

ACCOUNTABILITY IN TEXAS: PREDICTING DISTRICT RATINGS BASED  
UPON SELECT INDICATORS

by

SARAH FITZHUGH

Presented to the Faculty of the Graduate School of  
The University of Texas at Arlington in Partial Fulfillment  
of the Requirements  
for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

December 2011

Copyright © by Sarah Fitzhugh 2011

All Rights Reserved

## ACKNOWLEDGEMENTS

I would like to begin by thanking the fine folks at Panera, whose delicious macaroni and cheese fueled me to continue writing when I thought there was no more to write. To Dr. Hardy for putting up with my countless emails and questions, yet always providing a guide to the path necessary for completion, thank you. I would also like to thank Dr. Munoz and Dr. Casey for their time and input along the way. To my cohort and writing partner Ramona, thank you for all of your encouragement and support- for keeping me moving along when challenges arose. It is probably best to say thank you and I am sorry to all of those who endured with me through this process.

I would not have been able to complete this great undertaking without the support of my family who allowed me to spend my weekends eating macaroni and cheese and writing while they babysat and changed diapers. Thank you Brian and Dylan for supporting me in pursuit of my dream. To the many that inspired me and pushed me to continue my education, I say thank you. Not all who began this journey with me are still here in being, but I know they are here in spirit and smiling; thank you for always believing in me.

November 14, 2011

## ABSTRACT

### ACCOUNTABILITY IN TEXAS: PREDICTING DISTRICT RATINGS BASED UPON SELECT INDICATORS

Sarah Fitzhugh, PhD

The University of Texas at Arlington, 2011

Supervising Professor: Dr. James C. Hardy

The major purpose for the study was to determine if a relationship existed between district ratings, as assigned by the Texas Education Agency, and population groups, funding, and class size. The variables utilized for this study included the following number of student ethnicity/ race (African American, Hispanic, and White) utilized in a district rating as well as, the Economically Disadvantaged group indicator. The groups were coded to indicate each district that had a particular student group applied towards the district's accountability rating. It was then determined how many race/ ethnicity indicators counted in a rating for each district. In addition, teacher-to-student ratio and funds spent per student were analyzed. The study also explored the relationship between ratings and college readiness of students.

It is possible for a district in Texas to attain one of four ratings. Three separate logistic regressions were calculated using several independent variables as predictors. Two research questions examined the relationship between college readiness and ratings. A point biserial correlation was calculated to determine the relationship between ratings and college readiness as determined by the Reading test scores. A point biserial correlation was run to determine the relationship between ratings and college readiness as determined through Mathematics test scores.

No included variables were found to effectively discriminate between *Academically Acceptable* and *Academically Unacceptable* districts. The number of race/ethnic groups utilized in the rating as well as the economically disadvantaged indicator both contributed to a district being rated *Recognized* or *Academically Acceptable*. The economically disadvantaged indicator was found to predict a district being rated *Exemplary* or *Recognized*. Findings indicated that ratings and college readiness were positively correlated for both the Reading and Mathematics assessments.

## TABLE OF CONTENTS

ACKNOWLEDGEMENTS.....	iii
ABSTRACT .....	iv
LIST OF TABLES.....	ix
Chapter	Page
1. INTRODUCTION .....	1
Statement of the Problem.....	6
Accountability and Predictor Variables.....	6
Accountability and College Readiness .....	8
Purpose of the Study .....	9
Research Questions.....	10
Significance of the Study.....	11
Definitions of Terms.....	16
Limitations and Delimitations .....	19
Basic Assumptions.....	20
Organization of the Dissertation Chapters.....	20
2. LITERATURE REVIEW .....	21
History of Accountability .....	21
Accountability Systems .....	29
State of California.....	30

State of New York .....	33
State of Texas .....	35
Texas Assessment of Knowledge and Skills and Current Accountability System .....	36
Changing Demographics .....	42
Demographic Factors in Accountability .....	43
Class Size and Accountability .....	46
Educational Financing and Accountability.....	47
College and Career Readiness and Accountability.....	49
Summary of Literature Review .....	50
3. METHOD OF PROCEDURE .....	52
Research Questions.....	52
Research Design .....	53
Procedure .....	54
Collection of the Data .....	54
Treatment of the Data.....	55
Research Questions 1, 2, and 3 .....	56
Research Questions 4 and 5 .....	58
Summary of Methods .....	59
4. PRESENTATION AND ANALYSIS OF DATA.....	60
Data Characteristics .....	61

Descriptive Statistics by Rating Category .....	64
Analysis of Research Questions .....	67
Research Question 1 .....	67
Research Question 2 .....	68
Research Question 3 .....	72
Research Question 4 .....	75
Research Question 5 .....	76
5. SUMMARY, FINDINGS, CONCLUSIONS, IMPLICATIONS FOR PRACTICE AND RECOMMENDATIONS FOR FURTHER RESEARCH.....	77
Summary.....	77
Discussions .....	78
Findings .....	81
Conclusions.....	83
Implications for Practice.....	85
Recommendations for Further Study .....	86
APPENDIX	
A. TABLE A1 AND A2 .....	88
B. PERMISSION FROM TEA TO UTILIZE TABLES IN DISSERTATION.....	92
C. GUIDE TO CODING .....	94
D. IRB APPROVAL.....	97
REFERENCES .....	99
BIOGRAPHICAL INFORMATION.....	109

## LIST OF TABLES

Table	Page
1. Texas Assessment of Knowledge and Skills (TAKS) Test by Grade and Subject.....	37
2. Descriptive Statistics of Student Population Groups for All Districts Utilized in Study .....	61
3. Descriptive Statistics of Non Population Groups for All Districts Utilized in Study.....	62
4. Select Descriptive Statistics of Population Groups by Rating Category for All Districts Utilized in Study .....	63
5. Crosstab Calculation of Number of Race/ Ethnic Groups by Rating Category .....	64
6. Select Descriptive Statistics of Non Population Group Variables for Academically Unacceptable Districts Utilized in Study .....	64
7. Select Descriptive Statistics of Non Population Group Variables for Academically Acceptable Districts Utilized in Study .....	65
8. Select Descriptive Statistics of Non Population Group Variables for Recognized Districts Utilized in Study .....	66
9. Select Descriptive Statistics of Non Population Group Variables for Exemplary Districts Utilized in Study.....	66
10. Logistic Regression Analysis of Rating categories Academically Acceptable and Academically Unacceptable with All Variables.....	67
11. Logistic Regression Analysis of Rating categories Recognized and Academically Acceptable with All Variables.....	68
12. Logistic Regression Analysis of Rating categories Recognized and Academically Acceptable with Significant Variables .....	69

13. Predicted Probability of Being Rated Recognized versus Academically Acceptable for 10 Districts .....	70
14. Cross tabulation of Race/Ethnicity Indicators Utilized in Calculating Rating and Rating Assigned to District: Recognized to Academically Acceptable .....	71
15. Accuracy of Predicted Ratings According to Predicted Logit Equation: Recognized to Academically Acceptable .....	71
16. Logistic Regression Analysis of Rating Categories Exemplary and Recognized with All Variables.....	72
17. Predicted Probability of Being Rated Exemplary versus Recognized for 10 Districts .....	73
18. Logistic Regression Analysis of Rating Categories Recognized and Exemplary with Significant Variables .....	74
19. Cross tabulation of Race/Ethnicity Indicators Utilized in Calculating Rating and Rating Assigned to District: Exemplary and Recognized with Significant Variables .....	74
20. Accuracy of Predicted Ratings According to Predicted Logit Equation: Exemplary to Recognized.....	75
21. Correlation between Rating and College Career Readiness Mathematics Rates.....	76
22. Correlation between Rating and College Career Readiness Reading Rates .....	76
A1. Requirements for Each Rating Category of Districts in Texas.....	89
A2. Texas Overview of 2008 System Components.....	90

## CHAPTER 1

### INTRODUCTION

Striving to provide the most appropriate and necessary education to *all* students nationwide has long been a goal of the school system in the United States. Federal legislation from The Elementary and Secondary Education Act of 1965, created to help students living in poverty, to The No Child Left Behind (NCLB) legislation of 2001, created to hold schools accountable for the education of all students, has pushed for improving the education of all American students (Sass, 2010). Under NCLB, schools are evaluated based on student proficiency with the assumption that student performance on the exam is reflective of the caliber of education provided by the school (Hopmann, 2008). No Child Left Behind outlined specific criteria for schools regarding performance of the entire student population as well as evaluating specific target populations to ensure performance gaps do not exist among population groups such as White, Hispanic, African American, Economically Disadvantaged, Special Education, and Limited English Proficient students (NCLB, 2002). Schools that do not meet performance criteria face specified sanctions. Ultimately, such an accountability system is intended to expose low-performing schools and necessitate increasing and improving student performance (Kane & Staiger, 2003).

Historically, assessment scores of low-socioeconomic students and students of color are generally lower when compared with their middle class, Anglo peers (Nelson,

McGhee, Meno, & Slater, 2007). In addition to setting evaluation standards, NCLB outlines a protocol for schools that fail to meet standards. Based on the number of consecutive years a school fails to meet the NCLB standards, the federal government provides for a range of actions. Examples of actions allowed under NCLB include the option for students to transfer to another campus that meet standards and/or the district replacing the majority of the staff assigned to the underperforming campus (McDermott, 2003).

Assessments utilized under NCLB are created by each state with each state setting its own proficiency standard. States can set a low threshold for proficiency and avoid federal consequences (Jennings & Corcoran, 2009). The reasoning behind each state creating its own exam can be explained in part by local control. Maintaining local control, meaning individual states have the authority to determine what is appropriate for their schools, is a priority for educational organization within the United States. Ultimately, states continue to have the final say when setting performance standards, selecting curriculum, and selecting exams (Murnane, 2007).

The federal government not only has strived for change in the education system, but individual states have created systems to improve education for all students. The State of Texas is one such state. With the passing of Senate Bill 7 in 1993, Texas began a system of academic accountability for school districts and individual campuses (Texas Education Agency, 2010a). A number of states generated similar accountability systems based on the reported score increases in Texas (McNeil, 2000). The Texas system is also considered to be the model of the accountability system utilized in the

NCLB Act (Giambo, 2010). Darling-Hammond (2007) suggests that the reason behind creating such a school system for low-rated schools is to help the low-rated schools alter methodologies in order to better serve students; low ratings cause embarrassment. Parents, educators and students are impacted negatively by the stigma created with the label low performing. The school or district is seen as unsuccessful.

The Texas accountability system assigns one of four ratings to campuses and districts based upon dropout rates and the student performance on the state assessment, Texas Assessment of Knowledge and Skills (TAKS) instrument. The four ratings are *Academically Unacceptable*, *Academically Acceptable*, *Recognized* and *Exemplary*. Each rating is achieved based on performance of specified criteria on the Reading/English Language Arts assessment, Writing assessment, Mathematics assessment, Science assessment, and Social Studies<sup>1</sup> assessment in combination with having a dropout rate less than the designated percentage specified by the state. All measures are evaluated for specified indicators. Indicators exist for the total population of students and specified student population groups: White, African American, Hispanic, and Economically Disadvantaged students. The maximum number of indicators a district and/or school can be evaluated on is 35. For example, a campus could meet the criteria to be *Academically Acceptable* in all areas except White students' Writing scores. In that case, the campus rating would drop to *Academically Unacceptable* based on the one group student performance (Texas Education Agency

---

<sup>1</sup> Tested subjects will be capitalized throughout the dissertation for ease of reading.

Department of Assessment, Accountability, and Data Quality Division of Performance Reporting, 2008).

Accountability systems were created in an effort to close achievement gaps between various race and ethnicity groups (Kane & Staiger, 2003). Due to this accountability movement, districts are rated based on standardized test scores, which ultimately only reflect district demographics and the state created assessment program (Townsend, 2002). Each state must create an assessment program and decide how to set proficiency standards with the least unintended consequences. The higher the score set for meeting standards on an exam, the greater the likelihood that schools will face larger numbers of students not meeting the standard. In addition, setting higher standards increases the possibility that schools with large populations of traditionally underperforming student groups, that is, low income, Hispanic, and African American (Nelson et al., 2007) will face sanctions under NCLB due to the isolation of these groups as a performance indicator. If low performance standards are set in order to allow more campuses to appear to have students meeting standards, states run the risk of under teaching students and then compromising the integrity of the education students receive due to the low expectations (Kane & Staiger, 2003),

The stigma created through accountability systems that assign ratings can have long-term unintended consequences for schools and districts. Being labeled “low performing” stigmatizes a district and its campuses by implying that low-performing schools are not improving student performance and/or adequately teaching students (Nelson et al., 2007). The implication is an inappropriate assumption based on a

performance-based assessment that accounts for meeting standards rather than student growth. Students at low-performing schools can make great academic gains and not meet standards. For instance, a student might begin third grade not knowing how to read, but leave third grade reading at a first grade level. The student may have made great strides, yet will most likely not meet standards on the third-grade Reading assessment. The school may have made great strides in teaching the student, but is still negatively impacted by the student score because ratings do not take growth into account. In addition, the students, teachers, administrators, etc, associated with the school “wear” the rating. Being labeled low performing affects the morale of the campus and the community (McNeil, 2000). Furthermore, the public uses ratings for comparative discussions (McNeil, 2000), and to discuss the success of “their” school.

In addition to monitoring student learning, the State of Texas has also set forth standards to insure that students are ready for college and careers based on results of standardized testing. Texas developed a college career readiness indicator for the 11th grade English language arts and Mathematics tests in order to assist districts with monitoring their student preparation. Adopted in 2008, Texas was the first state to create such standards that were intended to help gauge a student’s preparedness after high school graduation. Standards were created by a committee of K-12 public educators, higher education professionals, and community members (Texas Education Agency, 2010b). Currently, the college and career readiness indicators are acknowledged publicly in the Texas accountability system via Gold Performance Acknowledgements. In 2008, there were 14 categories in which a district could receive

such acknowledgements. The following two categories applied to the college and career readiness indicator for the purpose of this study:

- Texas Success Initiative—Higher Education Readiness Component: English Language Arts
- Texas Success Initiative—Higher Education Readiness Component: Mathematics

Districts receive acknowledgement in Mathematics by having 55% of students in applicable student indicator groups achieve a score of 2200 or higher on the 11th grade Mathematics Texas Assessment of Knowledge and Skills (TAKS) test. Districts receive the acknowledgement in English Language Arts by having 55% of students in applicable student indicator groups achieve a score of 2200 or higher and a rating of 3 or higher on the essay portion of the 11th grade English Language Arts TAKS test (Texas Education Agency Department of Assessment, Accountability, and Data Quality Division of Performance Reporting, 2008). The percent of students achieving college and career readiness is also reported on the district's Academic Excellence Indicator System (AEIS) report.

### Statement of the Problem

#### Accountability and Predictor Variables

According to Linn (2008) the purpose of accountability systems is to evaluate the caliber of instruction and learning. Therefore, accountability ratings should reflect student learning as measured by performance on assessments. If, as suggested by Kane

and Staiger (2003), that ratings merely reflect district demographics, the accountability system that the United States is vested in does not fulfill its purpose. Accountability systems were created to determine school performance with regard to adequately educating all students. Should ratings be predicted based on criteria other than student performance, it will indicate a flaw in accountability systems necessitating a restructuring of said systems. Sergiovanni (2000) suggests that the accountability system utilized by Texas is a bureaucratic system that has gone awry. Schlechty (2011) furthered the discussion by stating that not only have accountability systems created bureaucratic systems but have also stripped local communities of their identity and responsibility to the local community by imposing governmental control and keeping schools accountable to the government instead of the local community. While the intent of the accountability system is to monitor student performance, the actual effects include low morale of stakeholders, teachers focusing on meeting standards versus educational growth, and an instructional calendar that focuses on the test instead of student performance (Sergiovanni, 2000). The accountability system dictates what is happening in classrooms rather than measuring the progress of each student in the classroom.

The rationale for developing accountability systems, as defined in this study, was to measure student progress and measure learning as stated by Linn (2008), Schlechty (2011), and Sergiovanni (2000). Therefore, ratings of districts and schools were based on student performance. Stevens, Estrada, and Parkes (2000) suggested that systems should include growth measurements in their model in addition to a variety of

components. This study sought to determine if district ratings could be predicted based on nonacademic variables rather than actual student performance. In order to be effective, accountability systems need to take into account the diversity of demographics and profiles in various school districts when systems are being implemented (Stevens et al., 2000).

Student performance on state-mandated assessments has been evaluated by many researchers. Studies previously conducted by other researchers have included evaluating financial allocations linked to ratings or campus demographics and student performance (Humiston, 2007; Starrett, 2008, Turley, 2009). According to Humiston, Starrett, and Turley, the results of their studies indicate a need for further research into factors affecting student and/or district performance.

