided to implement a grading reform policy in the 2010-2011 academic year. We made this decision because students were graduating with an average GPA over a 3.0 on a 4.0 scale, yet scored approximately 200 points below the average California state score on the SAT. Research attributes this discrepancy between students' GPA and SAT scores, especially for lower middle class Black and Latino students, to teacher grading practices (Kobrin et al., 2003; Koretz & Berends, 2001; U.S. Department of Education 1994). Therefore, in an attempt to align SAT scores and GPAs at SMJ, the school administration implemented new grading policies and professional development with teachers.

The first purpose of this research was to investigate the level of TBI to the grading reform at SMJ. The construct of TBI consisted of two variables: description of reported value (RV) of the school's grading policy; and the teachers' actual practice (AP) of the grading policy. The second purpose was to examine the convergence between the RV and AP variables. In other words, to what extent do teachers' actual grading practices reflect their reported value of the grading policy? This chapter re-introduces the research questions, describes the survey instruments, explains the data collection, and outlines the analytical plan used for this study.

Research Questions and Hypothesis

Three main questions guided the analysis and investigation of this dissertation:

1. To what extent did teachers report buy-in to the school grading policy?
2. To what extent did teachers actually practice the school grading policy?
3. To what extent was there convergence between what teachers reported to buy-in to the new grading policy and teachers' actual practice of the grading policy?

To address the research questions, this quantitative study utilized surveys and artifacts from 35 teachers at SMJ during the school-wide grade reform. The methodology was determined with the goal of describing the components of TBI to the grading policies. There were two key components of the methodology. The first was the use of a survey instrument to measure RV. Survey research aims to provide descriptions by collecting data that represents people's opinions, attitudes, and general dispositions on a topic (Gay, Mills, & Airasian, 2009). The study used SPSS to describe the RV of the components of the grading policy.

The second key component of this methodology was the collection and analysis of unobtrusive school artifacts to reveal the second variable, AP. The researcher collected teacher score sheet records and grade setup values from the school's online database, PowerSchool, with permission of the school principal. These records provided key insight into the actual grading practices implemented by the teachers throughout the school year. The teacher score sheets displayed a detailed account of every assignment recorded for each student in each course. Each teacher determined and categorized assignments. For example, teachers would name an assignment "Unit 1 Photosynthesis Exam" and provide a category "Exams" in the grade book. The grade setup values represented the weights or percentages of the final grade of each grading category established by the teacher. The researcher reviewed the two artifacts to identify what

weight each teacher assigned to their self-determined categories. These categories and weights were then recorded in SPSS and further analyzed to describe the AP variable.

Context

Grade Reform at St. Miguel Jose High

SMJ is one of 50 Catholic high schools in the Los Angeles Archdiocese. It is one of 21 diocesan high schools governed by the Department of Catholic Schools. Catholic high schools in Los Angeles when compared to the rest of the nation serve over twice the percentage (69.5%) of ethnic minority students. The success of Los Angeles Catholic schools when compared to Los Angeles public schools with ethnic minority populations in Los Angeles is overwhelming. SAT scores, graduation rates, and college entrances far exceed the public school trends when comparing ethnic minority students in Catholic schools to public schools (Litton, Martin, Higareda, & Mendoza, 2010).

St. Miguel Jose High School is located in a small city in the South of Los Angeles County. The mission of the school is one committed to providing a Catholic college preparatory high school experience and developing morally aware and academically strong individuals who are of service to society. The student body is predominantly African American and Latino. Additionally, over 60% of the students in the school receive financial assistance to pay the approximately $7000 annual cost to attend. For the 2010-2011 school year, SMJ had an enrollment of 655 students.

The incoming freshman class ranged in achievement as measured by the school's High School Placement Exam (HSPT). A score on the HSPT numerically represented an academic grade level. A score of 8.5 represented an academic grade level of eighth grade fifth month. The

incoming freshmen at SMJ high school ranged in scores from 5.5-10.5 and had an average score of 7.1. In other words, the average ninth grader at SMJ was two years behind in all subject matter.

Although SMJ high school served an "at risk" population of students, it celebrated success as a college prep high school. From 2007-2011, 100% of its seniors graduated and met the minimum requirements for entrance to the California State University system. The school enforced college-prep graduation requirements consistent with the California State and University of California systems. Classes at the school were A-G approved and the course sequence guaranteed that students who graduated met the minimum course requirements for University of California and California State University eligibility. From 2007-2011, an average of 75% of graduating students were offered admissions to four-year universities and the remaining 25% enrolled in community college.

In an effort to increase four-year college acceptances for its graduates, the administration of SMJ discovered that the most significant barrier were SAT scores. For the past five years, average SAT scores of graduating seniors remained at least 200 points below the California state average for White students. These scores were the starting point for the administration's effort to develop school wide reform to increase student achievement as measured by the SAT.

Table 4

Average SAT Score Comparison for 2010 College Bound Seniors

Population	Critical Reading	Mathematics	Writing	Total
Saint Miguel Jose High	456	452	422	1330
California (All)	501	516	500	1517
California (White)	546	553	542	1641
U.S.A. (All)	501	516	492	1509
U.S.A. (White)	528	536	516	1580

Note: Data taken from: SMJ database; 2010 College-Bound Seniors State Profile Report for California (2011); 2010 College-Bound Seniors State Profile Total Group Report (2011)

As a college-preparatory school, SMJ recently focused on strengthening its curriculum through re-sequencing the courses, adding new course offerings, offering instructional development, utilizing data driven teacher evaluation, and creating grading policies. The current school-wide efforts were an attempt to bring students' standardized test scores into alignment with their GPA.

New grading policies. The questions for this study emerged from the administration's aim to evaluate its recent efforts to improve the academic achievement of its students. Since 2009, the administration had transitioned its faculty to adopt a standards-based instructional model. All faculty members were required to backwards plan by submitting semester long curriculum maps that outlined lesson plans centered on weekly objectively measurable learning standards. The school adopted the use of the California state learning standards for its subject matter content, requiring that all teachers minimally cover the state content standards of instruction over the course. Final exams and curriculum maps became departmentalized, requiring instructors who taught the same subject matter to standardize their content objectives each semester. Students taking the same course should have been able to learn the same content standards across different teachers in the school.

In the 2010-2011 academic school year, new grading policies were introduced to the faculty members to complement the standardization reform efforts. The school eliminated the "D" grade, minus and plus marks, and adopted a four-mark grade reporting system: A, B, C, and F. Each mark corresponded with the following percent scale: A: 90-100; B: 80-89; C: 70-79: F: 69 and below.

The school wide policies placed grading into two categories: mastery of standards and practice of standards. The new policy stipulated that at least 60% of the students' final grade should have consisted of mastery of standards assignments, defined as any objectively measurable student actions showing that they had mastered a learning standard for the course. This category emerged from the product learning criteria. Product learning assignment types are performance-based tasks that could include writing an essay, playing a musical instrument, or taking a final exam (Guskey & Bailey, 2001). Examples of products that would fall into the mastery of standards grading criteria are exams, tests, final projects, and other performance based summative assessments.

At most 40% of the students' final grade should have consisted of practice of standards assignments. These assignments include any student behaviors that show they have practiced mastering a learning standard for the course. This category arose from the process learning criteria, or student action that evidences a students' attempt to learn (Guskey & Bailey, 2001). Examples of process learning grading criteria are effort, class work, homework, and other similar assessments.

Professional Development. The administration of SMJ incorporated two major professional development trainings to implement the new grading policies. The administration's

approach to establishing buy-in was based in the implications from Roorda's (2008) study of TBI. Roorda linked strong TBI to a reform effort at a school site in a series of professional developments that enabled teachers to have input and the rationale for school change. The purposes of the trainings were: to create awareness among the teachers of the discrepancy between their students' GPA and test scores; educate the teachers about the importance of grading within the premises of standards-based reform; review product learning grading criteria and process learning grading criteria in student evaluation practices, and present the new grading policies in an inclusive atmosphere. The ultimate goal of the professional development series was to establish strong TBI to the new grading policies.

The first professional development occurred during the spring of the previous school year. During a weekly faculty meeting, the Vice-Principal led a presentation and discussion sharing school wide trends of the past four graduating classes' GPA averages and SAT score averages. He compared these averages to the California and national averages to highlight a key discrepancy in the school. These data fueled a discussion among the faculty, offering them the opportunity to provide input to the administration regarding the cause and possible solutions to the phenomenon.

The Vice-Principal then incorporated Guskey and Bailey's (2001) book *Developing Grading and Reporting Systems for Student Learning* as a basis for establishing a framework for standards-based grading practices. Teachers reviewed school created handouts that presented the premises of SBR and grading, and outlined product versus process grading criteria. These handouts served as a source for small group discussions during the faculty meeting to create

dialogue. The faculty received an invitation to join the administration in developing new instructional and grading policies to attempt to address the GPA - SAT discrepancy.

After initial professional development and introduction of the problem, the administrative team of SMJ met regularly to discuss and develop the grading policies. They called neighboring schools, reviewed scholarly literature, and included faculty input to decide ultimately on the policy. Administrative members and individual faculty members also met informally to provide further opportunities for feedback on the pending reform. Before the faculty left for summer vacation, they were informed that new grading policies would be introduced for the upcoming school year and that they would be trained upon their return to school.

The second opportunity for professional development took place when the teachers returned from summer vacation at the inaugural in-service. They had already received information on the new grading policies over the summer through a mailing of the updated Faculty Handbook. The second in-service reviewed the details of the grading policy in depth and encouraged overall buy-in among the faculty. The in-service allowed for individual and group discussion regarding the new policies. The teachers learned that these policies were designed to increase student academic achievement at SMJ and could only be evaluated if teachers practiced the policies. To evaluate these grading reform efforts, this study attempted to measure the reported value (RV) of teachers and their actual grading practices (AP), to determine if they truly demonstrate buy-in to the new grading policy at SMJ.

The researcher's position. I was a lead administrator at SMJ during the grading reform policy. I was a member of the administrative team that led professional development and

teachers through the grading reform effort. The teachers were aware that I was conducting research at the school site with the permission of the principal.

Participants

At the time of the study, there were 35 full-time teaching and administrative staff members working at SMJ high school. About 9% (n = 3) of the staff members held doctoral degrees and 43% (n = 15) held master's degrees. Sixty percent (n = 21) of the teachers held or were earning a California Teaching Credential. The average number of years teaching at the school was 7.86, ranging from 1-34 years. Of the 35 teachers, 43% were female (n = 15) and 57% were male (n = 20), The average age of the teachers was 31 years. Each teacher taught five courses throughout the school year with an average class size of 28 students. About 31% of the teachers had taught an Honors or AP level course at the school. Teachers taught up to three subjects from the following disciplines: Mathematics, Science, English, Social Science, Visual Performing Art, Foreign Language, and Religion. All teachers at the school site had the opportunity to participate in this study. Twenty-eight of the teachers turned in the survey. The researcher examined all teachers' score sheets and grade setup values.

Measures

The study measured TBI to the grading policies in two ways. First, it collected and analyzed self-reported survey data about teachers' beliefs, attitudes and opinions about the policies. Second, it examined teacher artifacts to reveal teachers' actual practice of the grading policies.

Survey

The survey instrument used to identify RV was comprised of items from a version of the Teacher Survey on Grading Practices by Rich (2001). Rich's survey was piloted with a focus group of teachers before he conducted an investigation of teacher reported value (RV) on achievement-based and non-achievement based grading criteria in a school site. The achievement-based grading criteria used in the Rich (2001) survey represented student work, such as tests. that directly measured their achievement. The non-achievement based criteria on his survey represented student actions that did not directly measure their academic achievement. As such, this survey is an appropriate tool for the purpose of this study and has been field tested with a similar population. Items on the survey were tailored to assess the SMJ grading policies and identify teachers' value of the new grading policies. Rich's (2001) terms, achievement-based and non-achievement based grading criteria, were replaced by the terms product learning criteria and process learning criteria in order to be consistent with this study's terminology.

Demographic information. The first part of the survey asked participants to provide demographic information such as grade level currently teaching, gender, ethnicity, subject matter department, number of years teaching, credential, and degree status. This demographic information was collected to investigate if buy-in varied by teacher characteristics.

Reported value. The next part of Rich's Teacher Survey on Grading Practices (2001) used rank order and Likert scale items designed to assess teachers' reported beliefs, level of consideration, values, practices, and opinions. These data defined RV for this study. Each portion is described below.

Beliefs about the purposes of grading. Participants read six statements that presented beliefs about purposes of grading. Chapter two of this study discussed the purposes of grading. Participants ranked the statements according to their beliefs on the purposes of grading on a five point likert scale ranging from one, most important, to six, least important. The six purposes of grading presented to the participants to rank were as follows: to communicate the achievement level of students to parents; to provide information for students to use for self-evaluation; to select identify, or group students for certain educational paths or programs; to provide incentives for students to learn; to evaluate the effectiveness of instructional programs; to provide evidence of students' lack of effort or inappropriate responsibility.

Product learning grading criteria items. This survey expanded Rich's (2001) instrument by the addition of several product learning grading criteria items. His original survey included only one item to measure academic achievement when measuring teachers' reported values. For this survey, six product learning grading criteria were added to capture the participants' value in greater detail. These product learning grading criteria were determined from examining SMJ's grading policy that provided examples such as students' performance on tests, performance on presentations, and performance on the final exam.

The survey asked teachers to report the level of consideration they gave to specific grading criteria when calculating students' final grades: final rank in class; amount of improvement displayed during the term; amount of effort put forth; following classroom rules; behavior; amount of attendance; times being late to class; participation in the classroom; turning in work on time; completion of class work; completion of homework; performance on the final assessment; performance on the exams; performance on the quizzes; performance on the

projects; performance on presentations; performance on essays/reports; amount of extra credit a student completes. Teachers responded on a five point Likert scale including *none*, *little*, *some*, *moderate*, or *substantial*. Combining item types enabled for clear and concise data analysis of product learning and process learning criteria. The six items measuring participants' level of consideration when determining final grades of product learning criteria were tested and yielded strong internal reliability (Cronbach Alpha = .887). As such, the researcher computed a mean composite variable, product learning criteria (PLC), across the six items to reflect the total product learning criteria RV for analysis.

Process learning grading criteria items. Rich's (2001) survey instrument included several process learning grading criteria items. In conjunction with Rich's survey instrument and the types written in the school's policy, the survey instrument included a total of 11 items to measure participants' RV of process learning grading criteria. Among these items were: students' participation in class; attendance to class; behavior; and effort. Teachers responded on a five-point Likert scale including *none*, *little*, *some*, *moderate*, or *substantial* (see Table 5). Combining item types enabled for clear and concise data analysis of product learning and process learning criteria. The 11 items that measured process learning grading criteria were tested and also yielded strong internal reliability (Cronbach Alpha = .810). As such, a mean composite variable, process learning criteria (PRLC), across the six items was created to reflect the total process learning reported value for analysis.

Opinions. The researcher added a section to the survey instrument that measured teachers' opinions of the new policy. Opinion items provided an opportunity for teachers to provide input specific to the overall grading reform at the school. Teachers responded to nine

opinion items on a Likert scale ranging from strongly disagree, disagree, neutral, agree, and strongly agree. The nine items that measured teacher opinions of the school grading policy were tested and yielded strong internal reliability (Cronbach alpha = .82). As such, a mean composite variable, opinion of grading policy (OGP), across the nine items reflects the total teacher reported opinion for analysis. The opinion items were included to determine further convergence with AP. Opinion items included: the current school-wide grading policies benefit our students; I agree with the current school-wide grading policies; the grading policies of my classes follow the school-wide grading policies; the grading policies of all of my classes are the same; the current school-wide grading policies are fair; all of my grading policies are fair; the current school-wide grading policies help the school achieve its mission.

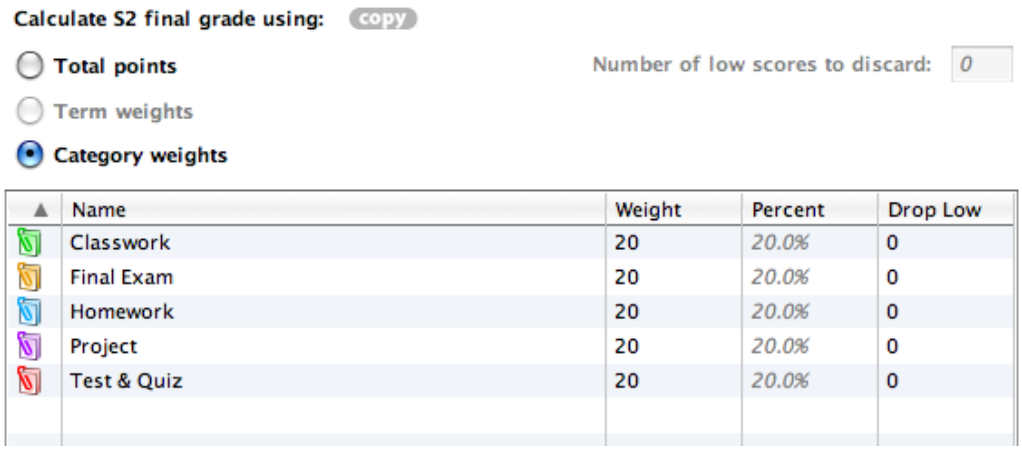
Grading criteria prioritization. The next part of the survey prompted teachers to rank the top three most important criteria they used when determining students' final grade. The final portion of the survey presented participants with a 17-item list of assessments. These terms represented the most common types of assessments identified through Rich's (2001) original survey and Guskey & Bailey's (2001) presentation of progress, product, and process grading criteria. Sample items included for ranking were effort, attendance, final exams, tests, and participation.