#### Accountability and College Readiness

In addition to accountability system evaluations, the extent to which students are prepared for college and careers has become a focus of public schools in the State of Texas. Texas has begun to evaluate student preparedness for success beyond high school by creating the college and career readiness indicator. The current focus of districts in regard to the accountability system in Texas is the rating campuses and districts receive. The indicators used for career readiness are simply gold stars for districts to achieve on top of the rating, which is the main focus. As part of the new accountability metrics, according to Schmidt (2011), the movement to restructure accountability calls for measuring students who graduate from high school for college and career readiness. Do the indicators correlate with the ratings assigned and truly

indicate a districts success? According to the National Center for Educational Statistics, income and race/ethnicity are associated with students' being college ready. For example, 63% of college students, in the lowest quintile, enrolled in remedial college courses. In addition, 62% of African American and 63% of Hispanic students must enroll in remedial courses (Wiley, Wyatt, & Camara, 2010). The National Assessment Governing Board (NAGB) is currently attempting to determine if it is possible to utilize student performance on the National Assessment of Educational Progress (NAEP) to indicate student college preparedness (Wiley & et al., 2010). According to Geiser (2007), high school grade point average is the strongest predictor of student success in college and the high school grade point average, when used for student admission to college, impacts economically disadvantaged and minority students less adversely (Geiser, 2007). Regression analysis was utilized to determine the predictors (Geiser, 2007). Martin (2010) recommends investigating the benefits of increased expectation and student performance. Conversely, Richey (2009) stated in his conclusion that, "high stakes testing in Texas ... could have influenced the rise in developmental reading and writing enrollment" (p. 148); the overall indication is that studies need to be conducted to evaluate the link between ratings based on state assessment and college preparedness.

#### Purpose of the Study

There were two purposes to this study:

1. To determine if the rating of a district, as issued by the State of Texas, can be predicted based on the district's student/teacher ratio, amount spent per

student, and/or specified student population groups: African American, Hispanic, White, and Economically Disadvantaged.

2. To determine if a relationship exists between district accountability ratings and the percent of students that consider college and career ready based on the 11th grade TAKS test in the areas of English Language Arts and Math. The intent of the study was to determine if higher ratings were correlated with the percent of students attaining this level of achievement.

#### Research Questions

The following five questions guided the study:

1. Can a Texas school district's rating of *Academically Acceptable* or *Academically Unacceptable* be predicted based upon one or more of the following variables: the district's number of race/ethnic groups utilized in calculating the rating, district's economically disadvantaged indicator, the district's money spent per student, and/or the district's teacher/student ratio?
2. Can a Texas school district's rating of *Recognized* or *Academically Acceptable* be predicted based upon one or more of the following variables: the district's number of race/ethnic groups utilized in calculating the rating, district's economically disadvantaged indicator, the district's money spent per student, and/ or the district's teacher/student ratio?
3. Can a Texas school district's rating of *Exemplary* or *Recognized* be predicted based upon one or more of the following variables: the district's number of race/ethnic groups utilized in calculating the rating, district's

economically disadvantaged indicator, the district's money spent per student, and/or the district's teacher/student ratio?

4. Is there a relationship between district accountability ratings and district Mathematics college and career readiness rates?
5. Is there a relationship between district accountability ratings and district English Language Arts College and Career Readiness rates?

#### Significance of the Study

According to Linn (2008), the goal of accountability systems is to distinguish between schools that employ effective instructional strategies versus those that do not. In the State of Texas, one of the purposes of the accountability system is for it to be "informative." Community members should be able to look at a school or district rating and clearly understand the quality of the education provided by the school. However, the four ratings *Academically Unacceptable*, *Academically Acceptable*, *Recognized*, and *Exemplary* were found not to be equitable or to measure effective practices (Nelson et al., 2007).

In the United States, most states utilize an accountability system that penalizes low-performing schools and rewards high-performing schools. Labeling schools as inferior in many cases is incorrect (Popham, 2001). Eleven states evaluated student subgroup performance and included this performance in the overall evaluation of the school (Kane & Staiger, 2002). Minority student scores applied towards accountability in multiple categories often lead to students in these groups to receive additional practice for state assessments which took students away from regular classroom

instruction (Baker, Linn, Herman, & Koretz, 2002; Jones, Jones, & Hargrove, 2003). From historical achievement gaps between White students and Hispanic and African American students, there is concern over performance of minority students. Accountability systems were designed to close these gaps by drawing attention to the performance of each specific group; however, the negative consequence from such accountability systems appears to encourage schools to segregate groups of students (Kane & Staiger, 2002). In addition, the state assessment test scores were not as reliable as one would have thought (Kane & Staiger, 2003). Schools and districts play a game by focusing attention on borderline students, who count in multiple accountability categories, in order to increase proficiency in multiple areas and improve ratings (Koretz, 2008a). Accountability ratings, based on test scores, reflected uncontrollable variables such as income level, race, and ethnicity (Kane & Staiger, 2002). If the conjecture that test scores were based on uncontrollable variables such as race, ethnicity, and income level is true, district ratings should be able to be predicted based upon a district's demographics.

Starrett (2008) conducted a descriptive statistical analysis of district size and diversity in relation to accountability ratings in the State of Texas. The findings indicated that district size and diversity impacted accountability ratings; however, degree of impact on ratings was not discussed (Starrett, 2008).

Turley's (2009) dissertation evaluated adequacy of spending and student achievement. Findings indicated that no significant correlation existed between spending and achievement as measured by TAKS scores. The study concluded by

recommending that further research be conducted regarding finance and student performance in order to promote informed decision making.

Humiston (2007) recommends future investigations be conducted regarding the relationship between accountability ratings and variables that are uncontrollable by school personnel. Variables that would be considered uncontrollable would include socioeconomic status and race. Knowing these factors are uncontrollable at the district level, Rouse and Barrow (2006) stated that, “evidence suggests that parental socioeconomic status has a causal effect on children’s educational outcomes” (p. 104). It would seem that rating schools accurately based solely on student performance would not be possible without accounting for these uncontrollable variables.

A high correlation exists between family income and test scores indicating that ratings could reflect the relationship between income and performance more than actual effectiveness of instruction and performance (Hursh, 2005). According to Jennings and Corcoran (2009), when discussing two schools, one serving advantaged students and the other disadvantaged,

It makes no sense to label some schools as ‘in need of improvement’ simply because their students were lower performing to begin with. Nor does it make sense to give other schools a free pass because their students were proficient when they walked through the door. (p. 637)

Based on these studies, the proposed study sought to determine the degree to which various student group indicators could predict district ratings in the State of Texas’ education system.

According to Giambo (2010), high-stakes testing is becoming more prevalent across the country. With accountability and high-stakes testing being used more prominently, predicting ratings based on indicators other than actual student scores could be useful to educators in several ways.

In Texas, a school district is assigned a rating each year based on specified data regarding student performance on state assessments, dropout rates, and completion rates (Texas Education Agency Department of Assessment, Accountability, and Data Quality Division of Performance Reporting, 2008). When created, the accountability system in the State of Texas was designed to demonstrate a district's academic performance on state-mandated assessments ((Nelson et al., 2007). Student performance on the various assessments was the basis for ratings according to the state published 2008 *Accountability Manual* (Texas Education Agency Department of Assessment, Accountability and Data Quality Division of Performance Reporting, 2008). However, research shows that children of families of high income perform better on assessments when compared to students who come from families of lower income households (Rouse & Barrow, 2006). A district's rating can be higher simply by nature of its demographics. If a district's ratings can be predicted by variables other than student performance on standardized assessments, the rating system, utilized as it is, is not serving its intended purpose of indicating a district's academic performance. Instead, the stigma associated with working at low-performing campuses then only serves to discourage qualified and motivated educators from working at campuses that have large populations of minority and poverty students (Nelson et al., 2007). The nature of the

system's design does not truly allow for such conclusions regarding effectiveness to be appropriately drawn, but the conclusions being drawn are leading to stigmatization of the learning organization (Yeow Meng, 2009).

In a 1994 study, Alspaugh (1994) noted an increase in achievement related to student teacher ratios. However, Chingos (2010) and Wößsmann and West (2002) observed no effect regarding class size reduction and student performance. Chingos noted that a study of the impact of class size reduction in the State of California indicated significant improvement in student performance. According to Chingos, conflicting studies regarding class size and student performance warrants further study. Knowing that such ratios can affect achievement in certain situations, the relationship between student teacher ratios and accountability ratings should be studied in order to determine the extent to which a rating can be predicted by the student teacher ratio.

An accountability system is only as good as the data it is based upon and the data it provides . It is important to begin to inspect school accountability systems to determine if the conclusions drawn linking highly effective schools with high rated schools are true or not true (Foote, 2007). If accountability system ratings can be predicted based on the demographics, money, and class size, then the current system does not give information regarding student performance and effectiveness, but merely tells about the district's composition. The ability to predict a rating would also discontinue the practice of inferring that a school with one rating is more effective than a school with a lesser rating because it is not scientifically justified (Linn, 2008).

Of further interest is the relationship between district accountability ratings and student college readiness. Richey (2009) and Martin (2010) both discussed the need for future research to be conducted in the area of college preparedness in relationship to student testing. Guerra (2009) recommended further studies to be conducted to evaluate student preparation and college readiness in Mathematics standards. Lim (2010) found that college preparedness was based more on family variables than schooling variables. Determining the relationship between college readiness and state ratings in conjunction with the previously discussed ratings in relationship to specified variables provided the educational community insight into the accuracy of the readiness measures.

#### Definitions of Terms

**Accountability.** “A policy of holding schools and teachers accountable for students' academic progress by linking such progress with funding for salaries, maintenance, etc.” (Dictionary.com, 2010).

**Adequacy.** According to Texas Association of School Boards (2008),  
In school finance, adequacy refers to whether the state system of finance provides sufficient resources to each district to operate all required programs and services so that all students meet state and federal education standards.

“Adequate” should mean more than just the minimum. Several methods of measuring adequacy have been devised, but experts disagree on the accuracy and usefulness of the results of various methods. (p. 23)

Chapter 41. According to Texas Association of School Boards (2008),

Referring to Chapter 41 of the Texas Education Code, a Chapter 41 school district that has property wealth in excess of \$364,000 per WADA is subject to recapture provisions in Tier 1 and Tier 2a. A district with wealth per WADA of \$319,500 is subject to recapture in Tier 2c. (p. 23)

General Revenue (GR) Fund. According to Texas Association of School Boards (2008), “the GR Fund is the state’s main funding source. Moneys collected from state taxes and other non-dedicated sources go into this fund, and it is used to provide revenue for most state programs” (p. 24).

Limited English Proficient. A student who enters a school in the United States speaking a language other than English.

Low Performing. The term low performing is used throughout this paper to indicate an entity that does not meet the standard. The term is synonymous with not meeting Adequate Yearly Progress (AYP) and other subpar labels. In the State of Texas accountability system, the term “Academically Unacceptable” is utilized rather than “low performing.”

Language Proficiency Assessment Committee (LPAC). A campus-based committee that determines accommodations and language (Grades 3-5) of assessments for a Limited English Proficient student.

No Child Left Behind (NCLB). Law passed by the federal government in 2002 that was a reauthorization of the Elementary and Secondary Education Act of 1965 to set national accountability standards.

Ratings. Assigned based upon student performance on state examinations.

Recapture. According to Texas Association of School Boards (2008), Recapture is a feature of school finance in which local districts give the state locally collected property tax revenue for redistribution through state aid formulas. The recapture provision in Chapter 41 of the Texas Education Code is a salient feature of the Texas school finance equalization system. It has become more complicated under HB 1 (2006). (p. 26)

Student Groups. A group of students, which is separated from the whole population based upon specific demographic characteristics, whose performance on state assessments is reported for accountability ratings.

Target Revenue. According to Texas Association of School Boards (2008), Target revenue is an amount of funding for a school district based on the revenue per WADA a district had in 2005-2006 (or would have had in 2006-2007 under law prior to HB 1). Target revenue includes Tier 1 and Tier 2a state aid, local M&O tax revenues (net of recapture) \$110 per weighted student, and additional aid for declining ADA, property value declines, and taxpayer protests. Target amounts also include the salary increase from HB 1 and the high school allotment. In 2007, the Legislature adjusted target revenue for changes in NIFA and transportation from the base year, changes in economic development tax credits from the base year, district changes in the optional homestead exemption (if more optional homestead exemptions are granted, the target is reduced) and expiring abatements. Currently, target revenue funding replaces formula-based

funding. School district target revenue has not been adjusted for inflation. (p. 26)

Texas Assessment of Knowledge and Skills (TAKS). The state mandated assessment utilized in Texas. It is based upon the Texas Essential Knowledge and Skills (TEKS).

Texas Education Agency (TEA). The agency that manages education in the State of Texas.

Texas Essential Knowledge and Skills. The State of Texas mandated curriculum standards which are utilized to create the TAKS test.

Weighted Average Daily Attendance (WADA). According to Texas Association of School Boards (2008),

In Texas, students with additional educational needs are weighted for funding purposes to help recognize the additional costs of educating those students.

Weighted programs include special education, vocational, bilingual, gifted and talented, and compensatory education. A weighted student count is used to distribute guaranteed yield funding. (p. 26)

#### Limitations and Delimitations

The current study is limited in the following ways:

1. The population of this study is limited to public school districts in the State of Texas.

2. Ratings assigned by the State of Texas in 2008 do not include certain special education populations that participated in TAKS Modified, TAKS Alternative, and TAKS Accommodated.
3. Ratings assigned by the State of Texas are based on assessments taken by students in Grades 3-12. Data utilized for teacher/student ratios and money spent per pupil accounted for Grades Prekindergarten-12th.

#### Basic Assumptions

All data used for this study was by downloaded from the TEA webpage. It is assumed that all data provided TEA are accurate.

#### Organization of the Dissertation Chapters

The study is organized into five chapters. Chapter I includes the introduction to the study, purpose of the study, and significance of the study. Chapter II provides the review of literature that pertains to accountability and ratings as well as education and testing in the State of Texas. Chapter III discusses the methodology utilized to conduct the study. It includes statistical method descriptions as well as collection and treatment of data information. Chapter IV provides analysis of the data based on the statistical procedures that are discussed in Chapter III. Chapter V summarizes the findings of the study and suggestions for future research.

## CHAPTER 2

### LITERATURE REVIEW

The review of literature explores sources that discussed accountability in the field of education in the United States, beginning with the formation of schools in the United States. The literature review discusses the accountability systems in California, New York, and Texas. In addition, the literature review examines the changing demographics of the United States and Texas, as well as accountability in relationship to race, ethnicity, socioeconomics, class size, and district finance, as well as an evaluation of K-12 accountability and impact on student readiness for higher education.

#### History of Accountability

The system known as “public education” in the United States has grown and transformed into the massive system that involves the participation of approximately 20% of Americans on any given school day (Lehman, as cited in Mondale & Patton, 2001). In the United States, the notion of schools for students to learn, dates back to the 1600s (Applied Research Center, n.d.). Depending on the state and town, school facilities and instruction varied. Between 1830-1840, Horace Mann proposed “common schools” that would serve all boys and girls regardless of wealth a common curriculum. It is estimated that only 50% of children attended school in the early 1900s (Mondale & Patton, 2001). By 1920, 17% of 17 year olds graduated from high school. In order to be efficient, a tracking system began where students took an Intelligence Quotient (I.Q.)

test to determine which track they would be placed on. According to Mondale and Patton (2001), psychologists determined based on the I.Q. test, field tested on Army recruits during World War I, “that the average mental age of American adults was 13.7 years and that ethnicity affected intelligence.” (p. 101). Students scoring higher on the IQ test would be placed in more rigorous coursework while those scoring low would be written off. The I.Q. test was only given in English, so students who spoke English as a second language were classified as slow learners more often. Dissenting opinion over the use of I.Q. testing and tracking caused some cities and states such as Los Angeles and Washington D.C. to ban the use of I.Q. tests by the 1970s. Many state continued to use the I.Q. test and the new Scholastic Aptitude Test (S.A.T.). In 1945, 51% of seventeen year old students graduated high school. Critics began to complain that students did not know how to think critically and problem solve. With the Soviet launch of *Sputnik* in 1957, President Eisenhower signed the National Defense Education Act which allocated \$100 million to education in order to support schools in teaching a more rigorous curriculum focusing on math and science (Mondale & Patton, 2001).

Twenty six years later, the United States Department of Education released *A Nation at Risk* in April 1983. The report alerted Americans to the risk of the failing U.S. educational system, (National Commission on Excellence in Education, 1983). *A Nation at Risk* (1983) states the following:

The people of the United States need to know that individuals in our society who do not possess the levels of skill, literacy and training essential to this new era will be effectively disenfranchised, not simply from the material rewards that

accompany competent performance, but also from the chance to participate fully in our national life. A high level of shared education is essential to a free, democratic society and to the fostering of a common culture, especially in a country that prides itself on pluralism and individual freedom. (§ 8)

*A Nation at Risk* detailed the academic areas in which the public school students in the United States were lagging educationally. In addition, it made several suggestions for remediating the deficits in student learning. Recommendations included strengthened curriculum, defined expectations, amount of time spent in school, and recruitment of qualified personnel (Jorgenson & Hoffman, 2003). *A Nation at Risk* went on to discuss how the American education system was failing to educate American students adequately in comparison to the students in the rest of the world. The report also discussed the repercussions of the failing U.S. education system, thus, becoming a catalyst for educational reform (National Commission on Excellence in Education, 1983). One of the main criticisms of *A Nation at Risk* is that mandated assessments are not offered worldwide. The report compared the performance results of students in the United States to the performance results of students of other countries that utilize different instruments and varying performance standards. Comparing data from varying sources without accounting for differences in the instrument of measure and performance standards is not an equitable comparison (Bracey, 2008).