Actual Practice

In addition to the survey, an analysis of grade setup values and score sheets revealed the actual teacher practice of the grading policy (AP). The teacher score sheets display a detailed account of every assignment recorded for each student in each course. These artifacts represent

actual teacher practice of grading students and were categorized into two subscales: Actual Practice Product Learning Criteria and Actual Process Learning Criteria

First, the PowerSchool grade book application allowed teachers to predetermine the weight given to each assignment used in course grading and calculated the final grade accordingly. Before the course began, teachers created assignment categories such as tests, quizzes, or final exam and assigned each a percentage weight that represented the relative value when calculating final course grades. After creating categories and assigning weights on the grade setup screen, all course assignments were classified within a category. For example, the teacher below assigned a total weight of 20% to all course class work. At the end of that course, all class work was worth 20% of the final grade no matter how many class work assignments and points for class work were given during the course.



Name	Weight	Percent	Drop Low
Classwork	20	20.0%	0
Final Exam	20	20.0%	0
Homework	20	20.0%	0
Project	20	20.0%	0
Test & Quiz	20	20.0%	0

Figure 1. Example screenshot of grading setup screen from PowerSchool Online Grade Book.

In the example above, there are five teacher generated grading criteria (see Figure 1), each worth 20%. By referring to the school grading policy, test & quiz, project, and final exam would

all be classified as product learning grading criteria, while classwork and homework would be classified as process learning. For example, if a grade setup value displayed a final exam to be worth twenty-percent of the final grade as shown in Figure 1, then a variable for final exam would be created in SPSS for that teachers' AP category and a number of 20 would be inputted. All percentages for each type of category type used by the teachers were entered into SPSS.

Product learning criteria. The school grading policy described mastery of standards assessments as any measurable student actions that evidenced a level of mastery of learning standard for the course. Categories and assignments teachers presented during coursework that fit this description were included as product learning grading criteria by the researcher. Examples of product learning grading criteria found from the artifacts were exams, tests, final projects, and other types of summative assessments. All product learning grading criteria category types entered into SPSS were combined into one variable in SPSS called actual practice product learning criteria (APLC).

Process learning criteria. The school grading policy described practice of standards Assessments as any student experience or assignment presented during coursework to students that facilitated understanding and practice of the learning objectives. Categories and assignments teachers presented during coursework that fit this description were included as process learning grading criteria. Examples of process learning grading criteria include participation, effort, class preparedness, or completion of assignments. All process-learning criteria category types entered into SPSS were combined into one variable in SPSS called, (APRLC).

Scoresheet. Teachers who did not predetermine the weight of an assignment and chose to use total points for their assignments were manually coded without the category of total reports. The researcher examined their score sheets and the types of assignments, coding them as either product learning criteria or process learning criteria. Score sheets (Appendix B) were printed from the PowerSchool online grade book for review. They provided additional evidence to reaffirm the values of the grade setup screen. The researcher identified product learning and process learning grading criteria from the teacher artifacts using the school's grading policy description as a rubric in the same fashion as the screenshots of the grade setup values.

Procedure and Data Collection

Obtaining all teacher input offered the most representative data for analysis. At the yearend weekly faculty meeting in May, the Principal of SJM explained to the teachers that the school was conducting research to assess teachers' opinions about the new grading policy. The purpose of the research was to gain input from the faculty about their grading practices and opinions of the new school grading policy to evaluate how to improve the policy for the next school year. Teachers were notified that the survey would be completely anonymous and would not be used in their evaluations.

Each teacher was asked to be a part of the study and fill out the survey. We administered the pen and paper survey in the school library. Each faculty member was asked to turn the survey into a box to maintain anonymity. The majority of the teachers, 28 out of 35, turned in the survey portion of the research at the end of the meeting. The rest noted that they would complete the survey on their own time later, but they did not return the survey.

To access the school artifacts, the principal granted access to the online PowerSchool Gradebook database. As standard grade recording procedure, teachers submitted grades using the PowerSchool Gradebook program. The school's online database kept teacher score sheets and grade setup values The researcher obtained these artifacts by taking screenshots of the grade setup values downloading reports of the teacher score sheets. They were examined to measure AP. This teacher-generated data included names and descriptions of assignments, student scores, and teacher comments and was the most representative of teachers' grading practices. Before analysis, the researcher removed all teacher names from each artifact to maintain anonymity.

Analytical Plan

To answer the first objective of this research and determine teachers RV of the grading policy, the researcher performed internal reliability tests to determine if item types on the survey could be combined. He then assessed RV of the new grading policy by examining mean responses of the two types of items on the survey, product learning criteria items (PLC), and process learning criteria items (PRLC).

An independent samples t-test compared the teachers' overall value of these two criteria to determine which criterion was valued more heavily and if the mean values significantly differed. Frequency distributions, ANOVAs and correlations were then performed in order to investigate if teachers' value of PLC and PRLC differed within their demographic subgroups of gender, ethnicity, years teaching, credential status, and grades taught.

To answer the second question of this research, the grade setup value screens and score sheets were examined to measure teachers' actual practice of the grading policy. The school

grading policy served as the rubric and lens to guide the researcher through the coding of each type of artifact. The researcher coded each artifact by identifying product learning and process learning categories present in the grade records.

Once identified, the weight of each category was recorded in SPSS for further analysis. The researcher ran descriptive statistics to describe the teachers' actual value of the new grading policy. He assessed the two types of categories present, product learning and process learning. Similar to RV, an independent samples t-test was run to determine which type of grading criteria, product learning or process learning, teachers reported to value more in their actual grading practices. Next, the researcher compared these findings to the school's new grading policy parameters, which set a 60% weight for product learning grading criteria and 40% weight for process grading criteria for students' final grades.

To answer the third research question regarding convergence between reported value and actual practice, the findings from the survey and artifact data were compared to determine if product learning or process learning criteria were both similarly described. The researcher calculated descriptive statistics, frequencies distributions, ANOVAs, and correlations of teacher opinion items on the survey in order to investigate to what extent teachers agreed with the grading policies within their demographic subgroups such as: gender, ethnicity, years teaching, credential status, and grades taught. These tests contributed further comparison and convergence to the data.

Conclusion

This chapter presented the context of the research site, introduced the measures used to collect data, outlined the data collection procedure, and described the analytical plan used to

describe the data. The research for this dissertation was conducted at a Los Angeles Archdiocesan Catholic high school with 35 teachers. It surveyed the teachers to measure their buy-in to the school's new grading policies. Grade book records identified actual practice of the grading policies. The researcher coded all data was coded and used SPSS to perform inferential statistics. The results section of this study will present the findings and answer the research questions.

CHAPTER FOUR

RESULTS

The purpose of this dissertation is to provide a deeper understanding of teacher buy-in to a school reform effort. The faculty at Saint Miguel Jose Catholic High School participated in survey research to provide self reported values and opinions of the new grading policies. The teachers' grade records were then examined and coded to reveal their actual grading practices. This chapter of the dissertation will provide the findings to the research questions.

1. To what extent did teachers report buy-in to the school grading policy?
2. To what extent did teachers actually practice the school grading policy?
3. To what extent was there convergence between what teachers reported to buy-in to the new grading policy and teachers' actual practice of the grading policy?

Reported Value

To answer the first research question, the researcher performed an independent samples T-test to examine mean differences between teachers' reported buy-in to product learning grading criteria and teachers' reported buy-in to process learning grading criteria. A significant finding emerged. On a five-point Likert scale, product learning grading criteria were significantly valued ($M = 4.17$, $SD = .74$) more than process learning grading criteria ($M= 3.48$; $SD = .76$) $t\ (28) = 29.6$, $p < .01$. In other words, teachers self-reported on the survey that they valued product-learning criteria, such as student performance on the final exams or average performance on tests, more than process learning criteria, such as effort or completion of class work. The teachers' reported values consistent with the aspects of the grading policy and

suggests TBI. The school grading policy aimed to deemphasize the overall value of process learning criteria when compared to product learning criteria in determining students' final grades.

Table 5

T-Test Results for Reported Value Given to Grading Criteria

Variable	*N*	Mean	Std. Deviation	Sig. (2-tailed)
Product Learning Criteria (PLC)	28	4.17	.75	.00
Process Learning Criteria (PRLC)	28	3.48	.76	.00

Note: Likert scale numerical representations: 1 = *none*, 2 = *little*, 3 = *some*, 4 = *moderate*, 5 = *substantial*; *N* = number of responses.

The first research question aims to describe to what extent the teachers reported buy-in. Comparing means and identifying relationships of RV among participant demographic groups would allow for a sense of whether reported buy-in differed by participant characteristics. The comparison provided a more nuanced answer to the first research question. So, the researcher performed ANOVAs and correlations. These would reveal significant differences, trends, or correlations for reported buy-in between participants' reported demographic subgroups.

One significant correlation emerged from participants' demographic makeup and their reported buy-in. A Pearson's r correlation coefficient was computed to assess if a relationship existed between number of years teaching and the two variables: RV of product learning grading criteria and RV of process learning grading criteria. A moderately positive correlation emerged between the number of years teachers' taught at the school and teachers' reported value of the grading policies [$r = .43$, $n = 28$, $p = .021$]. The longer a teacher reported to have been teaching, the more they valued product learning grading criteria. This finding suggests that there is an association between teaching experience and buy-in to the policy.

Means for reported buy-in did not significantly differ for the ANOVAs performed. There was no significant difference between the values of the two grading criteria by ethnicity, teachers' subject matter department, level of education attained, and credential status. This finding suggests that the demographic makeup of the teachers with the exception of teaching experience was not associated with the extent of their buy-in.

Table 6

ANOVA for PLC and PRLC by Demographic

Demographic Type	*df*	Mean Sqr.	*F*	Significance
Race				
PLC	4	.18	.29	.89
PRLC	4	.66	1.2	.35
Subject Matter				
PLC	5	.74	1.45	.25
PRLC	5	.92	1.86	1.43
Level of Education				
PLC	2	1.1	2.14	.14
PRLC	2	.8	1.4	.26
Credential Status				
PLC	1	.44	.76	.39
PRLC	1	.77	1.4	.25

Notes: Data reported is for between subjects analysis of variance.

Ranking of Purpose and Grading Criteria

The researcher calculated frequency distributions in SPSS to identify participant ranking of the six purposes of grading. Seventy-one percent of the participants' rankings (n = 25) chose the purposes that aligned with SBR as the first or second most important purposes of grading.

Table 7

Frequency Ranking for Purposes of Grading

Purpose	Mode	Frequency	Percent of Sample
Communicate the achievement level of students to	1	12	42.8
Provide information for students to use for self-	1	12	42.8
Select, identify, or group students for certain educational paths or programs	2, 6	6	21.4
Provide incentives for students to learn	4,6	6	21.4
Evaluate the effectiveness of instructional programs	4	8	28.6
Provide evidence of students' lack of effort or inappropriate responsibility	6	11	39.3

Table Notes: Mode denotes most frequently chosen rank of purpose(s). Frequency denotes amount of times mode value occurred.

Twelve participants chose "communicate the achievement level of students" and "provide information for students to use for self-evaluation" as the two most important purposes of grading. The rest of the purposes of grading provided on the survey were more frequently ranked lower. Overall, the teachers agreed with the purposes of grading that aligned with the school's grading policies.

Out of the 17 assignment type items presented for rank ordering, a frequency distribution from SPSS revealed the top four most frequently ranked criteria.

Table 8

Most Frequently Ranked Assignment Types

Assignment	Mode	Frequency	Percent of Sample
Tests	1	22	78.57
Effort	1	16	57.14
Class Work	3	16	57.14
Final Exam	3	15	53.57

Note: Mode denotes most frequently chosen rank of assignment among top choice from 1-3. Frequency denotes amount of times assignment was chosen by participants as a top choice.

Tests were most frequently chosen (n = 22) among participants from 1 to 3 representing their top choices for grading criteria. Effort and class work tied for the second most frequently ranked criteria (n = 16), and the final exam was the third most frequently ranked (n = 15) grading criteria. Overall, teachers more frequently chose product-grading criteria than purposes of grading. This aligned with the school's grading policies.

Actual Practice

To answer the second research question regarding teachers' actual grading practices, screenshots of the grade setup screen from the online PowerSchool Gradebook and printed score sheets identified actual grading practices. The data were coded from the score sheets and entered into SPSS for analysis. Two variables were created: actual product learning criteria (APLC) and actual process learning criteria (APRLC). Each variable represented the actual percentages that teachers weighted for product learning grading criteria and process learning grading criteria in calculating students' final grades. Once the researcher created both variables, he performed a t-test was performed to compare means of APLC and APRLC. Again, a significant finding emerged $t(36) = 28.84$, $p < .01$). Product learning grading criteria was significantly valued (M = 61.81, SD = 12.5) more than process learning grading criteria (M = 37.74; SD = 12.8).

On average, teachers weighted product learning grading criteria as 61.8% of the final grade. The school's policy stated that product learning grading criteria within the mastery of standards grading assessments should have been weighted at least 60% of the final grade. Furthermore, teachers weighted product-grading criteria to be 37.7% of the final grade. The school's policy stated that process learning grading criteria within the practice of standards

grading assessments should have been weighted at most 40% of the final grade. This finding suggests that teachers' actual practice aligned with the school's grading policy.

Convergence

To answer the third research question regarding convergence between the buy in that teachers reported and their actual practice of the grading policy, the researcher compared both the survey results and data from teacher artifacts. Convergence refers to the degree to which scores on two measurements of the same construct are related. For example, if student scores on the SAT are similar to student scores on the ACT, then the two tests have convergence. In this study, two measurements of the construct TBI were used. Examining reported buy-in of the grading policies (RV) through a survey instrument and actual practice of the grading policies (AP) through examining teacher artifacts revealed similar results.

Actual Practice

Findings confirm that the participants valued the product learning grading criteria significantly more than the process learning grading criteria, ultimately aligning with the goals of the grading reform. The measurement of teachers' actual practice clearly documented the teachers' grading practices throughout the course providing evidence consistent with the results determined from the analysis of the survey instrument.

Reported Value

Teachers' self-reported data across several portions of the survey supported this argument. First, they reported to consider product learning grading criteria significantly more than process learning grading criteria. Second, teachers rank-ordered product learning grading criteria higher and more frequently than process learning grading criteria when asked about

determining students final grade. Third, teachers agreed with the purposes of grading that most aligned with the premises of the school's grading policy reform. The opinion item portion of the survey yielded further evidence of TBI.

Opinions. On a five-point Likert scale, teachers rated their level of agreement with opinion statements about the school grading policies.

Table 9

Participants' Reported Agreement with Grading Policy

Opinions	Mean	Std.
The current school-wide grading policies benefit our students.	3.78	.93
I agree with the current school-wide grading policies.	3.93	.83
The grading policies of my classes follow the school-wide grading policies.	4.44	.51
The grading policies of all of my classes are the same.	4.33	.78
The current school-wide grading policies are fair.	4.00	.73
All of my grading policies are fair.	4.15	.82
The current school-wide grading policies help the school achieve its mission.	3.74	.81
The current school-wide grading policies are easy to implement.	4.11	.70
The current school-wide grading policies improve the way I teach.	3.52	.89

Teachers' overall agreement ($M = 4$, $SD = .5$) with the composite variable, OGP, was strong. There was also a moderately positive correlation between the number of years teachers taught at the school and their reported opinion of the grading policy [$r = .547$, $n = 28$, $p = .003$]. The longer a teacher reported to have been teaching, the more they reported to agree with the grading policy. This finding suggests that teaching experience is an area to further investigate in relation to teachers' opinions of standards-based grading practices. Teachers' reported buy-in (RV) to the grading policy was evident in the results on all portions of the survey instrument

suggesting overall construct validity and convergence with the teachers' actual practice (AP) of the grading policy.

Conclusion

This chapter presented findings from the analysis of data collected through survey and teacher artifacts. All data collected aimed to reveal the TBI to the grading policy at SMJ high school. The data and findings as presented and analyzed were divided into two categories to adequately answer the research questions. The researcher reported and explained teachers' attitudes, beliefs, opinions, and actual practice regarding product learning grading criteria and process learning criteria. Overall, teachers' reported to buy-in to the grading policies. Furthermore, their actual practice of the policies also aligned. Chapter five will discuss the significance and implications of the findings.

CHAPTER FIVE

DISCUSSION

The overall goal of this dissertation was to provide empirical evidence for a stronger definition of TBI and its relationship to a reform policy. This dissertation examined a school site policy reform effort and operationalized the definition of teacher buy-in as the strength of the convergence between reported teacher value and actual teacher practice in relation to the new grading policies.

The first purpose of this research aimed to identify and describe teachers' RV to the grading reform at SMJ high school. The second purpose of this research was to study teacher behavior and identify the teachers' AP of the policy. The final purpose of this study was to examine how the identified RV of the participants and their AP converged. Ultimately, these purposes combined to assess if the reform implementation was successful in achieving TBI. This chapter will review the research questions that guided this work, summarize the findings, discuss their significance, present contributions of this work, and provide recommendations for future research.

The research questions for this dissertation were:

1. To what extent did teachers report buy-in to the school grading policy?
2. To what extent did teachers actually practice the school grading policy?
3. To what extent was there convergence between what teachers reported to buy-in to the new grading policy and teachers' actual practice of the grading policy?

Research Question 1

The school's grading policy was based in the premises of SBR and focused on product learning versus process learning grading criteria. To fully capture teachers' RV of the grading policy, the study expanded a version of Rich's (2001) Teacher Survey of Grading Practices by adding new portions and items that measured teachers' opinions of the purpose of grading, levels of consideration of grading criteria, general opinions of the new policy, and priorities of grading criteria. A deeper investigation of reported TBI to the grading policy resulted in utilizing the new survey instrument.