As the catalyst for change, three reform movements evolved from *A Nation at Risk*: the excellence movement, the restructuring movement, and the standards movement. The excellence movement called to increase standards for teachers and

students by adapting and modifying the conditions of the education environment including the following length of school day, number of days students attend school in a year, frequency of assessment. The restructuring movement called for schools to have more power to make decisions and for less decision making at the level of superintendents and school boards. Restructuring also called for professional growth focused on an abundance of instructional strategies and educators trying new innovative ideas to educate students. The standards movement shifted the focus to student achievement on standardized tests. Rather than allowing teachers to assign grades to determine if a student was adequately performing, standardized assessments strive to provide an external exam that claims to truly measure student performance and prevent inflation of teacher assigned scores (Bishop & Mane, 2004). The movement emphasizes individual and group performance on state-created assessments, which emphasize mandated state-created curriculum and performance standards.

The movement to improve the nation's schools continued through the *Improving America's Schools Act* of 1994 (IASA) combined with the *Goals 2000: Educate America Act*. The two pieces of legislation focused on the academic achievement of all students. *Improving America's Schools Act* required all states to have a set curriculum, performance standard, aligned assessments, and an accountability system that would identify schools where students were not performing as expected on the state assessments. States had local control to make curriculum decisions as well as set performance standards. The understood emphases of IASA were to reform schools and educate all students. By year 2000, the U.S. Department of Education had approved 48

states to develop content standards and 24 states to develop performance standards. According to Jorgenson and Hoffman (2003), all of the states were looking to publishers for a quick way to administer a test and get results that could be examined to determine that students were being taught the designated curriculum.

By 2001, individual states had a variety of accountability systems. Some identified underperforming schools while others did not. Some allocated authority to states to reconstitute schools. Some allowed states to replace staff all based on school performance on standardized assessments (McDermott, 2003). In 2002, the No Child Left Behind Act went into effect nationally (Hunt, 2008). No Child Left Behind (2002) stipulated federal requirements for performance on state-created assessments. Assessments needed to be aligned to the state curriculum. It mandated that schools determine why students were not performing well and how best to remediate the education system. Adequately Yearly Progress (AYP) was measured based on the performance on the state assessment to determine if the educational institution was addressing students' individual needs and demonstrating growth in the percentage of students meeting standards. All states had to submit a plan that detailed their plans for success accountability. Plans were to include details of the accountability system, assessments and standards, and reporting adopted procedures (Jorgenson & Hoffman, 2003).

No Child Left Behind further linked funding to performance. No Child Left Behind created a federal system that allowed the federal government to impose a series of consequences based on the number of consecutive years a school failed to achieve a

satisfactory performance. In the first year of failure to meet AYP, students are allowed to transfer from the assigned campus to a campus that met AYP. In the second consecutive year, students are given the right to request and receive extra educational services at the school district's expense. In the third consecutive year of failing to meet AYP, the district must make staffing changes at the campus level. In the fourth consecutive year, district action must include one of several options that include replacing the majority of the school staff (McDermott, 2003).

A basic assumption of NCLB is that children are left behind because of poor teaching and the best way to measure the quality of education received at a school is the student's performance on the exam (Hopmann, 2008). Creation of accountability systems in education stemmed from the need for public schools to be held responsible for educating all students prior to allowing them to graduate. Prior to NCLB, there were instances of schools allowing students to graduate or promote up to the next grade level when they were not adequately prepared for the next grade. Therefore, it was believed that the solution was to create a system that would hold schools and districts accountable for teaching students appropriately. The reasoning for accountability is logical, but the efficacy in addressing public concern has not been determined (Johnson, Rochkind, & DuPont, 2011).

Limitations to consider when discussing NCLB include the fact that each state selects its own tests, performance standards, and curriculum (Murnane, 2007). According to Cawletti (2006, p.1), "NCLB is now the prescribed treatment for the achievement gaps in U.S. schools, but it has some serious side effects." Side effects

include skewed curriculum, discouraged teachers, and manipulation of numbers. Schools are focusing on Reading and Mathematics in order to raise achievement in those areas and neglecting other subjects. The looming test is driving teachers into utilizing more and more practice materials to ensure good scores and discouraging teachers from using their professional judgment to educate students. In addition, the ability of each state to set individual performance standards on the state assessment is creating an illusion. National Assessment of Education Progress (NAEP) is administered across the United States. State proficiency ratings often conflict with NAEP results because of states setting low passing standards or utilizing easier assessments (Cawelti, 2006).

Currently, NCLB is being evaluated for reauthorization. Discussion centers around modifying the expectation from 100% proficiency in Reading and Mathematics to focus on students graduating and readiness for career and collegiate coursework. Furthermore, evaluating student population groups is being addressed. Rather than focus on students meeting a standard, the focus would shift to individual student growth over time. Such a shift, would remove the emphasis on specified populations to merely achieve the standard and promote growth for all students; even students who are high performers (Schmidt, 2011). In addition, educators are calling to reform accountability systems (Public Education Visioning Institute, 2008).

Public Education Visioning Institute (2008) reports that, “accountability systems of themselves do not produce excellence. Excellence can only come from commitment and meaning” (p. 2). Koretz (2008b) suggests including out-of-school programs as part

of accountability measures since student performance is affected by factors other than those in school. Additionally, performance measures should be extended to other curricular areas besides Reading and Mathematics (Sunderman, 2008). Accountability systems should utilize a variety of assessments to determine student learning goals. Assessments should be formative and vary in form in order to allow students to demonstrate learning in a multitude of ways. Data from assessments should be utilized not only to monitor student growth, but should also be utilized to plan and create new learning goals (Public Education Visioning Institute, 2008). Kornharber (2008) states that,

Any system of educational accountability in the United States should be guided by two essential and equally important aims. First, it should be informative—it should allow the public to know the status and progress of students’ knowledge, skills, and understanding within and across schools. Second, the system should be cognitively constructive- it should advance all students’ learning and enable educators to improve instruction. (p. 43)

No Child Left Behind did not support or foster change in teaching methodology. It did not build capacity in teachers; NCLB focused on compliance. State agencies created systems to evaluate schools and impose sanctions on low-performing schools in order to comply with the law (Sunderman, 2008). According to Graue and Johnson (2011), the current accountability systems and indicators are troublesome. Rothstein, Jacobsen, and Wilder (2008) discussed that the current accountability system, based on standardized tests, is not adequate, but a new system could hold educators accountable for teaching.

The authors believed that the new system would be more expensive and not based on quantitative data but it would be more adequate than the current systems.

### Accountability Systems

No Child Left Behind requires each state to create an accountability system to evaluate student performance in its districts. No Child Left Behind mandates that states monitor African American, Hispanic, White, Economically Disadvantaged, and Limited English Proficient students' performance in Reading and Math. Test scores from students in Grades 3-8 and 10 are utilized in district and campus federal ratings (Texas Education Agency, 2011). When determining how well a school has performed, student diversity and other factors are not explained within the one-word rating assigned (Starrett, 2008). Accountability ratings should have been viewed as descriptive; a starting point for further inquiry into the functioning of districts (Linn, 2008). No Child Left Behind created federal accountability standards that require states to disaggregate schools according to specified student groups. Students are evaluated as an entire population as well as their membership in the following groups: African American, Hispanic, White, Limited English Proficient, Economically Disadvantaged, and Special Education (Texas Education Agency, 2010). The purpose of this form of disaggregation is to prevent the "hiding" of low-performing students who belong to groups that traditionally struggle within the larger group of all students (United States Department of Education, 2006). In addition to federal accountability, most states also have their own system of accountability (Starrett, 2008).

In 2006, every state including the District of Columbia had a testing policy in place. All states tested at minimum English and Writing. All states require students to test up until Grade 10 at a minimum. The states of Alabama, California, Georgia, Michigan, Montana, New York, South Dakota, Texas, and Utah require students in Grade 11 to take exams in the areas of English/Language Arts/Writing, Math, Science, and History/Social Studies (Grade by Grade State Testing Policies, 2006).

Under NCLB, states developed an accountability system to measure student performance. Each state created its own individual program. California, Texas and New York are the largest states in the United States based on population counts. Examining the state of California, New York, and Texas' accountability systems provides an insight into the differences between each state.

#### State of California

All students must pass the California High School Exit Exam (CAHSEE) in order to earn a high school diploma. The accountability program is called Standardized Testing and Reporting (STAR; California Department of Education, n.d. c). Students and groups are evaluated and assigned one of five performance levels: *advanced*, *proficient*, *basic*, *below basic* and *far below basic*. Student scores are reported on a scale of 150 to 600 (California Department of Education, n.d. a).

Data are disaggregated and reported based on gender, economic status, disability, and proficiency in English. California utilizes norm-referenced tests in Grades 2-11 as well as additional criterion-referenced assessment as determined by the California State Board of Education (California Department of Education, n.d.c).

In California, students are tested in Grades 2-11 for English Language Arts and Math. Grades 5, 8, and 10 participate in the Science exam in addition to course specific exams for students in Grades 9-11 in the areas of Biology, Chemistry, Earth Science, Physics, and Integrated/Coordinated Science 1-4. Students in Grades 8 and 11 participate in a History/Social Science exam. Students with disabilities are offered modified and alternate exams as applicable (California Department of Education, n.d. a).

Each school receives an Academic Performance Index (API; California Department of Education, n.d. a). Academic Performance Index ranges from 200 to 1000 and indicates the performance level of the school, district or population group. To calculate the API, student scores are converted into the API format and averaged across the population of the school/district (depending on what is being calculated) for all tests. Academic Performance Index is designed to measure growth and close achievement gaps (California Department of Education, 2010). The specific content areas are weighted in order to calculate the API per student. Depending on grade levels served at the campus for a school with students in Grades K-8, the English Language Arts (ELA) assessment score ranges from 51.4% to 56.5% of the API scores for the school. The Mathematics assessment accounts for anywhere from 34.3% to 37.6%. Science accounts for 5.9% to 7.1%. Social Science is between 3.2% and 7.1%. For schools serving Grades 9-12, ELA is 27.1% of the API score. Mathematics is 18.1%. Science is 22.9%. Social Science is 13.9%. Schools are rated based on the growth of API, which is calculated by subtracting the previous year's base API from the current year's

growth API. Academic Performance Index is calculated for individual schools per state law. In order to meet federal guidelines, it is reported for local educational agencies (California Department of Education, 2010). Schools that fail to meet set growth targets and who are in the lower portion of the state distribution may be selected to participate in programs designed to improve academic performance (California Department of Education, 2011).

As a general rule, the target API for schools in the State of California is 800 as a whole school and for each student group. Depending on previous performance, targets are set for each school and a student group in relation to the base API. For base APIs between 200 and 690, the growth required is 5% of the difference between the base API and 800. For groups between 691 and 795, the group is required to grow 5 points. If the group is between 796 and 799, the growth target is for the group to reach 800. For groups with an 800 API, their goal is to maintain that score. Student groups utilized for accountability in California include African American, American Indian, Asian, Filipino, Hispanic, Native Hawaiian, White, Two or More Races, Socioeconomically Disadvantaged, English Learners, and Students with Disabilities (California Department of Education, 2011). For the 2009-2010 school year, 4,775 out of 8,259 schools met the Target API in California (California Department of Education Assessment and Accountability Division, 2011).

California evaluates student groups in the rating of campus and local educational agency performance. California assesses similar content areas and grade levels as other

state systems; however, the California model of accountability focuses on student and school growth.

#### State of New York

In the State of New York, the accountability system is based on English Language Arts, Mathematics, and elementary and middle school Science exams. Based on assessments, districts are considered in need for improvement if they do not meet AYP for 2 years in English Language Arts or Mathematics at multiple grade levels or fail to meet AYP in Science. Districts are rated on a scale from 1 to 6 based on the number of consecutive years they fail to meet AYP. A rating of 1 indicates good standing. A rating of 2 indicates failure to achieve AYP for 1 year. A 6 indicates failure to meet AYP for 5 years.

Schools are assigned ratings for each measure reported. Areas evaluated with assessments include Grades 3-8 ELA, Mathematics, and Science. Secondary schools are evaluated and assigned a phase based on ELA, Math, and graduation rates. School phases include Good Standing, Improvement Year 1, Improvement Year 2, Corrective Action Year 1, Corrective Action Year 2, Restructuring Year 1, Restructuring Year 2, and Restructuring Advanced.

Schools identified as needing improvement are further categorized based on the initial label. Improvement schools are also labeled Basic, Focused, or Comprehensive. Basic indicates the school failed to meet AYP in one student group area. Focused indicates the school fails to meet AYP in more than one student group. Comprehensive indicates the school failed to meet AYP in the all student groups for English Language

Arts or Mathematics. For schools rated Corrective Action or Restructuring, subcategory options are Focused and Comprehensive. It is possible for schools to have multiple ratings depending on each area examined. To determine the overall rating, the school is labeled with the most progressive label. If a school is labeled in “good standing” for one measure and Corrective Action (Year 2)/Focused, then the overall school rating will be Corrective Action (Year 2)/Focused (New York State Education Department, 2010). The 2011-2012 status of schools in New York indicates that 3,633 of 5,372 schools are labeled as being in Good Standing (NYC Department of Education, n.d.). Student achievement is ranked on a scale from 1-4 with 1 being Basic, 2 indicating Basic Proficient, 3 being Proficient, and 4 meaning Advanced. The student score is then used to calculate the school Performance Index (PI). New York has established a state standard per subject tested. The standard is the lowest PI acceptable. The Science PI is set at 100. ELA and Mathematics PIs are set according to AYP requirements for the school year and increase according to federal guidelines. In addition, the state utilizes the Annual Measurable Objectives (AMO). The AMO measure is a set value for the PI that indicates acceptable progress is being made. Each group that contains 30 or more students must meet or exceed the effective AMO for ELA and Math. For Science, New York also utilizes a progress measure should a group fall below the state target (New York State Education Department, 2010).

## State of Texas

Beginning in 1979, the Legislature created a law which mandated assessment of grades 3, 5, and 9 in the areas of Mathematics, Reading, and Writing (Texas Education Agency, 2009). The Texas Assessment of Basic Skills (TABS) was created and utilized from 1979-1984. Since the state had no mandated curriculum, committees of educators created test objectives. Students were not required to meet standard in order to graduate, but beginning in 1983, ninth-grade students who failed to meet the standard were required to retest throughout high school until mastery was achieved (Texas Education Agency, 2009).

In 1984, the state transitioned to the Texas Educational Assessment of Minimum Skills Test (TEAMS). The change in assessment occurred because of the legislative shift from basic skills to minimum skill requirements. In addition, the legislature required that schools be held accountable and evaluated based on student performance. The state developed the Academic Excellence Indicator System (AEIS) as a result of the mandate with the first report being generated in 1990-1991 (Texas Education Agency, n.d.). The 1985-1986 school year marked the first administration of TEAMS. Objectives were tested at a more rigorous standard and level. In addition, Grades 1, 3, 5, 7, 9, and 11 were assessed in the areas of Reading, Math, and Writing. Meeting mastery level on the Grade 11 exam, considered an exit level exam, became mandatory in order to graduate high school (Texas Education Agency, 2009).

State education law changed in 1990 requiring a criterion referenced exam in Texas. A new instrument, Texas Assessment of Academic Skills (TAAS), was

developed and focused on academic skills based upon Essential Elements, the new state curriculum. Reading, Math, and Writing assessments were conducted in Grades 3, 5, 7, 9, and 11. Not only was TAAS more rigorous, but accountability stretched to campus, districts, and students as the state measured indicators at all levels. Texas Assessment of Academic Skills was utilized until 2002. Throughout the span of time, the test and accountability were altered as directed by the law. Changes included additional assessments for Grades 3-8 in Reading and Mathematics as well as Science and Social Studies in Grade 8; Grade 10 became the exit level and campus, and district ratings were created and yearly growth indicators were added (Texas Education Agency, 2009).

In 1999, Texas legislation mandated a halt to social promotion and focused on creating of a more rigorous assessment. Development of the Texas Assessment of Knowledge and Skills (TAKS) began. Field tested in 2002, the 2003 school year marked the first official administration of TAKS (Texas Education Agency, 2009).

#### Texas Assessment of Knowledge and Skills and Current Accountability System

Currently, Texas continues to utilize TAKS to measure student progress as well as comply with state and national legislation regarding student learning. The state assesses public schools in Grades 3 through 12. Subjects are tested, as shown in Table 1 (Texas Education Agency Department of Assessment, Accountability and Data Quality Division of Performance Reporting, 2009).

Texas state accountability system contains four rating categories. A campus and district can be Exemplary, Recognized, Academically Acceptable, or Academically Unacceptable. Ratings are based upon three domains: student performance on TAKS

exams, completion rates, and attendance rates. In order to receive a specific rating, a campus must meet the requirements for that rating in all three areas (Texas Education Agency Department of Assessment, Accountability and Data Quality Division of Performance Reporting, 2009).