The current findings evidenced that the teachers at SMJ reported to buy-in to the grading reform on every portion of the survey instrument. These findings suggest that the administration of SMJ was successful in its approach to establish TBI to the new grading policy reform by providing professional development and offering opportunities for teachers to provide feedback. Inter-item reliabilities emerged among survey items. Product learning grading criteria items, process learning grading items, and opinion items all yielded strong inter-item reliability within each type, enabling new variables to be created and tested in SPSS. The three occurrences of inter-item reliability from the survey items suggest strong overall reliability of the new survey instrument. Furthermore, the convergence from the survey results with the AP of the policy suggested construct validity for the survey instrument.

Findings for this question are consistent with previous literature. Roorda (2008) found that teachers who were subjected to professional development experiences about a school policy change reported buy-in to the school's reform effort. Assessing the buy-in teachers reported

during a school reform effort to change its grading policies established the necessary foundation for the development of a more complete definition of TBI.

Teachers' RV of product learning and process learning criteria items on the survey was compared within demographic subgroups of the sample to examine if significant differences or trends emerged. Teachers reported buy-in to the policy on the survey items of product grading criteria positively correlated with the number of years a teacher reported to be teaching. Due to a small sample size, $N = 28$, this finding should be noted as a potential association to investigate in future studies of TBI. A new question emerged from this trend: To what extent does the amount of teaching experience relate to teachers' reported value of product learning criteria?

Although the researcher further developed the survey instrument to measure a deeper understanding of TBI to the new grading policies, using self-reported instruments did not provide data of actual teacher behavior. Teachers may have reported to buy-in to a policy, but they may not have actually practiced it. This study provided an attempt to address this limitation and strengthen educational research on the second TBI question.

Research Question 2

Examining the teacher artifacts to identify teachers' actual practice of grading students provided rich data for this study and for all research on TBI. The data collection strategy was innovative and important to understand. Actual teacher behavior is recorded at school sites in grade book score sheets and other archival data. As a result of the increased use of online grading platforms in schools, teachers' grading practices can be conveniently obtained through computer access. The data collection procedures described in Chapter three of this dissertation are useful tools for educational leaders to consult and understand.

Online data is increasingly accessible and organized as schools use modern online educational software. This study realized some of the implications of online educational software databases, especially for those schools that employ the PowerSchool online grade book. This program standardizes the process for grade record keeping across schools, establishing a strong source for data collection and analysis. A significant amount of the data accessed for this study was through the school's computer programs using screenshots. These screenshots were captured at the school site then transferred to the researcher's computer storage for further analysis of teachers' actual grading practice.

Examination of teachers' actual grading practices of students after the implementation of the school grading policies converged with the findings from the survey instrument. Actual grading practices of students at SMJ aligned with the school's grading policies. Several implications and questions of this finding surfaced.

The alignment of teachers' actual practice of the grading policy suggests that teachers not only psychologically bought-in to the grading policies but also implemented the policy. Implementation of the policy required teachers to know how to grade according to the policy and to decide to grade according to the policy. If teachers' actual practice did not align with the policies while they reported to buy-in to the policies, then a discrepancy between RV and AP would exist, perhaps because teachers failed to understand how to implement the grading policy. This possibility would imply that professional development was only successful in buying the teachers into the policies in theory, but not practice. However, the current findings suggest that teachers' AP of grading students aligned with the school's grading policies.

Studying the actual practice of the teachers occurred after the implementation of the grading reform effort and policy change. How did the teachers' actual practice of grading students change after the grading reform compared to before? Studying the teachers' actual practice of grading before the effort and comparing it to their actual practice after the effort would provide insight on the relationship between the reform efforts and TBI.

Research Question 3

Convergence of teacher reported buy-in to the grading policies and actual grading practice of the policies was evident in the data analysis. This finding supported the claim that the teachers at SMJ bought-in to the grading policy. Teacher buy-in had been previously defined as teachers self-reported level of value or perception of a school rule, policy or change (Turnbull, 2002).

A notable element of this study is the study of both teacher RV and AP. Simply studying RV would not fully reveal if teachers' actually practiced the policy and implemented the change. Simply studying practice does not fully capture the construct of TBI because practice of a policy does not reveal beliefs. The current data provides evidence to create a multidimensional working definition of TBI for future educational research.

This finding supports this dissertation's definition of the construct of TBI as the strength of the convergence between teacher RV and teacher AP in relation to the new grading policies. This definition of buy-in offers a stronger understanding because of its inclusion of the teachers' actual behavior. By studying both RV and AP, the findings of the study describe a more complete representation of TBI.

Contributions and Future Recommendations

Teacher Buy-In

A new idea of TBI has been established through the evidence of convergence. The convergence of the RV of the grading policies with the AP of the policy created a relationship between self-reported perceptions and actual behavior. Actual teacher behavior provided insight to teachers' ideology, assumptions, and expectations of students. Teachers' beliefs about grading combined with their expectations of a student are all subjectively rooted and are taken into account when teaching and grading a student (Barnes et al., 1998). In their classrooms, teachers implicitly and explicitly act on preconceptions (Day-Vines, 2008; Langhout & Mitchell, 2008; Nieto, 2005).

This study has created an understanding of teachers' beliefs and assumptions about their students, the grading reform, and the premises of SBR. The guiding premises of SBR, academic expectations for all students, alignment of the key elements of the educational system, and assessing student achievement to measure outcomes (Hamilton et al., 2008) comprised the framework for creating the grading policies. The findings from this study suggest that teachers successfully bought into the guiding premises of SBR by reporting buy-in on the surveys and implementing the grading policies.

Methodology

Future research on grading policies can utilize this study's survey instrument to yield significant results. The survey instrument used was an expanded version of Richard Rich's (2001) Teacher Survey of Grading Practices. Findings of convergence with teachers' AP of the policies confirms Rich's (2001) study and suggest construct validity of this study's survey

instrument. Convergence also provides reason to believe that the data collected from the survey instrument represent not only the teachers' RV of the grading policies, but also their AP.

Future research on grading policies and school leaders can utilize this study's data collection procedure as a reference. School leaders can determine the trends in their teachers' grading practices by using this study's method of analyzing teacher score sheets and grade setup values. School leaders can investigate to what extent the teachers at their sites grade students based on product learning or process learning criteria. Having access to grading trends would provide insight and data to examine, discuss, and analyze so that future grading reform efforts or conversations in a school could be tailored specific to the identified trends. Furthermore, teachers' grading practices could become incorporated into teacher evaluations.

Currently, teacher observations are standard evaluation procedures used in high schools. Most of the research presented in the literature review was based on teacher observation data. These observations provided observers with teachers' actual instructional practices during class time to evidence if teachers are planning lessons, engaging students, and exhibiting behaviors that are deemed "good teaching" by the observer. Observational data provides rich data to the observer and was mentioned several times in the literature presented in Chapter two to highlight how teachers' acted on preconceptions inside of the classroom.

The data collection approach employed in this study revealed teachers' instructional practices outside of class when evaluating students. Schools that operate from a standards-based curricular framework will benefit from the process of data collection and analysis used in this study. Grade records provide rich to data to the school leader and give the school leader insight into evaluation practices. If in the future research is published on "relevant" or "effective"

grading policies, school leaders could utilize the collection and analysis of teacher actual grading used in this study to evaluate practices based on research.

The findings from the quantitative research approach employed for this dissertation should be further investigated through a qualitative approach. Future research on TBI and grading reform will yield more robust findings by incorporating in depth interviews or focus groups with teachers and provide a deeper understanding to the original research questions and survey instrument items. Qualitative research methodology would enable an exploration of open-ended questions that emerged from the findings of this study: Why did teacher report to buy-in to the grading policy? Why did teachers actually practice the policy? What did teachers think of professional development? What were the teachers' experiences and thoughts incorporating the school grading policies? Are there any changes teachers would make to the policies? A qualitative investigation of these questions and of TBI will offer future school reformers insight to consult and utilize when planning for school change.

School Reform

This dissertation documented an attempt at school reform. The school leaders identified a research-based reform plan that incorporated stakeholders in the organizational decision-making process. They communicated with teachers at SMJ in a transparent way several times during the development of the grading policies. They presented the problem first in the form of historical data that evidenced the misalignment between student GPA and SAT scores. Next, they implemented discussions and workshops with the teachers to outline solutions to the problem and encourage buy-in to proposed solutions. The teachers participated during the reform and, as established by the findings of this study, they bought in. This dissertation's

account of the implementation of school reform provides other school leaders with a framework to utilize in their own sites.

It is important to note my dual role as researcher and administrator at the school during the time of this study. My responsibility at the school during the time of the research was to supervise the teachers. We can only speculate about the effects of my role during the study. Questions for further research on school reform could bring to light the association of the researcher as practitioner. Would the findings have been different if the researcher had not worked at the school site and aided in the implementation of the reform effort?

The findings from this research provide hope for educational leaders at SMJ high school. Their reform effort to respond to ethnic minority students' GPAs misaligned with SAT scores generated strong TBI. Since actual practice aligned with the grading policies, data on student GPA and future SAT scores can now be collected, analyzed, and evaluated throughout the following years. Using the methodology established in this study, we can continue to monitor the relationship between grading policies and practices. Future research at the high school and for other educational leaders can focus on the effectiveness of the components of the grading policies designed to align GPA with SAT scores among ethnic minority students.

Final Thoughts

The standards-based movement is on the national stage in an unprecedented and powerful way. States have adopted common core standards and assessments and the educational landscape is becoming increasingly standardized. School leaders must consider this context and respond to the standards movement in order to better serve students, especially ethnic minority students that are consistently performing below their White and higher income counterparts.

This dissertation can be useful for school site leaders in their grading reform efforts. The achievement disparities between lower income ethnic minority students and higher income White students continues to exist in the context of global competition, national learning standards, and high stakes testing. However, this disparity can be addressed and even abolished by changing one school at a time.

This research topic provides a response the underlying significance of previous research on grading. The College Board have released several research articles evidencing the inconsistency of high school grading (Camara, 1998; Camara, Kimmel, Scheuneman, & Sawtell, 2003; Godfrey, 2011; Kobrin et al., 2002; Mattern, Shaw, & Kobrin, 2010; Ramist, Lewis, & Jenkins, 1997). The underlying theme of these publications advocated that high school grading is an inconsistent and invalid measure of student achievement, therefore promoting the College Board SAT and other standardized assessments. This history of research along with the many SAT validity research studies, (Kobrin, Patterson, Shaw, Mattern, & Barbuti, 2011), consistently established the rational for colleges and universities to strongly weigh students SAT score in college admissions.

This study attempted to address student grading at a school site in an effort to spark the dialogue needed among schools, districts, states, and the nation about grading reform. Future research on grading practices is needed to identify reliable and valid student grading practices and policies that align with achievement, predict college success, and offer an alternative method for evaluating students.

Structures and instruments of the schooling system must be examined to identify how they influence and ultimately teach students to internalize particular ideologies, behaviors, and

achievement outcomes. We must create dialogue, awareness, policy, an understanding of resistance, and the process of liberation (Apple, 1980). It is our hope that school leaders will believe that local change can and will be the catalyst for successful reform in the context of national standards. This dissertation offers a study of teacher buy-in and grading policy reform to begin answering the real question for school leaders of the United States. The question is not what do we teach our children; it is how do we teach them?

Appendix A

St. Miguel Jose High School Grading Practices
Teacher Questionnaire
Adapted from Rich's Teacher Survey on Grading Practices

BASIC INFORMATION

Grades I Currently Teach: 9___ 10___ 11___12__ **Years Teaching: _____**

Sex: Male ___ Female___ **Ethnicity: _____**

Department(s) _____

1) I am currently teaching honors level/AP courses. ☐Yes ☐ No

2) I have a California Teaching Credential ☐Yes ☐ No

3) Highest Level of Education Degree Earned: _____

PART I

How should grades be used at this school? Please rank your value of importance of the following 6 purposes of grading from 1-6 (1 is most important):

____ To communicate the achievement status of students to parents and others.
____ To provide information that students can use for self-evaluation.
____ To select, identify, or group students for certain educational paths or programs.
____ To provide incentives for students to learn.
____ To evaluate the effectives of instructional programs.
____ To provide evidence of students' lack of effort or inappropriate responsibility.

Directions: Teachers consider many factors when determining a student's final grade. This survey will be used to better inform the administration of best grading practices you utilize in your classes to evaluate the school's grading policies for the upcoming school year. In no way will your responses affect your evaluation. ***Please check the box that most corresponds with your opinion on the level of consideration that should be given to each criteria below.***

Criteria Used when Grading a Student	**Level of Consideration**				
	one	*ittle*	*ome*	*M oderate*	*Sub stantial*
Final rank in the class					
Amount of improvement displayed during the term					
Amount of effort put forth					
Following the classroom rules					
Behavior					
Amount of attendance					
Times being late to class					
Participation in the classroom					
Turning in work on time					
Completion of class work					
Completion of homework					
Performance on the final assessment					
Performance on the exams during the term					
Performance on the quizzes during the term					
Performance on the projects during the term					
Performance on the presentations during the term					
Performance on the essays/reports during the term					
Amount of extra credit a student completes					

Please check the box that most corresponds with your opinion on the level of consideration that should be given to each criteria below.

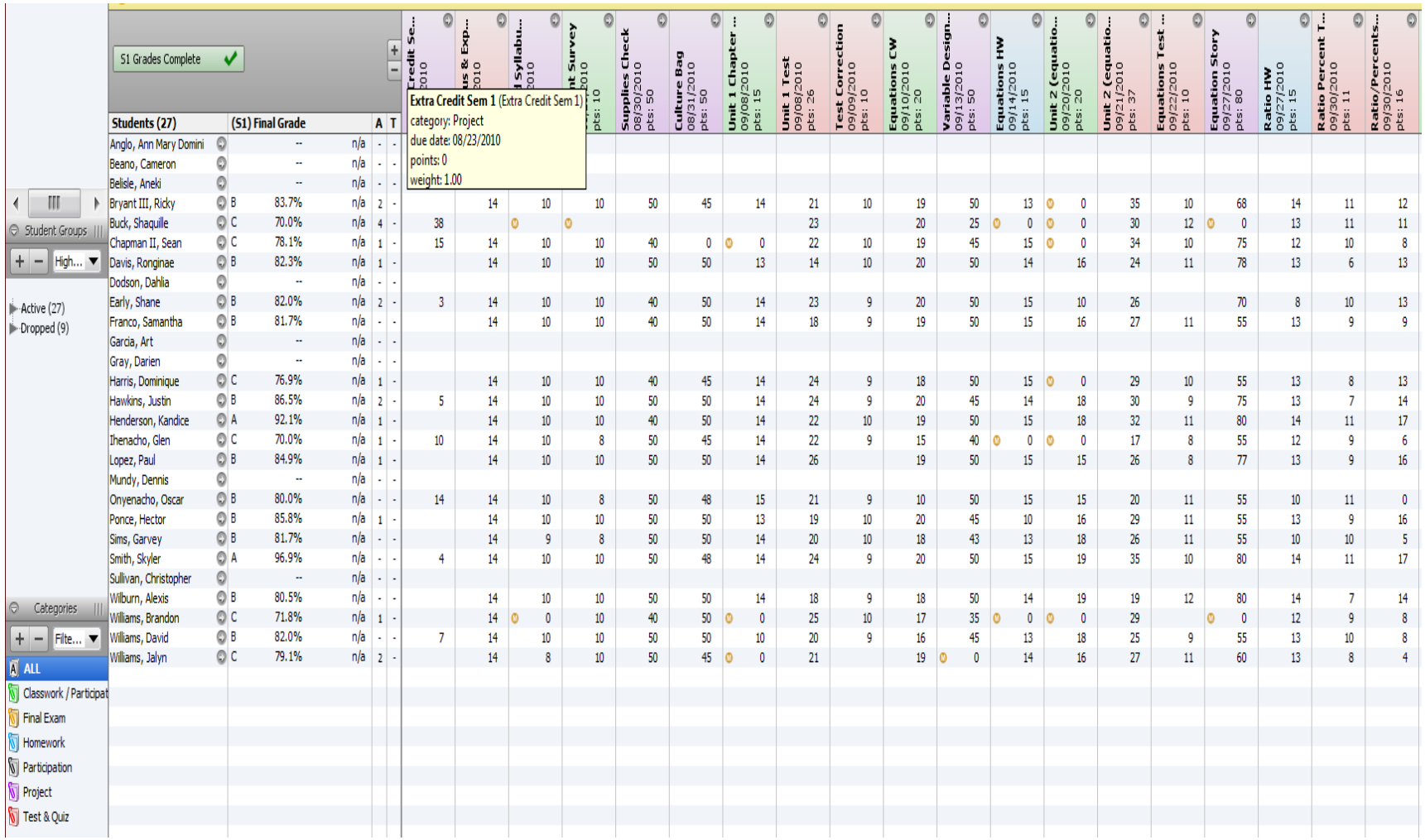
Opinions of New Grading Policy	**Level of Agreement**				
	Strongly Disagree	*Disagree*	*Neutral*	*Agree*	*Strongly Agree*
The current school-wide grading policies benefit our students.					
I agree with the current school-wide grading policies.					
The grading policies of my classes follow the school-wide grading policies.					
The grading policies of all of my classes are the same.					
The current school-wide grading policies are fair.					
All of my grading policies are fair.					
The current school-wide grading policies help the school achieve its mission.					
The current school-wide grading policies are easy to implement.					
The current school-wide grading policies improve the way I teach.					

Please rank order from 1-3, the top 3 most important criteria you use when determining the final grade for your students.

a) Effort _____
b) Work Completion _____
c) Attendance _____
d) Participation _____
e) Extra Credit _____
f) Behavior _____
g) Tests _____
h) Final Exams _____
i) Quizzes _____
j) Class work _____
i) Homework _____
k) Projects _____
l) Reports _____
m) Work Habits _____
n) Journals _____
o) In Class Observation _____
p) Portfolios _____
q) Other: _____ _____

Appendix B

Sample Score Sheet Screenshot



References

Andrade, H. G. (2000). Using rubrics to promote thinking and learning. *Educational Leadership, 57*(5), 13-18.