Table 1

*Texas Assessment of Knowledge and Skills (TAKS) Test by Grade and Subject*

Grade	Reading	Writing	English Language Arts (ELA)	Math	Science	Social Studies
Third	X			X		
Fourth	X	X		X		
Fifth	X			X	X	
Sixth	X			X		
Seventh	X	X		X		
Eighth	X			X	X	X
Ninth	X			X		
Tenth			X	X	X	X
Exit Level			X	X	X	X

Texas Assessment of Knowledge and Skills performance is a focus of campuses and districts. The five subjects evaluated through assessment for ratings are Reading and ELA (combined), Math, Science, Social Studies, and Writing. Ratings are assigned based on the percentage of students that meet set standards. Standards are set per test and per grade level. Within each subject, a maximum of five student populations groups are assessed. Groups include: All students, Economically Disadvantaged

students, White students, Hispanic students, and African American students. If a campus or district does not have enough students within a specific student population, that student group is not individually evaluated for campus/district ratings. Table A2 discusses criteria for minimum group size. Depending upon the specific district evaluated, the number of indicators utilized in the rating may vary. One student's score may count in up to three indicators; for example, a White, Economically Disadvantaged student's seventh grade Mathematics results will count in the All Student, Economically Disadvantaged, and White categories (see Appendix A, Table A1). The table indicates the criterion each campus and district must attain in order to receive a specific rating (Texas Education Agency Department of Assessment, Accountability and Data Quality Division of Performance Reporting, 2009). District ratings are published and compared (McNeil, 2000), by those who read the reports, regardless of indicator variance between the districts.

In addition to the passing percentage establishing the district ranking, Texas created three alternative measures to allow districts to achieve a higher level rating. The first is called Recognized Improvement (RI). Required Improvement measures growth in a category from one year to the next. Should a campus achieve half the distance between the previous year's rating and the next higher rating in one academic year, the state awards the higher rating based on the improvement (Texas Education Agency Department of Assessment, Accountability and Data Quality Division of Performance Reporting, 2009).

The second measure is called Texas Projection Measure (TPM). Texas Projection Measure is based upon a multiple regression and projects students who will attain set standards by specified cut years. If a student, who is currently below standard, is projected to reach standard, the score is counted as a met standard (Zyskowski & O'Malley, 2009).

The final alternative is called an "Exemption." Based upon the number of indicators a district is ranked on, a set number of exemptions are available. Should a district miss achieving the next level ranking by specific indicators, a district can use an exemption for specified categories as long as they are within the number of exemptions allocated. Appendix A illustrates guidelines for all components (Texas Education Agency Department of Assessment, Accountability and Data Quality Division of Performance Reporting, 2009).

Students with special needs can receive accommodation based on their needs as well as an alternate version of the TAKS test. Qualifying students must meet a specific criterion and a committee must agree to the accommodation. Texas Assessment of Knowledge and Skills versions include TAKS Accommodated, TAKS Modified, and TAKS-Alt. In 2009, TAKS Accommodated results were included in campus ratings only in specified subjects and grades (Appendix A, Table A2). In subsequent administrations, TAKS Accommodated is included in additional areas. Texas Assessment of Knowledge and Skills Modified and TAKS-Alt apply to federal ratings (Texas Education Agency Department of Assessment, Accountability and Data Quality Division of Performance Reporting, 2009).

The Texas accountability system currently takes into account dropout rates as part of the district and campus accountability ratings. An increase in the number of poverty students and minority students who drop out of high school has been noted since the onset of this program (Foote, 2007), presumably because they cannot meet this standard. Current accountability systems are responsible for a reduction in career and technical programs in many districts and states. Every state that has reduced these programs to focus on testing has experienced a rise in the student dropout rate (Lewis, 2005).

As of 2000, there was a 22% difference in graduation rates of Hispanic students when compared to White students with 88% of White students graduating while only 63% of Hispanic students graduated (Tressler, 2010, p. 29). According to the 2009 dropout report, the Hispanic dropout rate for 2006-2007 was 5.6%, African American was 5.8%, and White was 1.9% (U.S. Department of Education, 2009).

Accountability ratings were created to demonstrate a district's performance in regards to meeting state standards. Now ratings are a badge of honor. Administrators want to work at Recognized and Exemplary buildings and campuses (Nelson et al., 2007). Achieving the rating has become a large focus for Texas school districts with districts investing in data analysts and specialists to help districts achieve higher ratings through strategies rather than focus on improved teaching technique (Nelson et al., 2007). Student performance and population group performance are evaluated thoroughly. The system, while based on student performance, is further complicated by allowing some student scores to be counted multiple times in the ratings process; for

instance, if a student is African American and Economically Disadvantaged, his/her score will apply for each subject area test taken three times—once for the all student indicators, once for the African American indicator, and once for the Economically Disadvantaged indicator. In comparison, an Asian student who is not Economically Disadvantaged will have his/her score count once per subject area test taken in the all-student category. In essence, some student scores are more significant on a rating scale than others based on the fact that the score is utilized up to three times (Texas Education Agency Department of Assessment, Accountability and Data Quality Division of Performance Reporting, 2008).

Critics of the Texas and NCLB accountability system cite the lack of growth measure as a main drawback to the current system. Poor schools often demonstrate the most growth, but are not rewarded because they were so far behind to begin with. In contrast, wealthier schools performed well from the beginning and continue to perform at the same standard each year. Some high-performing schools' scores drop each year, but they maintain a high rating simply because of where the scores start. These schools are awarded with top rankings while truly achieving very little growth or change in student performance (Hursh, 2005).

To summarize, the states of California, New York, and Texas all evaluate districts based on standardized assessments. California utilizes a numeric reporting index. Texas utilizes a categorical rating. New York utilizes both a numeric and categorical rating. The states look at student demographic in the analysis. California

disaggregates the data for more student population groups when compared to New York and Texas. Student demographics impact publicized ratings.

### Changing Demographics

From 2000 to 2010, the United States population increased by 9.7%. During the same time frame, the Hispanic population grew by 43%: “More than half of the growth in the total population of the United States between 2000 and 2010 was due to the increase in the Hispanic population” (Humes, Jones, & Ramirez, 2011, p. 3). Also during the past decade, Texas showed an overall population increase of 20.6 % (Mackun & Wilson, 2011). In Texas, those claiming race and ethnicity as being anything other than non-Hispanic, White increased by 38.6% moving Texas into a “majority-minority” state (Humes et al., 2011). The 2009 census data indicate that 17.1% of Texas residents live below the poverty level compared with the nation’s average of 14.3% (United States Census Bureau, 2011). In 2000, the national poverty rate was 11.3% while Texas reported a poverty rate of 14.9% (Dalaker, 2001). In the last decade, the Texas rate of poverty has increased by 2.2%.

The demographic profile of students is changing as evidenced by the 2010 census data. According to Murdock (2006), the demographic profile will continue to change with projections for 2040 showing that nearly 70% of children under the age of 5 will be Hispanic while nearly 18% will be Anglo. The performance of students based on demographic grouping is going to become more popular with the specified groups containing larger numbers of students that will count in multiple accountability indicator groups under the Texas accountability plan, ultimately leading to a stronger

push to close educational gaps between students of various demographic backgrounds. These shifts in demographics coupled with gaps in level of proficient performance amongst various ethnic/racial groups could ultimately impact the United States' work force negatively in future years (Callan, Finney, Kirst, Usdan, & Venezia, 2006).

### Demographic Factors in Accountability

According to Derthick and Dunn (2009),

The trouble, keenly felt by the nation's teachers, is that the performance of school children depends not only on the motivation, effort, and skill of their teachers, but also on a host of social, economic, cultural, and psychological factors that are beyond the reach of the schools. (p. 1031-1032)

Closing achievement gaps between White students and other ethnicities is a goal of education legislation (Tressler, 2010). No Child Left Behind was created to remedy the educational crisis created by advocating for improving education of those students who are traditionally left behind (Darling-Hammond, 2007). Proponents of accountability state that current systems are the only way to guarantee that students of poverty and color receive adequate educations and are held to the same high standards (Hursh, 2005). State and federal accountability advocates for closing performance gaps among specified historically underperforming population groups. These groups include Hispanic, African American, and low-socioeconomic students (Yeop Kim, Zabel, Stiefel, & Schwartz, 2006). Accountability systems have been created to address concerns and assure that students from specified groups receive an adequate education.

The assumption of current legislation is that the children of poverty and from minority populations will benefit greatly from the accountability movement (Townsend, 2002). Historically, children from low-income families are less educated (Rouse & Barrow, 2006). Family income and test scores tend to be highly correlated, according to Hursh (2005). Hanushek (2004) noted that a causal link between student performance and poverty had not been established.

Texas is one state where the growth of the Hispanic student population is increasing rapidly. When combined with the fact that the Hispanic student is more likely to come from low income, less educated home when compared to a White student (Tressler, 2010), schools are working feverishly to educate a population that is changing. The Texas accountability system was developed in order to help create equity between White and Hispanic populations (Flinders, 2005). Florida officials cited their accountability system as beneficial based partially on gains made by low income and African American student populations who attended schools that were labeled low performing the previous year (as cited in West & Peterson, 2005). The low rating encouraged the schools to make improvements that positively affected the two student groups in the State of Florida.

Ultimately, the goal of testing and accountability is to narrow the educational achievement gap between various student groups. “The current conditions of schooling for many students of color and low-income students in the United States strongly resemble those that existed before *Brown vs. Board of Education* (1954) sought to end separate and unequal education” (Darling-Hammond, 2007, p. 225). Despite increased

accountability, research shows that high-performing schools do not always meet the needs of these special population groups (Yeop Kim et al., 2006). In addition, assumptions are made based on ratings. One such assumption is that schools with a large population of low income and/or minority students that have been labeled as “low performing” based on students not meeting state standards creates the illusion that students are effectively taught but incapable of mastering the content (Townsend, 2002). Furthermore, racial composition of schools impacts African American student performance but not the performance of White and Hispanic students (Hanushek, 2004). Many programs have been created to accommodate for low-income students; however, these programs have not eliminated the disparity in educational outcomes between those who come from affluence and those who do not (Rouse & Barrow, 2006).

Accountability systems also could encourage districts to “play the numbers” regarding minority students by gerrymandering school boundaries. By shuffling students around, a district can create a school where minority populations are not large enough to be considered as an individual group, or the district could choose to group all of one population together so that the larger number of students in that group will cover for those who struggle (Kane & Staiger, 2002).

Current accountability models lead to a disproportionate number of predominately minority and impoverished schools not meeting national and state accountability criteria (Townsend, 2002), fueling concern that the poorest schools will be denied funding under the NCLB legislation. Ironically, the wealthier schools under these circumstances will receive more funding because of satisfactory performance,

while the poor schools will only get poorer, thus sorely increasing the performance gaps between the two-indicator groups (Flinders, 2005). To compound this fear, the best educators are not drawn to the campuses with the greatest need because of fears of being labeled low performing (Nelson et al., 2007).

### Class Size and Accountability

Research suggests that class size as well as family income influences student performance (Rouse & Barrow, 2006). In a study on student to teacher ratios and student performance, Alspaugh (1994) found an increase in achievement at the one grade level that was examined. Further research should be done to see if the increase holds true for multiple grade levels. In 2002, Wößsmann and West (2002) concluded that the effect of class size on student performance was not large enough to justify the costs incurred by reducing class size. They further noted that both Greece and Iceland exhibited improvement in student performance with smaller class sizes along with discussion on teacher training as the main factor. As cited in Chingos' 2010 class reduction study, Jepsen and Rivkin (2009) found class size affected performance in California, but the influx of teachers needed to meet the reduction mandate allowed for less qualified teachers to be hired. The lack of quality teachers mainly impacted the minority populations and therefore skewed the results. The State of Florida enacted a similar class size reduction mandate in which research indicates that the reduction had no significant impact on student performance (Chingos, 2010). Low student to teacher ratios have been shown to close educational gaps among students (Nelson et al., 2007).

## Educational Financing and Accountability

Linking financial data to performance outcomes is rarely calculated by districts (Hess & Fullerton, 2009). In 2004-2005, it was estimated that \$536 billion was contributed by the government towards educating K-12 students (U.S. Department of Education, 2007). According to court rulings, states are required to fund education adequately as defined by their state constitution. In Texas, school finance has been an important topic over the past several years. Court battles and lack of finances have contributed to the debate of how to adequately fund education (Nelson et al., 2007).

The Texas constitution provides for an “efficient” education. Court rulings then hold the state to funding education adequately so that an efficient education can be attained by *all* students in the state. Texas utilizes both vertical equity and locality supplementation in the state funding formula for education (Umpstead, 2007). The Texas formula for funding education is quite complex, with funds being generated through the general revenue fund as well as local property taxes. Districts with large property values generate more money as a district than those with lesser property values. Texas created a recapture program under Chapter 41 of the Texas Education Code that redistributes this wealth among districts. The program has become commonly known as “Robin Hood” because it “takes from the rich and gives to the poor.” Texas also utilizes a target revenue system in which the state funding is based on the weighted average daily attendance (WADA) of students; each district has a different target revenue set by the state. The differing target revenues contribute to

districts having a variance in dollar allotment per student (Texas Association of School Boards, 2008).

Accountability systems are used by states to determine whether or not schools are meeting student needs and fulfilling their requirement to adequately fund education (Umpstead, 2007). According to Nelson et al. (2007), “to uphold the integrity of the educational accountability system, adequate funding to maintain support systems must be afforded to public schools” (p. 708). Jaekyung (2010) found that funding gaps could be a result of the performance standard gaps that exists within accountability systems. With estimates ranging from schools needing 20-35% more funding to meet the mandates set by NCLB (Peyser & Costrell, 2004), funding and accountability ratings should be related.

Education policy cannot account for all of the variables that students face outside of school. Accountability systems can only generate small improvements; they cannot close all performance gaps by applying pressure to districts in the form of ratings. Closing achievement gaps despite background is a goal that the U.S. education system is still far from (Rouse & Barrow, 2006).

As described in the previous sections, demographics (Derthick & Dunn, 2011), class size and funding (Nelson et al., 2007) have all been linked with student performance. Improving student performance was the purpose of the accountability system (Hopmann, 2008; Johnson et al., 2011).

## College and Career Readiness and Accountability

In the educational community, there is much discussion regarding K-12 preparation of students for higher education. In 2005, it was estimated that nearly half of all students graduating were unprepared for collegiate level coursework in the United States (Lewin, 2005). Research further indicates that student perceived preparedness for college was related to the courses students participated in while in high school (Goodwin, 2008). According to Edge (2009), fewer post graduate Economically Disadvantaged students were ready for college level Mathematics courses when compared with their peers. In addition, it is estimated that 84% of Hispanic students leave high school unprepared for higher education and an estimated 80% of African American students leave high school unprepared as well (Greene & Forster, 2003). Historically, schools were created to provide minimal education, but today's society encourages more students to aspire to college-level coursework in order to attain skills necessary for employment, but based on demand, more students must be prepared for collegiate work, and the old system of minimal education should be reevaluated (Callan et al., 2004).

In an effort to address the growing concerns regarding student preparation for college, the State of Texas began several initiatives: House Bill 1, passed in 2006, allocates \$275 per student to support college readiness goals. In addition, curriculum redesign has been focused on preparing students for coursework (Early Colleges, 2009). Texas is also addressing college readiness through the state-testing program. The Texas accountability system began issuing Gold Performance acknowledgements based on

district and campus performance in the area of college readiness (Texas Education Agency Department of Assessment, Accountability and Data Quality Division of Performance Reporting, 2008). Reported under Texas Success Initiative (TSI) ELA or Texas Success Initiative (TSI) Math, the measures indicate the percentage of students considered to be ready for college-level coursework in the respective subject areas based upon student performance in Grade 11 on the Exit Level ELA and/or Mathematics TAKS test. The TSI was first reported via the AEIS report in 2006 (Texas Education Agency, 2011b). Districts receive information regarding student performance in relationship to college and career readiness for the Exit Level Reading and Mathematics TAKS tests.

Ongoing discussion regarding reauthorization of NCLB and accountability focuses on linking accountability to student college and career readiness (Schmidt, 2011). Proponents are calling for an alignment of K-12 standards with collegiate standards (Callan et al., 2004).

#### Summary of Literature Review

Assessment and accountability are both necessary to ensure that students in the United States are receiving the education promised to them. The current systems in place in the State of Texas—state and federal—both monitor accountability in the same manner. Both systems rank campuses based on performance on TAKS. Furthermore, both systems isolate student groups and their performance to determine ratings. Schools, based on demographics, are often held to varying standards based on the nature of the student population served.

As indicated by Rouse and Barrow (2006), Hursh (2005), Townsend (2002), and Yeop Kim et al. (2006) students of color and poverty face more obstacles in demonstrating proficiency on standardized assessments. Schools with higher populations of these student groups must overcome great obstacles to achieve ratings similar to their counterparts that serve middle class White populations.

The current accountability system was created to help erase disparity in the education students in Texas receive; however, the current model does not account for growth and student improvement. It is simply a measure that is compared with a set standard—growth is not acknowledged. No measure of accountability demonstrates if schools are truly improving. The system needs further examination to determine if it measures instructional efficacy and teaching adequately.

## CHAPTER 3

### METHOD OF PROCEDURE

This study explored whether district accountability ratings issued by the State of Texas could be predicted by a district's funding per student, class-size ratio the number of student ethnicity/ race (African American, Hispanic, and White) utilized in a district rating as well as, the Economically Disadvantaged group indicator. The study also explored the relationship between district rating and class size ratio, as well as the relationship between the district rating and funding per student. In addition, the study evaluated the relationship between a district's college and career readiness statistics and district ratings.

#### Research Questions

1. Can a Texas school district's rating of *Academically Acceptable* or *Academically Unacceptable* be predicted based upon one or more of the following variables: the district's number of race/ethnic groups utilized in calculating the rating, district's economically disadvantaged indicator, the district's money spent per student, and/or the district's teacher/student ratio?
2. Can a Texas school district's rating of *Recognized* or *Academically Acceptable* be predicted based upon one or more of the following variables: the district's number of race/ethnic groups utilized in calculating the rating,

district's economically disadvantaged indicator, the district's money spent per student, and/ or the district's teacher/student ratio?

3. Can a Texas school district's rating of *Exemplary* or *Recognized* be predicted based upon one or more of the following variables: the district's number of race/ethnic groups utilized in calculating the rating, district's economically disadvantaged indicator, the district's money spent per student, and/or the district's teacher/student ratio?
4. Is there a relationship between district accountability ratings and district Mathematics College and Career Readiness rates?
5. Is there a relationship between district accountability ratings and district English Language Arts College and Career Readiness rates?

#### Research Design

This study utilized two separate statistical procedures. Logistic regressions were used as the statistical procedure for Research Questions 1 through 3. Point biserial correlation was utilized for Research Questions 4 and 5.

Logistic regression was the appropriate statistical method for Research Questions 1 through 3 based on the fact that the analysis was concerned with exploring the relationship between several independent variables and their ability to predict the outcome of a dependent variable that was dichotomous. By structuring each logistic regression so that only two ratings were being compared at a time, the dependent variable was dichotomous. The intention of the study was to find the combination of independent variables that allowed for the best prediction of the rating of a district

(Hosmer & Lemeshow, 2000). Furthermore, logistic regression is the appropriate statistical procedure to utilize because it is flexible and requires that no assumptions be made regarding the distribution of the predictor variables. Logistic regression yields adequate and appropriate results for a multitude of variables at differing statistical levels (Tabachnick & Fidell, 2007).

Point biserial correlation was the appropriate measure for Research Questions 4 and 5 due to the fact that one of the variables in each question was coded to be dichotomous (Salkind, 2004). The percentage of students ready for college was an interval-level continuous variable.

#### Procedure

The data utilized for this study were the 2008 Texas district ratings and the indicators from the 2007-2008 school year for all public school districts in the State of Texas. Permission to proceed with research from the University of Texas at Arlington was received in May 2011 through the Institutional Review Board (IRB). The 2008 data were utilized because of the creation and implementation of the Texas Projection Measure (TPM) in 2009. TPM was utilized in both years of 2009 and 2010 ultimately altering district ratings based on projected student growth in the future. For the purpose of this study, TPM inflated the data and skewed ratings, so 2008 data were utilized in this study. TPM was removed from use in 2011 due to a public outcry.

#### Collection of the Data

Data utilized for the study was available through the TEA website, and was available as part of public record. Data were collected from the Texas Education

Agency 2008 Accountability System Data Download site and the 2007-2008 Academic Excellence Indicator System Data Download Options site. Five separate spreadsheets were downloaded from these sites in order to gather the data necessary to complete the study. The five spreadsheets consisted of the following:

1. A spreadsheet that contained the county district number of each district and 2008 district ratings.
2. A spreadsheet that contained the county district number of each district and the indicators that were utilized for that district rating. The spreadsheet provided the information needed for the African American, Hispanic, White, and Economically Disadvantaged variables. As defined by TEA in the *2008 Accountability Manual*, the use of race and Economically Disadvantaged indicators was determined by the number of students testing in the district that belong to the group. If the district had between 30-49 students in a category, and that represented 10% of the total testing population, then the indicator applied towards ratings. If the district had 50 students or more in the whole testing group, the indicator was applied towards those ratings. If neither scenario applied towards a district, the indicator was not utilized for that district. The race/ethnic indicators were then totaled to determine the number of student race/ethnic groups utilized in the rating calculation.
3. A spreadsheet contained the county district number of each district and the financial statistics reported in the 2007-2008 AEIS report. The spreadsheet provided the cost spent per pupil by each district.

4. A spreadsheet contained the county district number for each district and the staffing statistics. The spreadsheet provided the student/teacher ratio data.
5. A spreadsheet contained the county district number for each district and the college and career readiness percentages for Reading and Math.

For the study, the five spreadsheets were combined utilizing the state issued six digit district-county numbers. Once combined, the data were evaluated for completeness. Only districts with data available for all variables were utilized; districts with missing data were removed from the database.

#### Treatment of Data

Statistical Package for Social Sciences (SPSS) version 19.0 was utilized for all statistical analyses conducted as part of this study. Logistic regressions were utilized for Research Questions 1 through 3. Point biserial correlations were utilized for Research Questions 4 and 5.

#### Research Questions 1, 2, and 3

After data was downloaded from the TEA website and merged into one database, the data necessary for each research question was identified and placed in an SPSS data file. Five SPSS data files were created and coded as detailed below. The dependent variable, district ratings, was coded as follows:

Research Question 1. *Academically Acceptable* = 1 and *Academically Unacceptable* = 0.

Research Question 2. *Recognized* = 1 and *Academically Acceptable* = 0.

Research Question 3. *Exemplary* = 1 and *Recognized* = 0.

For all databases, the independent variable, *Economically Disadvantaged* indicator was coded as follows:

0 = Did not count for district as an accountability indicator

1 = Did count for district as an accountability indicator

Race/ethnicity codes for the African American, Hispanic, and White students groups were then combined together and assigned one of three coding values based on the number of race/ethnicity groups that were utilized in the district rating were as follows:

1 = one race/ethnicity group

2 = two race/ethnicity groups

3 = three race/ethnicity groups

The student demographic indicators for each district were determined based on the Mathematics TAKS. The choice to utilize the Mathematics TAKS test student group indicators were based on the fact that all grade levels third through eleventh participated in Mathematics exams. Other subject areas do not test all grade levels. Refer to Table 1. The use of the Mathematics test allowed for the largest student population possible to be included in the study because all students in grades 3 through 11 participate in a Mathematics TAKS test.

The remaining two independent variables, student teacher ratio and amount spent per student, are each reported by districts and then made available to the public by TEA. The actual value reported by TEA was utilized in this study. A logistic regression was run to determine which of the six independent variables, if any, could

predict most accurately between the two ratings stated in the Research Questions 1 through 3.

SPSS was utilized to run descriptive statistics and a logistic regression for each research question. The output of the regression was analyzed to determine which variables were significant in the model. The variables that were not significant were then removed from the model. The logistic regression was run a second time containing only significant variables. The Hosmer and Lemeshow goodness-of-fit test was calculated to determine if the results modeled were adequate. A nonsignificant result confirmed the model was sufficient (Tabachnick & Fidell, 2007).

#### Research Questions 4 and 5

For Research Questions 4 and 5, the following methodology was utilized. Data were downloaded from the TEA website. The variable, district rating, was coded as follows:

0 = *Academically Unacceptable* and *Academically Acceptable*

1 = *Recognized* and *Exemplary*

The variable, percentage of students who were college and career ready as indicated by the Texas success initiative—higher education readiness components for English Language Arts and Math, is reported by TEA. The actual value reported by TEA was utilized in this study. A point biserial correlation was run to determine the correlation coefficient that existed between college and career readiness percentages and the district rating.

## Summary of Methods

This study evaluated the relationship between accountability indicators and district accountability ratings. It also explored teacher/student ratio, district finances, and college and career readiness indicators and the relationship to accountability ratings. The study utilized each public school districts with complete data available.

## CHAPTER 4

### PRESENTATION AND ANALYSIS OF DATA

The major purpose of this study was to determine what relationship exists between student to teacher ratio, funding per student, the number of student ethnicity/ race (African American, Hispanic, and White) utilized in a district rating as well as, the Economically Disadvantaged group indicator and district ratings in the State of Texas as well as explore the relationship between ratings and college readiness in the areas of Mathematics and Reading. Logistic regression was utilized to determine the relationship between the district's number of race/ethnic groups utilized in calculating the rating, district's economically disadvantaged indicator, the district's money spent per student, and/or the district's teacher/student ratio and ratings. Point biserial correlations were calculated to determine the relationships between ratings and college readiness.

From databases provided online by TEA, data of 967 districts were utilized in this study. Data included each district's 2008 accountability rating as well as information regarding usage of each student group in factoring the rating. Three separate databases reported the district's allotment per pupil, student-to-teacher ratio, and college readiness rates. The data from the databases were merged utilizing the county district number. Districts that did not have all data available were not utilized.

## Data Characteristics

SPSS version 19.0 was utilized in the analysis of the data. Of the 967 districts the study included, 1.2% of the districts were *Academically Unacceptable*, 70.4% *Academically Acceptable*, 26.5% *Recognized*, and 1.9% *Exemplary*. Of districts studied, 94.7% districts had a White student group apply toward their accountability rating, 97% of districts had an Economically Disadvantaged group apply towards their rating, 38.6% of districts had an African American student group, and 75.6% of districts had a Hispanic student group. Table 2 shows specific values for each student group. The average amount spent by each district per student was \$10,950.71. The average student to teacher ratio per district was 1 teacher for every 12.38 students. Tables 2 and 3 display the descriptive statistics for these categories.

Table 2

*Descriptive Statistics of Student Population Groups for All Districts Utilized in Study*

Group Size	<i>N</i> = 967	AA group met size requirement	ED group met size requirement	Hispanic group met size requirement	White group met size requirement
Group met minimum size		373	940	731	916
Group did not meet minimum size		594	27	236	51

*Note.* AA represents the African American group and ED represents the Economically Disadvantaged student group.

Table 3

*Descriptive Statistics of Non Population Group Variables for All Districts Utilized in the Study*

Variable	<i>N</i> = 967	Mean	<i>SD</i>
Student to teacher ratio		12.38	2.26
Amount spent per pupil		\$10,950.71	\$3,233.38

Table 4 displays data regarding the district rating and descriptive statistics of the districts with each rating. Of the 12 *Academically Unacceptable* districts, 58.3% of the districts had an African American group apply towards the rating; 91.7% of *Academically Unacceptable* districts had an Economically Disadvantaged student group. Hispanic student groups were present for 75% of *Academically Unacceptable* group, 681 districts were rated *Academically Acceptable* in 2008, 45.5% of these districts had an African American student group, 99.1% had an Economically Disadvantaged student group, 81.9% of *Academically Acceptable* districts had a Hispanic student group, 94.0% had a White student group. For the districts rated *Recognized*, 20.7% had an African American student group, 94.1% had an Economically Disadvantaged student group, 61.7% had a Hispanic student group, and 97.7% had a White student group. Of the 18 districts rated Exemplary, 16.7% had an African American student group, 72.2% had an Economically Disadvantaged group, 33.3% had a Hispanic group, and 100% had a White group.

Table 4

*Select Descriptive Statistics of Population Groups by Rating Category for All Districts Utilized in Study*

Rating	N	% of total	% AA group met size requirement	% ED group met size requirement	% Hispanic group met size requirement	% White group met size requirement
Academically Unacceptable	12	1.2%	58.3%	91.7%	75%	66.7%
Academically Acceptable	681	70.4%	45.5%	99.1%	81.9%	94.0%
Recognized	256	26.5%	20.7%	94.1%	61.7%	97.7%
Exemplary	18	1.9%	16.7%	72.2%	33.3%	100%

*Note.* AA represents the African American group and ED represents the Economically Disadvantaged student group.

As seen in Table 5, 32.4% of the districts utilized in this study had 3 race/ ethnic groups indicators utilized in calculating their rating, 44.2% of the districts utilized in this study had 2 race/ ethnic groups indicators utilized in calculating their rating, and 23.5% of the districts utilized in this study had 1 race/ ethnic group indicator utilized in calculating their rating.

Table 5

*Crosstab Calculation of Number of Race/ Ethnic Groups by Rating Category*

Rating	N	1 race/ ethnic group	2 race/ ethnic groups	3 race/ ethnic groups
Academically Unacceptable	12	41.7%	17.7%	41.7%
Academically Acceptable	681	16.3%	46.0%	37.7%
Recognized	256	38.7%	42.6%	18.8%
Exemplary	18	66.7%	16.7%	16.7%
Total	967	23.5%	44.2%	32.4%

#### Descriptive Statistics by Rating Category

Table 6 displays that the average amount spent per student in an *Academically Unacceptable* district was \$11,202. The average student to teacher ratio for an *Academically Unacceptable* district was 11.7 students for every 1 teacher.

Table 6

*Select Descriptive Statistics of Non Population Group Variables for Academically Unacceptable Districts Utilized in Study*

Variable	N = 12	Mean	SD
Amount spent per pupil		\$11,202	\$4,077.29
Student to Teacher Ratio		11.7	3.57

Table 7 displays that the average amount spent per student in an *Academically Acceptable* district was \$10,804.28. The average student to teacher ratio for *Academically Acceptable* districts was 12.57 students for every 1 teacher.

Table 7

*Select Descriptive Statistics of Non Population Group Variables for Academically Acceptable Districts Utilized in Study*

Variable	<i>N</i> = 681	Mean	<i>SD</i>
Amount spent per pupil		\$10,804.28	\$2,813.74
Student to Teacher Ratio		12.57	2.16

Table 8 shows that the average amount spent per student in a *Recognized* district was \$11,306.29. The average student to teacher ratio for *Recognized* districts was 12 students to every 1 teacher.

Table 8

*Select Descriptive Statistics of Non Population Group Variables for Recognized Districts Utilized in Study*

Variable	<i>N</i> = 256	Mean	<i>SD</i>
Amount spent per pupil		\$11,306.29	\$3,984.63
Student to Teacher Ratio		12.0	2.36

Table 9 shows that the average amount spent per student in an *Exemplary* district was \$11,265.94. The average student to teacher ratio in *Exemplary* districts ratio was 11.3 students to every teacher.

Table 9

*Select Descriptive Statistics of Non Population Group Variables for Exemplary Districts Utilized in Study*

Variable	<i>N</i> = 18	Mean	<i>SD</i>
Amount spent per pupil		\$11,265.94	\$5,055.62
Student to Teacher Ratio		11.3	2.83

## Analysis of Research Questions

### Research Question 1

Can a Texas school district's rating of *Academically Acceptable* or *Academically Unacceptable* be predicted based upon one or more of the following variables: the district's number of race/ethnic groups utilized in calculating the rating, district's economically disadvantaged indicator, the district's money spent per student, and/or the district's teacher/student ratio?

Table 10 displays the results of the logistic regression performed regarding *Academically Acceptable* and *Academically Unacceptable* ratings. Of the variables selected for the study, none were significant in predicting the ratings of *Academically Acceptable* or *Academically Unacceptable*.

Table 10

*Logistic Regression Analysis of Rating Categories Academically Acceptable and Academically Unacceptable with All Variables*

Predictor	B	S.E.	Wald	df	Sig.	Exp(B)
ethgrp#	.164	.484	.115	1	.734	1.179
Econ. Disadv.	1.758	1.364	1.661	1	.197	5.801
\$ Spent	.000	.000	.039	1	.843	1.000
Teacher Ratio	.086	.170	.255	1	.614	1.090
Constant	.695	2.588	.072	1	.788	2.005
Goodness of Fit Test			X2	df	Sig.	
			10.862	8	.210	

*Note.* Where ethgrp# equaled the number of race/ethnic groups a district had applied in their rating, Econ. Disadv. represents the Economically Disadvantaged indicator, \$spent represented the funds spent per pupil, and Teacher Ratio represents the number of students for every teacher.

## Research Question 2

Can a Texas school district's rating of *Recognized* or *Academically Acceptable* be predicted based upon one or more of the following variables: the district's number of race/ethnic groups utilized in calculating the rating, district's economically disadvantaged indicator, the district's money spent per student, and/ or the district's teacher/student ratio?

When including all variables, the logistic regression two variables were identified as significant. The variables number of race/ethnic groups utilized in calculating the rating and economically disadvantaged being statistically significant (Table 11). Race/ethnic group (Wald (1) = 41.475,  $p = .000$ ) and economically disadvantaged (Wald (1) = 7.082,  $p = .008$ ) were then entered into a subsequent logistic regression in order to create a predictor equation.