Anyon, J. (1997). *Ghetto schooling: A political economy of urban educational reform*. New York: Teachers College Press.

Anyon, J. (2006). Social class and the hidden curriculum of work. In G. Handel and G. Handel (Eds.), *Childhood socialization* (2nd ed.) (pp. 369-394). New Brunswick, NJ US: Aldine Transaction.

Apple, M. W. (1980). The other side of the hidden curriculum: Correspondence theories and the labor process. *Journal of Education*, *162*(1), 47.

Baird, L. L. (1976). *Using self-reports to predict student performance* (Research Monograph No. 7). New York: The College Board.

Banks, J. A., & Banks, C. (2004). *Multicultural education: Issues and perspectives. 5th Edition, Update*. Jossey-Bass, An Imprint of Wiley.

Barnes, L. L. B., Bull, K. S., Perry, K., & Campbell, N. J. (1998). *Discipline-related differences in teaching and grading philosophies among undergraduate teaching faculty*. Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA.

Berry, C. (2008). Toward an understanding of evidence of differential validity of cognitive ability tests for racial/ethnic subgroups. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 68(8-B), 5620.

Bloom, B. (1956). *Taxonomy of educational objectives: Book 1 cognitive domain.* White Plains, NY: Longman.

Bloom, B. S., Madaus, G. F., & Hastings, J. T. (1981). *Evaluation to improve learning*. New York: McGraw-Hill.

Bowles, S., & Gintis, H. (1976). *Schooling in capitalist America: Economic reform and the contradictions of economic life.* New York: Basic Books.

Brill, S. (2010, May 23). The Teachers' unions' last stand. *New York Times*, Retrieved from http://www.nytimes.com/2010/05/23/magazine/23Race-t.html?pagewanted=3&_r=1

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

Brookhart, S. M. (2009). *Grading*. Upper Saddle River, NJ: Pearson Education.

Camara, W. (1998, May). *High school grading policies* (Research note RN:04). New York: The College Board. Retrieved from http://research.collegeboard.org/sites/default/files/publications/2012/7/researchnote-1998-4-high-school-grading-policies.pdf

Camara, W., Kimmel, E. W., Scheuneman, J.,& Sawtell, E. A. (2003). *Why are grades inflated?* (Research Note RN:2003-4). New York: The College Board. Retrieved from: http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2003-4-whose-grades-are-inflated.pdf

Clune, W. H. (2001). Towards a theory of standards-based reform: The case of nine NSF state-wide systematic initiatives. In S.H. Fuhrman (Ed.), *From the capitol to the classroom: Standards-based reform in the states* (pp. 13-38). Chicago, IL: The University of Chicago Press.

Codding, J. B., & Tucker, M. S. (1998). *Standards for our schools: How to set them, measure them, and reach them*. San Francisco, CA: Jossey-Bass Publishers.

Congress of the U.S., W. r. (1991). *National Council on Education Standards and Testing Act. Report by Mr. Ford of Michigan, Committee on Education and Labor, House of Representatives, 102nd Congress, 1st Session, June 10, 1991. Report to accompany H.R. 2435, the National Council on Education Standards and Testing Act*.

Cummings, J. (1996). *Negotiating identities: education for empowerment in a diverse society.* California Association for Bilingual Education: Ontario, Canada.

Day-Vines, N.L., & Terriquez, V. (2008). A strength-based approach to promoting discipline among African American and Latino students. *Professional School Counseling, 12*, 170-175.

Dewey, J. (1956). *The child and the curriculum, and the school and society.* Chicago, IL: University of Chicago Press.

Dweck, C. S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry*, *6*(4), 267-285.

Earl, L. M. (2003). *Assessment as learning: Using classroom assessment to maximize student learning*. Thousand Oaks, CA: Corwin Press Inc.

Evans, R. (1996). *The human side of school change*. San Francisco, CA: Jossey-Bass Publishers.

Feldmesser, R. A. (1971). *The positive functions of grades*. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Frisbie, D. A., Waltman, K. K. (1992). Developing a personal grading plan. *Educational Measurement: Issues and Practices. 11*(3), 35-42.

Garcia, S. B., & Guerra, P. L. (2004). Deconstructing deficit thinking: Working with educators to create more equitable learning environments. *Education and Urban Society*, *36*(2), 150-168.

Gay, L. R., Mills, E., & Airasian, P. W. (2009). *Educational research: competencies for analysis and applications.* Upper Saddle River, NJ: Pearson Education, Inc.

Giroux, H. A. (1983). *Theory and resistance in education: a pedagogy for the opposition.* South Hadley, MA: Bergin & Garvey.

Godfrey, K. E. (2011). *Investigating grade inflation and non-equivalence.* (Research note RN: 2011-2). New York: The College Board. Retrieved from: http://research.collegeboard.org /sites/default/files/publications/2012/7/researchreport-2011-2-investigating-grade-inflation-non-equivalence.pdf

Guskey, T., & Bailey, J. (2001). *Developing grading and reporting systems for student learning*. Thousand Oaks, CA: Corwin Press, Inc.

Haladyna, T. (1999). *A complete guide to student grading.* Needham Heights, MA: Allyn & Bacon.

Hamilton, L. S., Stecher, B. M., Yuan, K. (2008). *Standards-Based Reform in the United States: History, Research, and Future Directions* (Document No. RP-1384).Washington, D.C.: Center On Education Policy. Retrieved from http://www.rand.org/pubs/reprints/RP1384.html

Hatt-Echeverria, B., & Jo, J. (2005). Understanding the new racism through an urban charter school. *Educational Foundations*, *19*(1-2), 51-65.

Hodgkinson, H. (1999). State differences: The key to demographics. *School Business Affairs*, *65*(5), 32-36.

Jenson, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review,* Vol. 39*,* 1-123.

Kirschenbaum, H., Napier, R., & Simon, S. (1971). *Wad-Ja-Get? The grading game in American education*. New York: Hart Publishing.

Kobrin, J. L., Camara, W. J., & Milewski, G. B. (2002). *Students with discrepant high school GPA and SAT I scores* (Research Note No. RN-15). New York: The College Board. Retrieved from: http://research.collegeboard.org/publications/content/2012/05/students-discrepant-high-school-gpa-and-sat-scores

Kobrin, J. L., Milewski, G. B., Everson, H., & Zhou, Y. (2003, April). *An investigation of school-level factors for students with discrepant high school GPA and SAT scores.* Paper presented at the Annual Meeting of the National Council on Measurement in Education, Chicago, IL.

Kobrin, J. L., Patterson, B.F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2011). *Validity of the SAT for predicting first-year college grade point average.* (Research Note RN 2008-5). New York: The College Board. Retrieved from: http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2008-5-validity-sat-predicting-first-year-college-grade-point-average.pdf

Kobrin, J. L., & Patterson, B. F. (2010, July). *Exploring the variability in the validity of SAT scores and high school GPA for predicting first-year college grades at different colleges and universities.* Paper presented at the meeting of the American Educational Research Association, New Orleans, LA. Retrieved from: http://research.collegeboard.org/sites/default/files/publications/2012/7/presentation-2010-6-variability-validity-sat-hs-gpa-fy-gpa.pdf

Kohl, H. (1994). *"I won't learn from you": And other thoughts on creative maladjustment.* New York: New Press.

Koretz, D., & Berends, M. (2001). *Changes in high school grading standards in mathematics, 1982-1992* (Research Report MR-1445-CB). Santa Monica, CA: RAND Corporation. Retrieved from http://www.rand.org/publications/MR/MR1445

Kovas, M. A. (1993). Make your grading motivating: Keys to performance based evaluation. *Quill and Scroll, 68*(1), 10-11.

Kulik, C. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990a). Effectiveness of mastery learning programs: A meta-analysis. *Review of Educational Research, 60*(2), 265-299.

Kulik, C. C., Kulik, J. A., & Bangert-Drowns, R. L. (1990b). Is there better evidence on mastery learning? A response to Slavin. *Review of Educational Research, 60*(2), 303-307.

Kuncel, N. R., Credé, M., & Thomas, L. L. (2005). The validity of self reported grade point average, class ranks, and test scores: A meta-analysis and review of the literature. *Review of Educational Research*, *75,* 63-82.

Ladson-Billings, G. (2006). "Yes, but how we do it?": Practicing culturally relevant pedagogy. In J. Landsman & C. Lewis (Eds.), *White teachers/diverse classrooms: A guide to building inclusive schools, promoting high expectations, and eliminating racism* (pp. 1-26). Sterling, VA: Stylus.

Langhout, R. D., & Mitchell, C. A. (2008). Engaging contexts: Drawing the link between student and teacher experiences of the hidden curriculum. *Journal of Community & Applied Social Psychology*, *18*(6), 593-614.

Lawrence, I., Rigol, G. W., Essen, T. V., & Jackson, C.A. (2002). *A Historical Perspective on the SAT 1926-2001.* (Research Note No. 2002-7). New York: The College Board.

Linn, R. L. (1983). Testing and instruction: Links and distinctions. *Journal of Educational Measurement, 20*(2), 179-189.

Litton, E. F., Martin, S. R., Higareda, I., & Mendoza, J. A. (2010). The promise of Catholic schools for educating the future of Los Angeles. *Catholic education: A Journal of Inquiry & Practice*, *13*(3), 350-367.

Loyola Marymount University School of Education. (2009, March 30). *Conceptual framework*. Retrieved from: http://soe.lmu.edu/Assets/Colleges+Schools/SOE+2009/framework.pdf

Loeb, H., Knapp, M. S., & Elfers, A. M. (2008). Teachers' response to standards-based reform: Probing reform assumptions in Washington state. *Education Policy Analysis Archives*, *16*(8), 1-29.

Marzano, R. J., (2000). *Transforming classroom grading*. Alexandria, VA: Association for Supervision and Curriculum Development.

Mattern, K. D., & Shaw, E. J. (2009). *Examining the accuracy of self-reported high school grade point average.* (Research Report No. 2009-5). New York: The College Board. Retrieved from: http://research.collegeboard.org/sites/default/files/publications/2012/7/researchreport-2009-5-examining-accuracy-self-reported-high-school-grade-point-average.pdf

Mattern, K. D., Shaw, E. J., & Kobrin, J. L. (2010, April). *A case for not going SAT-optional: Students with discrepant SAT and HSGPA performance.* Paper presented at the meeting of the American Educational Research Association, Denver, CO. Retrieved from: http://research.collegeboard.org/sites/default/files/publications/2012/7/presentation-2010-3-case-for-not-going-sat-optional.pdf

Maxey, E. J., & Ormsby, V. J. (1971). *The accuracy of self-report information collected on the ACT test battery: High school grades and items of nonacademic achievement* (ACT Research Report No. 45). Iowa City, IA: The American College Testing Program.

McLaren, P. (1999). *Schooling as a ritual performance: Towards a political economy of educational symbols and gestures*. Boulder, CO: Rowman & Littlefield Publishers, Inc.

Mickelson, R. (1980). Social stratification processes in secondary schools: A comparison of Beverly Hills High School and Morningside High School. *Journal of Education*, *62*(4), 83-112.

Morris, E. W. (2005). "Tuck in that shirt!" Race, class, gender, and discipline in an urban school. *Sociological Perspectives*, *48*(1), 25-48.

National Governors Association Center for Best Practices, Council of Chief State School Officers (2010). *Common core state standards initiative standards-setting criteria*. Washington D.C. : Author. Retrieved from common core standards website: http://www.corestandards.org/assets/Criteria.pdf

Nieto, S. (2005). Public education in the twentieth century and beyond: High hopes, broken promises and an uncertain future. *Harvard Educational Review;* Vol 75, 1.

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

Ornstein, A. C. (1994). Grading practices and policies: An overview and some suggestions. *NASSP Bulletin, 78*(559), 55-64.

Polloway, E. A., & Epstein, M. H. (1994). Classroom grading: A national survey of policies. *Remedial & Special Education*, *15*(3), 162.

Ramist, L., Lewis, C., & Jenkins, L. M. (1997). *Students with discrepant high school GPA and SAT scores.* (Research Report RS-01). New York: The College Board. Retrieved from: http://research.collegeboard.org/sites/default/files/publications/2012/7/researchinreview-1997-1-students-discrepant-hs-gpa-sat.pdf

Ravitch, D. (1995). *National standards in American education*. The Brookings Institution: Washington D.C.

Berry, C. (2008). Toward an understanding of evidence of differential validity of cognitive ability tests for racial/ethnic subgroups. *Dissertation Abstracts International: Section B: The Sciences and Engineering*, 68(8-B), 5620.

Rich, R. H. (2001). *Hidden factors in teachers' secondary grading practices (*Doctoral Dissertation). Retrieved from ProQuest Information and Learning database. (UMI 3036925).

Roorda, N. (2008). *The impact of professional development on assessment and grading practices for secondary teachers* (Doctoral Dissertation). Retrieved from Dissertations & Theses: Full Text. (Publication No. AAT 3342447).

Sawyer, R., Laing, J., & Houston, M. (1988). *Accuracy of self-reported high school courses and grades of college-bound students* (ACT Research Report 88-1). Iowa City, IA: The American College Testing Program.

Schiel, J., & Noble, J. (1991). *Accuracy of self-reported course work and grade information of high school sophomores* (ACT Research Report 91-6). Iowa City, IA: The American College Testing Program.

Skibia, R. J., Knesting, K., & Bush, L. D. (2002). Culturally competent assessment: More than nonbiased tests. *Journal of Child and Family Studies*. *11*(1).

Skrla, L., & Scheurich, J. (2001). Displacing deficit thinking in school district leadership. *Education and Urban Society, 33*, 239-259.

Spring, J. (1972). *The rise and fall of the corporate state*. Boston: Beacon Press.

Starch, D., & Elliott, E. C. (1912). Reliability of the grading of high-school work in English. *School Review*, *20*, 442-457

Thadani, V. V., Breland, W. W., & Dewar, J. J. (2010). College instructors' implicit theories about teaching skills and their relationship to professional development choices. *Journal on Excellence in College Teaching*, *21*(2), 113-131.

The College Board. (2001). *National report 2001 college-bound seniors. A profile for SAT program test takers.* Retrieved from http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2001/NATL.pdf

The College Board. (2002). *National report 2002 college-bound seniors. A profile for SAT program test takers.* Retrieved from http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2002/2002_TOTAL_GROUP_REPORT.pdf

The College Board. (2003). *National report 2003 college-bound seniors. A profile for SAT program test takers.* Retrieved from http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2003/2003_TOTALGRP_PRD.pdf

The College Board. (2004). *National report 2004 college-bound seniors. A profile for SAT program test takers.* Retrieved from http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2004/2004_CBSNR_total_group.pdf

The College Board. (2005). *2005 college-bound seniors total group profile report.* Retrieved from http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2005/2005-college-bound-seniors.pdf

The College Board. (2006). *2006 college-bound seniors total group profile report* .Retrieved from http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2006/national-report.pdf

The College Board. (2007). *2007 college-bound seniors total group profile report.* Retrieved from http://www.collegeboard.com/prod_downloads/about/news_info/cbsenior/yr2007/national-report.pdf

The College Board. (2008). *2008 college-bound seniors total group profile report.* Retrieved from http://professionals.collegeboard.com/profdownload/Total_Group_Report.pdf

The College Board. (2009). *2009 college-bound seniors total group profile report* .Retrieved from http://professionals.collegeboard.com/profdownload/cbs-2009-national-TOTAL-GROUP.pdf

The College Board. (2010). *2010 college-bound seniors total group profile report.* Retrieved from http://professionals.collegeboard.com/profdownload/2010-total-group-profile-report-cbs.pdf

The College Board. (2011a). *2010 College bound-seniors state profile report California.* Retrieved from http://professionals.collegeboard.com/profdownload/CA_10_03_03_01.pdf

The College Board. (2011b). *What is the SAT*. Retrieved from http://professionals.collegeboard.com/testing/sat-reasoning

The College Board. (2011c). *SAT Validity Studies*. Retrieved from http://research.collegeboard.org/programs/sat/data/validity-studies

Thorndike, R. (2005). *Measurement and evaluation in psychology and education* (7th ed.) Upper Saddle River, NJ: Pearson Education, Inc.

Trumbull, E., & Farr, B. P. (2000). *Grading and reporting student progress in the age of standards*. Norwood, MA: Christopher-Gordon Publishers, Inc.

Turnbull, B. (2002). Teacher participation and buy-in: Implications for school reform initiatives. *Learning Environments Research, 5*(3), 235-252.

U.S. Department of Education. (1993). *Improving America's schools act of 1993: The reauthorization of the Elementary and Secondary Education Act and amendments to other acts.* Washington, D.C.

U.S. Department of Education, National Commission for Excellence in Education. (1983). *A nation at risk: The imperative for educational reform.* Washington, D.C.: Gardner, D.P.

U.S. Department of Education. (2009). *Race to the top executive summary.* Retrieved from: http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

U.S. Department of Education, Office of Educational Research and Improvement. (1994). *What do student grades mean? Differences across schools*. Washington, D.C. Retrieved from http://www.ed.gov/pubs/OR/ResearchRpts/grades.html.

Valencia, R. R. (1997). *The evolution of deficit thinking: Educational thought and practice*. London: Falmer.

Vinovskis, M. (1999). *The road to Charlottesville: The 1989 education summit. A publication of the National Education Goals Panel.* Retrieved From:http://govinfo.library.unt.edu/negp/reports/negp30.pdf

Wiggins, G. (1994). Toward better report cards. *Educational Leadership, 52*(2), 28-35.

Willingham, W.W., Pollack, J. M., & Lewis, C. (2002). Grades and test scores: Accounting for observed differences. *Journal of Educational Measurement*, *39,* 1-37.