Table 11

*Logistic Regression Analysis of Rating Categories Recognized and Academically Acceptable with All Variables*

Predictor	B	S.E.	Wald	df	Sig.	Exp(B)
ethgrp#	-.782	.121	41.475	1	.000	.457
Econ. Disadv.	-1.360	.511	7.082	1	.008	.257
\$ Spent	.000	.000	1.706	1	.192	1.000
Teacher Ratio	.045	.043	1.095	1	.295	1.046
Constant	1.009	.742	1.847	1	.174	2.742
Goodness of Fit Test			X2	df	Sig.	
			13.974	8	.082	

*Note.* Where ethgrp# equaled the number of race/ethnic groups a district had applied in their rating, Econ. Disadv. represents the Economically Disadvantaged indicator, \$spent represented the funds spent per pupil, and Teacher Ratio represents the number of students for every teacher.

Table 12 displays the results of the logistic regression with non significant variables removed. The number of race/ethnic groups (Wald (1) = 46.646,  $p = .000$ ) impacted the odds of being rated *Recognized* compared to *Academically Acceptable* . With one unit increase in the number of ethnic groups applied to their rating, the odds of being rated *Recognized* compared to *Academically Acceptable* was decreased using a factor of 0.479. In addition, the odds of being rated *Recognized* versus *Academically Acceptable* decreased using a factor of 0.271 if the district has a disadvantaged group. Nagelkerke  $R$  square was computed in an effort to account for variance. It is a pseudo  $R^2$  (Tabachnick & Fidell, 2007). When comparing *Recognized* to *Academically Acceptable*, Nagelkerke  $R$  squared = .100, indicating that 10% of the variance in ratings could be attributed to the variables selected.

Table 12

*Logistic Regression Analysis of Rating Categories Recognized and Academically Acceptable with Significant Variables*

Predictor	B	S.E.	Wald	df	Sig.	Exp(B)
ethgrp#	-.736	.108	46.646	1	.000	.479
Econ. Disadv.	-1.307	.504	6.729	1	.009	.271
Constant	1.774	.507	12.253	1	.000	5.894
Goodness of Fit Test			X2	df	Sig.	
			2.029	2	.363	

*Note.* Where ethgrp# equaled the number of race/ethnic groups a district had applied in their rating, Econ. Disadv. represent the Economically Disadvantaged indicator..

Table 13 illustrates the relationship between actual observations and predicted outcome for the first ten districts utilized. A 2-variable predictor model was utilized to generate a predictor equation concerning the relationship between the rating of *Recognized* and *Academically Acceptable*. The results generated the following equation where ethgrp# represents the number of race/ethnic groups applied towards rating and Econ. Disadv. represents if the district had an economically disadvantaged student group on the Mathematics assessment:

$$\text{Predicted logit of (RATING)} = 1.774 + (-0.736) * \text{ethgrp\#} + (-1.307) * \text{Econ. Disadv.}$$

The equation can be utilized to help predict a rating of *Recognized* or *Academically Acceptable*.

Table 13

*Predicted Probability of Being Rated Recognized versus Academically Acceptable for 10 Districts*

Case Num.	# race/ ethnic groups	EcDis $\beta = -1.044$	Intercept $= -2.731$	Pred Prob	Actual Rating 1= Rec 0=Acc
1	2.00	1	1.774	.26808	1
2	1.00	1	1.774	.43324	0
3	1.00	1	1.774	.43324	0
4	3.00	1	1.774	.14930	0
5	3.00	1	1.774	.14930	0
6	1.00	1	1.774	.43324	1
7	2.00	1	1.774	.26808	0
8	2.00	1	1.774	.26808	1
9	3.00	1	1.774	.14930	0
10	1.00	1	1.774	.43324	1

Further exploration of the race/ethnic group category was conducted. The crosstabs outcomes of the distribution of race/ethnicity groups between the two ratings are provided in Table 14. When all three race/ethnicity groups count in a rating, 48 out of 305 districts received a *Recognized* rating. When all three race/ethnicity groups counted in a rating, 257 out of 305 were *Academically Acceptable*.

Table 14

*Cross Tabulation of Race/Ethnicity Indicators Utilized in Calculating Rating and Rating Assigned to District: Recognized to Academically Acceptable*

Number of race/ ethnicity groups	Rating		Total
	Recognized	Academically Acceptable	
1 race/ethnic group	99	111	210
2 race/ethnic groups	109	313	422
3 race/ethnic groups	48	257	305
Total	256	681	937

Overall prediction revealed that on the basis of 2 variables the predicted ratings were correct for 5.5% of *Recognized* districts and 99.1% of *Academically Acceptable*.

Overall the model was correct 73.5% of the time. The model over classified *Recognized* districts as *Academically Acceptable* as illustrated by Table 15.

Table 15

*Accuracy of Predicted Ratings According to Predicted Logit Equation: Recognized to Academically Acceptable*

Observed	Predicted		Percentage Correct
	Recognized	Academically Acceptable	
Recognized	14	242	5.5%
Academically Acceptable	6	675	99.1%
Overall Percentage			73.5%

### Research Question 3

Can a Texas school district's rating of *Exemplary* or *Recognized* be predicted based upon one or more of the following variables: the district's number of race/ethnic groups utilized in calculating the rating, district's economically disadvantaged indicator, the district's money spent per student, and/or the district's teacher/student ratio?

When all variables were loaded into the model, the logistic regression revealed that only the economically disadvantaged indicator (Wald (1) =6.553,  $p = .010$ ) was significant. See Table 16 for further details.

Table 16

*Logistic Regression Analysis of Rating Categories Exemplary and Recognized with All Variables*

Predictor	B	S.E.	Wald	Df	Sig.	Exp(B)
ethgrp#	-.378	.452	.700	1	.403	.685
Econ. Disadv.	.1.620	.633	6.553	1	.010	.198
\$ Spent	.000	.000	.089	1	.766	1.000
Teacher Ratio	-.008	.144	.003	1	.955	.992
Constant	-.292	2.002	.021	1	.403	.747
Goodness of Fit Test			X2	Df	Sig.	
			5.202	8	.736	

*Note.* Where ethgrp# equaled the number of race/ethnic groups a district had applied in their rating, Econ. Disadv. represents the Economically Disadvantaged indicator, \$ spent represents the funds spent per pupil, and Teacher Ratio represents the number of students for every teacher.

The nonsignificant predictors were removed and the logistic regression was repeated. The odds of being rated *Exemplary* versus *Recognized* decreased using a factor of 0.162 if the district has a disadvantaged group. The results generated the following equation:

$$\text{Predicted logit of (RATING)} = -1.099 + (-1.821) * \text{Econ. Disadv.}$$

The equation can be utilized to help predict a rating of *Exemplary* or *Recognized*.

SPSS calculated predicted probabilities for each of the districts in the study.

Table 17 illustrates the relationship between actual observations and predicted outcome for the first ten districts utilized in the study. A 1-variable predictor model was utilized to generate a predictor equation concerning the relationship between the rating of *Exemplary* and *Recognized*.

Table 17

*Predicted Probability of Being Rated Exemplary versus Recognized for 10 Districts*

Case Num.	EcDis $\beta =$ -1.424	Intercept = -1.390	Pred Prob	Actual Rating 1= Exe 0=Rec
1	1	-1.099	.05118	0
2	1	-1.099	.05118	0
3	1	-1.099	.05118	0
4	1	-1.099	.05118	0
5	1	-1.099	.05118	0
6	1	-1.099	.05118	0
7	1	-1.099	.05118	0
8	1	-1.099	.05118	0
9	1	-1.099	.05118	0
10	1	-1.099	.05118	0

When exploring the ratings of *Exemplary* and *Recognized*, a relationship was found in the category of the Economically Disadvantaged group and rating category. Economically Disadvantaged was related (Wald (1) =9.538,  $p = .002$ ). The odds of

being rated *Exemplary* versus being rated *Recognized* is decreased by using a factor of .162 if the district has an Economically Disadvantaged group as compared to not having an Economically Disadvantaged group as illustrated in Table 18. Nagelkerke  $R^2$  was computed in an effort to account for variance. When comparing *Exemplary* to *Recognized*, Nagelkerke  $R^2 = .072$ , indicating that 7.2% of the variance in a rating of *Exemplary* versus *Recognized* could be attributed to the variables selected.

Table 18

*Logistic Regression Analysis of Rating Categories Recognized and Exemplary with Significant Variables*

Predictor	B	S.E.	Wald	df	Sig.	Exp(B)
Econ. Disadv.	-1.821	.590	9.538	1	.002	.162
Constant	-1.099	.516	4.526	1	.033	.333

*Note.* Where Econ. Disadv. represents the Economically Disadvantaged indicator.

A descriptive analysis using crosstabs revealed that only six districts in the State of Texas were rated *Exemplary* and had two or more race/ethnic groups. See the following Table 19.

Table 19

*Cross tabulation of Race/Ethnicity Indicators Utilized in Calculating Rating and Rating Assigned to District: Exemplary and Recognized*

Number of Race/Ethnic Groups	Rating		Total
	Exemplary	Recognized	
1 race/ethnic group	12	99	111
2 race/ethnic groups	3	109	112
3 race/ethnic groups	3	48	51
Total	18	256	274

Overall prediction revealed that on the basis of 1 variable, the predicted ratings were correct for 0% of *Exemplary* districts and 100% of *Recognized*. Overall the model was correct 93.4% of the time. The model over classified *Exemplary* districts as *Recognized* as illustrated by Table 20.

Table 20

*Accuracy of Predicted Ratings According to Predicted Logit Equation: Exemplary to Recognized*

Observed	Predicted		Percentage Correct
	Exemplary	Recognized	
Exemplary	0	18	0.0%
Recognized	0	256	100%
Overall Percentage			93.4%

#### Research Question 4

Is there a relationship between district accountability ratings and district Mathematics college and career readiness rates?

The correlation between College Career Readiness rates in Mathematics were determined utilizing data provided by TEA regarding eleventh grade students performance on the 11th grade TAKS test. Ratings were grouped into two categories: *Academically Unacceptable/Academically Acceptable* and *Recognized/Exemplary*. A point biserial correlation was run to determine the correlation between the rating and population of students considered to be College Ready. There was a significant positive correlation between college readiness Mathematics rates and ratings,  $r_{pb} = .29, p < .01$  as shown in Table 21. Districts that are rated *Exemplary* or *Recognized* tend to have more students prepared for college in the area of Mathematics.

Table 21

*Correlation Between Rating and College Career Readiness Mathematics Rates*

	Results	CCR M	Rating
CCR M	Correlation	1	.291**
	Sig. (2-tailed)		.000
	<i>N</i>	967	967

*Note.* \*\*Correlation is significant at  $p < .01$  (2-tailed).

### Research Question 5

Is there a relationship between district accountability ratings and district English Language Arts College and Career Readiness rates?

The correlation between College Career Readiness reading rates were correlated with the 2008 district ratings. Ratings were grouped into two categories: *Academically Unacceptable/Academically Acceptable* and *Recognized/ Exemplary*. A point biserial correlation was run to determine the correlation between the rating and population of students considered to be college ready. There was a significant correlation between ratings and college readiness for reading,  $r_{pb} = .176$ ,  $p < .01$  as shown in Table 22.. Districts that are rated *Exemplary* or *Recognized* tend to have more students prepared for college in the area of Mathematics.

Table 22

*Correlation Between Rating and College Career Readiness Reading Rates*

	Results	Rating	CCR R
Rating	Correlation	1	.176**
	Sig. (2-tailed)		.000
	<i>N</i>	967	967

*Note.* \*\*Correlation is significant at  $p < .01$  (2-tailed).

## CHAPTER 5

### SUMMARY, FINDINGS, CONCLUSIONS, IMPLICATIONS FOR PRACTICE, AND RECOMMENDATIONS FOR FURTHER RESEARCH

Chapter V summarizes and reports all findings and conclusions. Suggestions for practice and future research are also included in this chapter.

#### Summary

Accountability in the State of Texas has become a major focus of districts. Research indicates that not all districts have the same opportunity to attain ratings (Starrett, 2008). The stigma associated with failure to achieve the desired rating can be devastating to district morale. The ability to predict a district's rating based on demographics would help diffuse the emphasis placed on rating and turn the focus to student growth and achievement. Stevens et al. (2000) suggest a multi-tier approach to accountability systems that includes measures for student growth.

In this study, three logistic regressions were performed comparing ratings: *Academically Acceptable* to *Academically Unacceptable*, *Recognized* to *Academically Acceptable*, and *Exemplary* to *Recognized*. The predictor variables for each logistic regression for the district were the number of student ethnicity/ race (African American, Hispanic, and White) utilized in a district rating as well as, the Economically Disadvantaged group indicator. Student demographic groups were evaluated based on the number of race/ethnicity groups that were counted towards a rating: Results showed that for each rating comparison, the variables studied had a different impact. For the

comparisons of *Academically Acceptable* to *Academically Unacceptable*, none of the selected variables were found to be significant and thus applicable to forming an equation to predict rating. For *Recognized* to *Academically Acceptable*, the number of race/ethnic groups and economically disadvantaged indicator were both significant. For the comparison of *Exemplary* to *Recognized*, the results indicated that having an economically disadvantaged group was related to the rating.

Two point biserial correlations were run to further evaluate the ratings and their relationship to students being prepared for college. A correlation was run evaluating the relationship between the two rating groups of *Academically Unacceptable* and *Academically Acceptable* to *Recognized* and *Exemplary* when compared to the percentage of students college ready for Mathematics. A second correlation was run evaluating the relationship between the two rating groups of *Academically Unacceptable* and *Academically Acceptable* to *Recognized* and *Exemplary* when compared to the percentage of students who are college ready for reading according to the TAKS test. Both correlations showed strong relationships between the ratings and percentage of students who are considered college ready.

### Discussion

The intent of this research was to determine if district ratings could be predicted by the following variables: number of race/ ethnicity indicators, Economically Disadvantaged student indicator, student to teacher ratio, and funds spent per student. In addition, the study explored the relationship between ratings and college readiness of

students. The study utilized data from the spring 2008 TAKS testing with the data being evaluated in the summer of 2011.

Data were utilized from several TEA-supplied public databases. Databases were compressed into one file by utilizing the county district number of each district. In 2008, Texas had 1158 districts that received a rating. Before analysis was run, 191 districts were removed from the research population because of incomplete data being available; 967 of those districts were utilized in this study.

Results of the study indicate that the number of race/ ethnicity group indicators and economically disadvantaged student group indicator can be utilized to predict accountability ratings when comparing *Recognized to Academically Acceptable*. In addition, the economically disadvantage indicator can be utilized to predict ratings when comparing *Exemplary to Recognized*. For each comparison the specified variables, if counted toward the district rating, decreases the likelihood of attaining the higher rating. These results reflect the research discussed in the literature review. According to Rouse & Barrow, (2006), Hursh, (2005), Hanushek, (2004), Kane & Staiger (2002), Yeop Kim & et al. (2006), various groups underperform on assessments and performance gaps exist. The current accountability system penalizes districts for this fact. Even if districts close gaps between student groups, the rating is impacted due to the underperforming. Test bias, academic language, and test anxiety could all impact student performance.

The results of this study reflect the conclusion of Wößsmann and West (2002). Class size did not prove to impact the accountability rating. While neither class size nor

amount spent per student showed a significant relationship in regard to predicting ratings, there has to be some minimum level in which a large impact occurs. In Texas, the school funding is equitable enough amongst the various rating group that ratings are not able to be predicted based on the differences in funding or teacher-to-student ratio.

Accountability serves to hold educators accountable for teaching every student. The problems with the current system in Texas lie within the usage of the system. While serving to guarantee that no students are overlooked based on race or economic factors, schools with diversity and poor students are penalized by the fact that students count multiple times in a rating. Pair this with the fact that students from poverty-level households come to school less prepared than their non poverty peers, and schools teaching low-socioeconomic clientele face a tremendous challenge. Accountability should measure student learning and growth. Each student should be treated individually with goals set each year based on where the student starts the year and where the student should finish.

Accountability should encompass more than a one-day assessment. Attempting to sum up a year's worth of learning based on a 50-question multiple choice test does not seem adequate or fair to students. "Differentiation" has been a buzz word in education. Differentiation basically means that teachers teach each student based on the student's individual needs. One student might work on a poem while another writes a story. Both assignments are created to measure the same learning goals. An accountability system should be created that allows for differentiation while measuring student learning.

## Findings

When comparing *Academically Acceptable* to *Academically Unacceptable*, none of the variables were found to be significant. For this study, 681 districts were rated *Academically Acceptable* while 12 were rated *Academically Unacceptable*. A prediction equation was not developed indicating that the number of race/ ethnicity groups, the presence of an economically disadvantaged student group, student to teacher ratio and the amount spent per student cannot predict a district's rating according to this study.

In the comparison of *Recognized* to *Academically Acceptable*, the number of race/ethnicity groups and Economically Disadvantage student group indicator were both related to the ratings at  $p < .05$ . Both indicators impacted the odds of attaining a *Recognized* rating when compared to an *Academically Acceptable* rating. With one unit increase in the number of race/ ethnic groups , the odds of being rated *Recognized* compared to *Academically Acceptable* was decreased using a factor of 0.479. In addition, the odds of being rated *Recognized* versus *Academically Acceptable* decreased using a factor of 0.271 if the district has a disadvantaged group. The findings directly reflect what the research indicates regarding students from homes of poverty underperforming those students who are not from poverty (Rouse & Barrow, 2006, Hursh, 2005, Hanushek, 2004). The findings also reflect what research states regarding performance of students and diversity. The more race/ ethnic groups a district has, the less likely the district is to attain a *Recognized* rating. As Hanushek (2004), Kane &

Staiger (2002), Yeop Kim et al. (2006), discuss conclusions drawn from ratings might reflect more than what is being taught in the classroom.

The exploration of *Exemplary to Recognized* ratings yielded that the Economically Disadvantaged student indicator was related to the rating at  $p < .05$ . The odds of being rated *Exemplary* versus being rated *Recognized* is decreased by using a factor of .162 if the district has an Economically Disadvantaged group as compared to not having an Economically Disadvantaged group. Districts of wealthy students have an advantage over those that serve low income students when attempting to be rated *Exemplary*. According to Rouse & Barrow, 2006, Hursh, 2005, and Hanushek, 2004, these results reflect what would be expected. Flinders (2005) discusses the concern that the wealthy schools will get wealthier based on ratings while Nelson et al. (2007) discusses the negative impact of low ratings on teacher recruitment. Both are valid concerns based on the results of this study. It is more difficult to attain a higher rating when working with students of poverty and diversity. The poverty and diversity do not cause the rating, but both lead to more obstacles. Test bias and lacking academic language should both be explored to determine their possible impact on ratings.

In addition, the correlation between College Readiness and ratings demonstrated a positive relationship that rating category increased as college readiness percentages increased in both areas of Reading and Mathematics. The larger the percentage of students college ready in the district; the higher the rating. However, the correlation did not account for a large portion of variance. Correlation does not imply causation; the percentage of student college ready must be studied further to determine causation.

Cross tabulations were performed for the comparison of *Academically Acceptable* to *Academically Unacceptable*, *Recognized* to *Academically Acceptable* and *Exemplary* to *Academically Acceptable*. Results, displayed in Table 5, indicated that 5 out of 12 districts rated *Academically Unacceptable*, had all three race/ethnicity groups (African American, Hispanic, and White) count individually towards the rating. Three districts rated *Academically Unacceptable* had only the Hispanic indicator applied for the race/ ethnicity possibility of combinations.

### Conclusions

Research indicates that ratings are a reflection of student populations as well as performance (Humiston, 2007; Kane & Staiger, 2003, Starrett, 2008). With so much attention focusing on ratings, it is essential that the rating actually reflect growth and teaching versus student population as otherwise the emphasis placed is counter intuitive to the goal of public education. Districts have begun focusing on ratings as the prime display of superior teaching (Popham, 2001), but this study can suggest that the rating attained by each district is a reflection of much more than academic excellence. The rating can be a reflection of district demographics and funding (Kane & Staiger, 2003; Turley, 2007). The accountability ranking seeks to be a one-word evaluation of the quality of a school, but it does not address the inequality of the status quo. Schools have disparate levels of funding and some population groups come to school undereducated, struggling with language acquisition, and/or poverty.

As discussed earlier, morale of schools is impacted by the rating system utilized. Based on the study conducted, the ability to attain a higher rating might be

predetermined by district demographics. Teachers at diverse campuses face greater challenges when one recognizes the fact that the system penalizes campuses for being diverse. Recruiting teachers to high-need schools can also become difficult because of the stigmatization associated with working at a low-performing school.

This study does not establish that diversity causes certain ratings, but to evaluate if ratings could be predicted. As seen in Table 5, it is important to note that 42% of districts that were rated *Academically Unacceptable* had all possible race and ethnicity groups count separately in the rating. That being said, 308 districts were rated *Academically Acceptable, Recognized* or *Exemplary* while having all race and ethnicity indicators count individually in the rating. It is important to be cautious in drawing conclusions about these data due to the restrictions. The small number of *Academically Unacceptable* districts prevented establishing a relationship between the variables and the rating. Further research should be conducted to determine if having all race and ethnicity indicators applied towards a rating can predict being *Academically Unacceptable*.

Too much emphasis is placed on the accountability rating of a district in the State of Texas. As shown in this study, it is not always an equitable measure- allowing all districts equal access to each rating despite various factors. Districts should be rated with an instrument that is equitable for all districts and accounts for the variance in demographics, funding, and staffing.

In addition, the ratings are strongly correlated with college readiness as measured by performance on the Exit level TAKS test. It is important to note that

correlation does not imply causation. With recent reports suggesting more and more students each year graduate high schools under prepared for college, the strong correlation implies that the districts that attain higher ratings are preparing students better for college and the other districts are not. However; when pairing the two studies together, it seems that possibly the students who are prepared best for college at higher rated districts come from wealthier and less diverse districts. This then begs the question: Are those high-performing districts excelling in teaching?

#### Implications for Practice

Achieving a particular rating is the goal for many districts. While this is an admirable goal, this study indicates that districts should be leery in assuming that ratings reflect teaching effectiveness. In some instances, the demographics of the district impact the rating as well. This study found that demographics can produce different odds of attaining some rating categories in the following circumstances: when comparing *Academically Acceptable* and *Recognized* and when comparing *Recognized* to *Exemplary*. Districts should continue to push to achieve recognition by the state, but in addition, the districts should seek alternate means of evaluating effective practices. Teacher and curriculum evaluation should look at growth made in combination with the percentage of students meeting standards. Should students make large gains each year, but fail to pass the state assessment, the growth should be considered as a positive step and indication of effectiveness. Rating a district and school based on one test is not the best method, nor should it be the total representation of a healthy program. Texas

should look at evaluating districts using alternative measures as well as standardized assessments.

All stake holders should be educated on the accountability system in Texas. Residents should be made aware that student groups impact accountability ratings and that the district ratings could be reflective of diversity in the community more so than academic achievement.

The findings of this study should be used to continue to educate all stakeholders.

Findings could be utilized to help create a new accountability system for the State of Texas that takes diversity into account as well as growth measures. It is hoped that districts will understand their rating in relationship to the entire state and recognize the effect diversity has on ratings; especially, when districts are looking to achieve a jump in rating. The specific situations addressed in this study can point out areas where making that rating change will be difficult based on the demographics.

Districts and campuses can utilize these results to improve the morale of campuses when the goal rating is not achieved. Creating an excuse is not the purpose of the study, but it is the intent to help prevent large-scale public criticism and stigmatization. Lower rated districts should strive to improve while remembering that the measure of effectiveness is not solely the state rating.

#### Recommendations for Further Study

Based on the results of this study, further research should be conducted. Future research on this topic should include ratings from another year prior to 2008 to see if the same conclusions are arrived upon. Research should also be conducted for future

accountability models. The State of Texas is currently working to create a new accountability system based on State of Texas Assessments of Academic Readiness (STAAR), the state mandated assessment that will replace TAKS. The study should then be replicated utilizing the new accountability system. Additional research should be conducted to evaluate ratings and growth measures to see if ratings correlate with student academic growth. Research also indicates a need for further evaluation of student college readiness and preparedness of students exiting high school. Correlating student success at the collegiate level with performance on TAKS/STAAR should be conducted. Further research should be conducted to utilizing districts rated *Academically Unacceptable* to determine which variables can predict the risk of attaining this rating most accurately. In addition, further studied should determine if test bias exists and the impact the bias has on accountability ratings.

APPENDIX A  
TABLES A1 AND A2

Table A1

*Requirements for Each Rating Category of Districts in Texas*

	<b>Academically Acceptable</b>	<b>Recognized</b>	<b>Exemplary</b>
<b>Base Indicators</b>			
<b>TAKS (2007-08)*</b> • All students <i>and each student group meeting minimum size:</i> • African American • Hispanic • White • Econ. Disadv. * <i>TAKS (Accommodated) included for some grades and subjects. See Table 3.</i>	Meets each standard: • Reading/ELA ... <b>70%</b> • Writing ..... <b>65%</b> • Social Studies.. <b>65%</b> • Mathematics .... <b>50%</b> • Science ..... <b>45%</b> <b>OR</b> Meets Required Improvement	Meets 75% standard for each subject <b>OR</b> Meets 70% floor and Required Improvement	Meets <b>90%</b> standard for each subject
<b>Completion Rate I (Class of 2007)</b> • All students <i>and each student group meeting minimum size:</i> • African American • Hispanic • White • Econ. Disadv.	Meets <b>75.0%</b> standard <b>or</b> Meets Required Improvement	Meets <b>85.0%</b> standard <b>or</b> Meets floor of 75.0% and Required Improvement	Meets <b>95.0%</b> standard
<b>Annual Dropout Rate (2006-07)</b> • All students <i>and each student group meeting minimum size:</i> • African American • Hispanic • White • Econ. Disadv.	Meets <b>2.0%</b> standard <b>or</b> Meets Required Improvement	Meets <b>2.0%</b> standard <b>or</b> Meets Required Improvement	Meets <b>2.0%</b> standard <b>or</b> Meets Required Improvement
<b>Additional Provisions</b>			
<b>Exceptions</b>	Applied if district/campus would be <i>AU</i> due to not meeting <i>AA</i> criteria. (See <i>detailed explanation.</i> )	Applied if district/campus would be <i>AA</i> due to not meeting <i>Recognized</i> criteria. (See <i>detailed explanation.</i> )	Applied if district/campus would be <i>Recognized</i> due to not meeting <i>Exemplary</i> criteria. (See <i>detailed explanation.</i> )
<b>Check for Academically Unacceptable Campuses (District only)</b>	Does not apply to <i>Academically Acceptable</i> districts.	A district with a campus rated <i>Academically Unacceptable</i> cannot be rated <i>Recognized</i> .	A district with a campus rated <i>Academically Unacceptable</i> cannot be rated <i>Exemplary</i> .
<b>Check for Underreported Students (District only)</b>	Does not apply to <i>Academically Acceptable</i> districts.	A district that underreports more than <b>200</b> students or more than <b>5.0%</b> of its prior year students cannot be rated <i>Recognized</i> .	A district that underreports more than <b>200</b> students or more than <b>5.0%</b> of its prior year students cannot be rated <i>Exemplary</i> .
<b>School Leaver Provision for 2008</b>	A campus or district annual dropout rate, completion rate, and/or underreported student measures cannot be the cause for a lowered rating.		

<sup>1</sup>. From 2008 Accountability Manual (p. 42), by Texas Education Agency, 2008, Austin, TX: Texas Education Agency. Copyright 2008 by Texas Education Agency. Reprinted with permission.

Table A2

Texas Overview of 2008 System Components

	TAKS TAKS (Accommodated)	Completion Rate I	Dropout Rate
<b>Definition</b>	Results (gr. 3-11) summed across grades by subject. ELA & reading results are combined. Cumulative results used for first two administrations of gr. 3, 5, & 8 reading; gr. 5 & 8 math. Included in TAKS in the following subjects and grades: ELA (gr. 11) Mathematics (gr. 11) Science (gr. 5, 8, 10, and 11) Science (gr. 5 Spanish) Social Studies (gr. 8, 10, & 11)	Grads & continuers expressed as a % of total students in the class. Campuses serving any of gr. 9-12 w/out a completion rate are not evaluated.	Gr. 7 and 8 dropouts as a % of students who were in attendance any time during the prior school year.
<b>Rounding</b>	Whole Numbers	One Decimal	
<b>Standards</b>	<i>Exemplary</i> :..... All Subjects..... ≥ 90% <i>Recognized</i> :..... All Subjects..... ≥ 75% <i>Acceptable</i> :..... Reading/ELA..... ≥ 70% Writ./Soc St..... ≥ 65% Mathematics ..... ≥ 50% Science..... ≥ 45%	EX: ≥ 95.0% RE: ≥ 85.0% AA: ≥ 75.0%	EX: ≤ 2.0% RE: ≤ 2.0% AA: ≤ 2.0%
<b>Mobility Adjustment (Accountability Subset)</b>	District ratings: results for students enrolled in the district in the fall and tested in the same district. Campus ratings: results for students enrolled in the campus in the fall and tested in the same campus.	None	
<b>Subjects</b>	Reading/ELA .....gr. 3-11 Writing.....gr. 4, 7 Mathematics .....gr. 3-11 Social Studies.....gr. 8, 10, 11 Science .....gr. 5, 8, 10, 11 ELA .....gr. 11 Writing .....N/A Mathematics.....gr. 11 Social Studies ..... gr. 8, 10, 11 Science ..... gr. 5, 8, 10, 11	N/A	
<b>Student Groups</b>	All & Student Grps: African American Hispanic White Econ. Disadv.	All & Student Grps: African American Hispanic White Econ. Disadv.	
<b>Minimum Size Criteria for All Students</b>	No minimum size requirement—special analysis for small numbers	≥ 5 dropouts AND ≥ 10 students	
<b>Minimum Size Criteria for Groups</b>	30/10%/50	≥ 5 dropouts AND 30/10%/50	
<b>Required Improvement (RI)</b>			
<b>Actual Chg</b>	2008 minus 2007 performance	Class of 2007 rate minus Class of 2006 rate	2006-07 rate minus 2005-06 rate
<b>RI</b>	Gain needed to reach standard in 2 years		
<b>Use</b>	As a gate up to <i>Academically Acceptable</i> or <i>Recognized</i>		
<b>Floor</b>	≥ 70% for <i>Recognized</i> , no floor for <i>Academically Acceptable</i>	≥ 75.0% for <i>Recognized</i>	No floor
<b>Minimum Size</b>	Meets minimum size in current year and has ≥ 10 students tested in prior year	Meets min. size current year and has ≥ 10 in prior year class.	Meets min. size current year and has ≥ 10 7 <sup>th</sup> – 8 <sup>th</sup> grade students the prior year.
<b>Exceptions</b>	Applies to TAKS measures only		
<b>Use</b>	As a gate to <i>Acceptable</i> , <i>Recognized</i> , or <i>Exemplary</i>		
<b>Floor</b>	<i>Academically Acceptable</i>	<i>Recognized</i>	<i>Exemplary</i>
R/W/SS	5 pts.	5 pts.	5 pts.
M/Sc	10 pts.	5 pts.	5 pts.
<b>Number of Exceptions Allowed</b>	1 – 4 measures evaluated ..... 0 allowed 5 – 8 measures evaluated ..... 1 allowed 9 – 11 measures evaluated ..... 2 allowed 12 – 15 measures evaluated ..... 3 allowed 16+ measures evaluated ..... 4 allowed	If 10 or more measures, one exception allowed	
Exceptions are Not Applicable to Completion Rate or Dropout Rate			

Table A2 – continued

<b>School Leaver Provision for 2008</b>	<p style="text-align: center;">N/A</p>	<p style="text-align: center;">Campus/ District rating will not be lowered due to annual dropout or completion rates.</p>
---	--	---

<sup>1.</sup> From 2008 Accountability Manual (p. 43), by Texas Education Agency, 2008, Austin, TX: Texas Education Agency. Copyright 2008 by Texas Education Agency. Reprinted with permission.

APPENDIX B

PERMISSION FROM TEA TO UTILIZE TABLES IN DISSERTATION

**From:** Copyrights  
**To:** Fitzhugh, Sarah; Copyrights  
**Subject:** RE: Tables from accountability manual  
**Date:** Thursday, May 19, 2011 9:47:30 AM

Dear Ms Fitzhugh, Your request to use the tables from the Accountability Manual for your dissertation is approved. We wish you well in your upcoming project. Sincerely,  
Norma Barrera Director, Purchasing and Contracts

**From:** Fitzhugh, Sarah [mailto:fitzhughsr@lisd.net]  
**Sent:** Wednesday, May 18, 2011 3:36 PM **To:** Copyrights **Subject:** Tables from accountability manual

I am currently writing a dissertation to complete the requirements for my doctoral program. I would like to include tables from the Accountability manual in the dissertation. I am a TX resident, but was not sure if a dissertation fell under personal use per the Copyright. Can I utilize the tables in my dissertation without violating the copyright?

Thank you,

Sarah Fitzhugh

APPENDIX C  
GUIDE TO CODING

### Research Question 1

#### Rating:

Academically Acceptable=1

Academically Unacceptable= 0

#### Indicators

Economically Disadvantaged

Met minimum size criteria= 1

Did not meet minimum size criteria= 0

Combination of Race/ Ethnicity

One race/ ethnicity group=1

Two race/ ethnicity groups=2

Three race/ ethnicity groups=3

### Research Question 2

#### Rating:

Recognized=1

Academically Acceptable= 0

#### Indicators

Economically Disadvantaged

Met minimum size criteria= 1

Did not meet minimum size criteria= 0

Combination of Race/ Ethnicity

One race/ ethnicity group=1

Two race/ ethnicity groups=2

Three race/ ethnicity groups=3

### Research Question 3

#### Rating:

Exemplary=1

Recognized= 0

#### Indicators

Economically Disadvantaged

Met minimum size criteria= 1

Did not meet minimum size criteria= 0

Combination of Race/ Ethnicity

One race/ ethnicity group=1

Two race/ ethnicity groups=2

Three race/ ethnicity groups=3

Research Question 7

Rating:

Exemplary and Recognized= 1

Academically Acceptable and Academically Unacceptable= 0

Research Question 8

Rating:

Exemplary and Recognized= 1

Academically Acceptable and Academically Unacceptable= 0

APPENDIX D  
IRB APPROVAL



THE UNIVERSITY  
OF TEXAS AT  
ARLINGTON

Office of Research Administration  
Box 19188 202 E. Border St., Suite  
214 Arlington, Texas 76019-0188

T 817.272.3723 F  
817.272.1111

<http://www.uta.edu/research>

[scorntia@uta.edu](mailto:scorntia@uta.edu)

<http://www.uta.edu/icoer/ta>

BaAHvawrticlw

May 19, 2011

Sarah Fitzhugh Dr. James Hardy  
Educational Leadership and Policy

Box 19575	
Protocol Title:	ACCOUNTABILITY IN TEXAS: PREDICTING DISTRICT RATINGS BASED ON INDICATORS
RE:	Exempt Approval Letter
IRB No.:	2011-0529e

The UT Arlington Institutional Review Board (UTA IRB) Chair (or designee) has reviewed the above-referenced study and found that it qualified as exempt from coverage under the federal guidelines for the protection of human subjects as referenced at Title 45 Part 46.101(b)(4). You are therefore authorized to begin the research as of May 17, 2011.

Please be advised that as the principal investigator, you are required to report local adverse (unanticipated) events to this office within 24 hours. In addition, pursuant to Title 45 CFR 46.103(b)(4)(iii), investigators are required to, "promptly report to the IRB any proposed changes in the research activity, and to ensure that such changes in approved research, during the period for which IRB approval has already been given, are **not initiated without IRB review and approval** except when necessary to eliminate apparent immediate hazards

to the subject."

All investigators and key personnel identified in the protocol must have documented Human Subject Protection Training (HSP) or *CITI Training* on file with this office. The UT Arlington Office of Research Administration Regulatory Services appreciates your continuing commitment to the protection of human research subjects. Should you have questions or require further assistance, please contact Robin Dickey at [robind@uta.edu](mailto:robind@uta.edu) or you may contact the Office of Regulatory Services at 817-272-3723.

Sincerely,

Patricia G. Turpin, PhD, RN, NEA-BC  
Clinical Associate Professor UT Arlington  
IRB Chair

## REFERENCES

- Alspaugh, J. (1994). The relationship between school size, student teacher ratio and school efficiency. *Education*, 114(4), 593. Retrieved from Academic Search Complete database.
- Applied Research Center (n.d.). *Historical Timeline of Public Education in the US*. Retrieved from <http://www.arc.org/content/view/100/217/>
- Baker, E., Linn, R., Herman, J., W. & Koretz, D. (2002). Standards for educational accountability systems. Policy Brief 5. CRESST: National Center for Research on Evaluation, Standards and Student Testing. Retrieved from <http://www.cse.ucla.edu/products/policy.html>
- Bishop, J., & Mane, F. (2004). *Educational reform and disadvantaged students: Are they better off or worse off?* Retrieved from Program on Education Policy and Governance website: <http://www.hks.harvard.edu/pepg/research.htm#10-17>
- Bracey, G. (2008). Disastrous legacy: Aftermath of a nation at risk. *Dissent* (00123846), 55(4), 80-83. Retrieved from Academic Search Complete database.
- California Department of Education (2010). *2010-11 Academic performance index reports: Information guide*. Retrieved from <http://www.cde.ca.gov/ta/ac/ap/documents/infoguide10.pdf>
- California Department of Education. (2011). *Overview of California's 2010-11 accountability progress reporting system*, 2011. Retrieved from <http://www.cde.ca.gov/ta/ac/ay/documents/overview11.pdf>
- California Department of Education. (n.d. a) *2010 STAR Test Results*. Retrieved August 20, 2011 from <http://star.cde.ca.gov/star2010/aboutSTAR.asp>
- California Department of Education. (n.d. b) *Standardized testing and reporting-CalEdFacts*. Retrieved from <http://www.cde.ca.gov/ta/tg/sr/cefstar.asp>
- California Department of Education. (n.d. c) *Testing and accountability*. Retrieved from <http://www.cde.ca.gov/ta/>

- California Department of Education Assessment and Accountability Division. (2011). *2009-2010 Accountability progress reporting (APR)*. Retrieved from <http://api.cde.ca.gov/AcntRpt2010/2010GrthAPIdst.aspx?cYear=&allcds=4168999&cChoice=2010GDst1>
- Callan, P., Finney, J., Kirst, M., Usdan, M., & Venezia, A. (2004). *Claiming common ground: State policymaking for improving college readiness and success*. Retrieved from The Institute for Educational Leadership website: <http://eric.ed.gov/PDFS/ED491187.pdf>
- Chingos, M. (2010). *The impact of a universal class-size reduction policy: Evidence from Florida's statewide mandate*. Retrieved from Program on Education Policy and Governance website: <http://www.hks.harvard.edu/pepg/research.htm#10-17>
- Dalaker, J. (2001). *Poverty in the United States:2000*. Retrieved from U.S. Department of Commerce website: <http://www.census.gov/prod/2001pubs/p60-214.pdf>
- Darling-Hammond, L. (2007). Race, inequality and educational accountability: The irony of 'No Child Left Behind'. *Race, Ethnicity & Education*, 10(3), 245-260. doi:10.1080/13613320701503207.
- Derthick, M., & Dunn, J. (2009). False premises: The accountability fetish in education. *Harvard Journal of Law & Public Policy*, 32(3), 1015-1034. Retrieved from Academic Search Complete database.
- Dictionary.com (n.d.). Accountability. Retrieved from <http://dictionary.reference.com/browse/accountability>
- Edge, D. (2009). *Mathliteracy: The relationship of algebra, gender, ethnicity, socioeconomic status, and AVID enrollment with high school Mathcourse completion and college readiness*. (Ed.D. dissertation, University of North Texas, United States—Texas). Retrieved from Dissertations & Theses: Full Text. (Publication No. AAT 3385784)
- Flinders, D. (2005). The failings of NCLB. *Curriculum & Teaching Dialogue*, 7(1/2), 1-9. Retrieved from Academic Search Complete database.
- Foote, M. (2007). Keeping accountability systems accountable. *Phi Delta Kappan*, 88(5), 359-363. Retrieved from Academic Search Complete database.
- Geiser, Saul; & Maria Veronica Santelices. (2007). *Validity Of High-School Grades In Predicting Student Success Beyond The Freshman Year: High-School Record vs. Standardized Tests as Indicators of Four-Year College Outcomes*. UC Berkeley:

Center for Studies in Higher Education. Retrieved from:  
<http://escholarship.org/uc/item/7306z0zf>

Giambo, D. (2010). High stakes testing, high school graduation, and limited English proficient students: a case study. *American Secondary Education*, 38(2), 44-56. Retrieved from Academic Search Complete database.

Edweek. (2006). *Grade by grade state testing policies quality counts at 10: A decade of standard-based education*. Retrieved from <http://www.edweek.org/media/ew/qc/2006/17sos.h25.ggtp.pdf>

Graue, E., & Johnson, E. (2011). Reclaiming assessment through accountability that is “just right”. *Teachers College Record*, 13(8), 1827-1862.

Greene, J., & Forster, G. (2003). *Public high school graduation and college readiness rates in the United States*. Retrieved from Manhattan Institute website  
[http://www.manhattan-institute.org/html/ewp\\_03.htm](http://www.manhattan-institute.org/html/ewp_03.htm)

Guerra, G. (2009). *A study of first-time-in-college students' college readiness: Using standards from the Texas Assessment of Knowledge and Skills and the Elementary Algebra ACCUPLACER*. (Ed.D. dissertation, The University of Texas at San Antonio, United States—Texas). Retrieved from Dissertations & Theses: Full Text. (Publication No. AAT 3354809)

Hanushek, E. (2004). *Distributional outcomes of the organization of the u.s. schools: peers, school quality and achievement*. Retrieved from Program on Education Policy and Governance website: <http://www.hks.harvard.edu/pepg/research.htm#10-17>

Hess, F., & Fullerton, J. (2009). The numbers we need. *Phi Delta Kappan*, 90(9), 665-669. Retrieved from Academic Search Complete database.

Hopmann, S. (2008). No child, no school, no state left behind: Schooling in the age of accountability. *Journal of Curriculum Studies*, 40(4), 417-456. doi: 10.1080/00220270801989818

Hosmer, D., & Lemeshow, S. (2000). *Applied logistic regression: Second edition*. New York, NY: John Wiley and Sons.

Humes, K., Jones, N., & Ramirez, R. (2011). Overview of race and Hispanic origin: 2010. *2010 Census Briefs*. Retrieved from <http://www.census.gov/prod/cen2010/briefs/c2010br-02.pdf>

- Humiston, S. (2007). *Texas accountability system: Exploring the relationship between student achievement and financial management*. (Ed.D. dissertation, Tarleton State University—Stephenville, United States—Texas). Retrieved from Dissertations and Theses: Full Text. (Publication No. AAT 3259617)
- Hunt, J. (2008). A Nation at Risk and No Child Left Behind: Déjà vu for administrators? *Phi Delta Kappan*, 89(8), 580-585. Retrieved from Academic Search Complete database.
- Hursh, D. (2005). The growth of high-stakes testing in the USA: Accountability, markets and the decline in educational equality. *British Educational Research Journal*, 31(5), 605-622. doi: 10.1080/01411920500240767
- Jaekyung, L. (2010). Dual standards of school performance and funding? Empirical searches of school funding adequacy in Kentucky and Maine. *Education Economics*, 18(2), 207-228. doi:10.1080/09645290902796415
- Jennings, J., & Corcoran, S. (2009). Beware of geeks bearing formulas. *Phi Delta Kappan*, 90(9), 635-639. Retrieved from Academic Search Complete database.
- Jones, M. G., Jones, B. D., & Hargrove, T. (2003). *The unintended consequences of high stakes testing*. Lanham, MD: Rowman & Littlefield Publishing.
- Johnson, J., Rochkind, J., & DuPont, S. (2011). *Don't count us out: How an overreliance on accountability could undermine the public's confidence in schools, business, government and more*. Editor Ilse Tebbetts. Kettering Foundation and Public Agenda.
- Jorgenson, M., & Hoffman, J. (2003). *History of the No Child Left Behind Act of 2001 (NCLB): Assessment report*. Retrieved from Pearson Education website: [http://www.pearsonassessments.com/NR/rdonlyres/D8E33AAE-BED1-4743-98A1-BDF4D49D7274/0/HistoryofNCLB\\_Rev2\\_Final.pdf](http://www.pearsonassessments.com/NR/rdonlyres/D8E33AAE-BED1-4743-98A1-BDF4D49D7274/0/HistoryofNCLB_Rev2_Final.pdf)
- Kane, T., & Staiger, D. (2002). The promise and pitfalls of using imprecise school accountability measures. *Journal of Economic Perspectives*, 16(4), 91-114. Retrieved from <http://www.aeaweb.org/articles.php?doi=10.1257/089533002320950993>
- Kane, T., & Staiger, D. (2003). Unintended consequences of racial subgroup rules. In P. E. Peterson & M. R. West (Eds.), *No Child Left Behind?: The politics and practice of school accountability* (pp.152-176). Washington DC: Brookings Institution Press.

- Koretz, D. (2008a). *Measuring up. What educational testing really tells us*. Cambridge, MA: Harvard University Press.
- Koretz, D. (2008b). The pending reauthorization of NCLB: An opportunity to rethink the basic strategy. In G. L. Sunderman (Ed.),  *Holding NCLB accountable: Achieving accountability, equity, & school reform* (pp. 9-26). Thousand Oaks, CA: Corwin Press.
- Kornhaber, M. L. (2008). The pending reauthorization of NCLB: An opportunity to rethink the basic strategy. In G. L. Sunderman (Ed.),  *Holding NCLB accountable: Achieving accountability, equity, & school reform* (pp.43-56). Thousand Oaks, CA: Corwin Press.
- Lewin, T. (2005, August 17). Many going to college are not ready, report says. *The New York Times*. Retrieved from <http://www.nytimes.com/2005/08/17/education/17scores.html>
- Lewis, A. C. (2005). Sound and fury. *Phi Delta Kappan*, 86(9), 643-644. Retrieved from EBSCOhost
- Lim, S. (2010). *Ethnic and racial difference in higher education access: Effects of family and school resources*. (Ph.D. dissertation, University of California, Santa Barbara, United States—California). Retrieved from Dissertations & Theses: Full Text. (Publication No. AAT 3422477)
- Linn, R. (2008). Methodological issues in achieving school accountability. *Journal of Curriculum Studies*, 40(6), 699-711. doi: 10.1080/00220270802105729.
- Mackun, P., & Wilson, S. (2011). Population distribution and change: 2000 to 2010. *2010 Census Briefs*. Retrieved from <http://www.census.gov/prod/cen2010/briefs/c2010br-01.pdf>
- Martin, R.(2010). *Relationship between college readiness, Oklahoma State Testing Program, and EXPLORE*. (Ed.D. dissertation, Oklahoma State University, United States—Oklahoma). Retrieved from Dissertations & Theses: Full Text. (Publication No. AAT 3422278)
- McDermott, K. (2003). What causes variation in states' accountability policies? *Peabody Journal of Education* (0161956X), 78(4), 153.
- McNeil, L. (2000). Creating new inequalities: Contradictions of reform. *Phi Delta Kappan*. 81(10), 729-734.

- Mondale, S., & Patton, S. (Eds.). (2001). *School: The story of American public education*. Boston, MA: Beacon Press.
- Murdock, S. (2006). *Population Change in Texas and Central Texas: Implications for Human and Socioeconomic Resources in the 21<sup>st</sup> Century*. Changing Demographics in Central Texas: Riding the Wave or Wiping Out. Envision Central Texas.
- Murnane, R. (2007). Improving the education of children living in poverty. *Future of Children*, 17(2), 161-182. Retrieved from Academic Search Complete database.
- National Commission on Excellence in Education. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- Nelson, S., McGhee, M., Meno, L., & Slater, C. (2007). Fulfilling the promise of educational accountability. *Phi Delta Kappan*, 88(9), 702-709. Retrieved from Academic Search Complete database.
- New York State Education Department. (2010). *How No Child Left Behind (NCLB) accountability works in New York State: Determining 2010-11 status based on 2009-10 results*. Retrieved from <http://www.p12.nysed.gov/irs/accountability/>
- NYC Department of Education. (n.d.). State and federal accountability. Retrieved from <http://schools.nyc.gov/Accountability/tools/accountability/default.htm>
- No Child Left Behind Act. (2002). Retrieved from <http://www.ed.gov/policy/elsec/leg/esea02/107-110.pdf>
- Peng, C. Y.J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1), 3-14.
- Popham, W.J. (2001). *The truth about testing: An educator's call to action*. Alexandria, VA: Association for Supervision and Curriculum Development.
- Public Education Visioning Institute. (2008). *Creating a new vision for public education in Texas: A work in progress for conversation and further development*. Austin, TX: Texas Association of School Administrators.
- Richey, J. (2009). *The TAKS test and teaching of the formal research paper: Implications for student success in higher education*. (Ed.D. dissertation, Texas

- A&M University—Commerce, United States—Texas. Retrieved from Dissertations & Theses: Full Text. (Publication No. AAT 3405803)
- Rothstein, R., Jacobsen, R., & Wilder, T. (2008). *Grading education: Getting accountability right*. New York, NY: Teachers College Press.
- Rouse, C., & Barrow, L. (2006). U.S. elementary and secondary schools: Equalizing opportunity or replicating the status quo? *Future of Children*, *16*(2), 99-123. Retrieved from Academic Search Complete database.
- Salkind, N. (2004). *Statistics for people who think they hate statistics* (2nd ed.). Thousand Oaks, CA: SAGE Publications.
- Sass, E. (2010, August 21). *American educational timeline: A hypertext timeline*. Retrieved from <http://www.cloudnet.com/~edrbsass/educationhistorytimeline.html#1900>
- Schlechty, P. (2011). The threat of accountabilism. *Educational Leadership*, *69*(1), 80-81.
- Schmidt, J. (2011). *The impact of federal and state accountability policies on the breadth of education in low-income schools*. (Thesis, California State University, Sacramento, United States—California). Retrieved from <http://csus-dspace.calstate.edu/xmlui/bitstream/handle/10211.9/1381/Final%20PDF.pdf?sequence=1>
- Sergiovanni, T. (2000). *The lifeworld of leadership: Creating culture, community and personal meaning in our schools*. San Francisco, CA: Jossey Bass.
- Starrett, T. (2008). *State accountability ratings as related to district size and diversity*. (Ed.D. dissertation, University of North Texas, United States—Texas) Retrieved from Dissertations & Theses: Full Text. (Publication No. AAT 3326825)
- Stevens, J., Estrada, S., & Parkes, J. (2000). *Measurement issues in the design of state accountability systems*. Retrieved from University of New Mexico website <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.24.2172>
- Sunderman, G. L. (Ed.). (2008). *Holding NCLB accountable: Achieving accountability, equity, & school reform*. Thousand Oaks, CA: Corwin Press.
- Tabachnick, B., & Fidell, L. (2007). *Using multivariate statistics: Fifth edition* Boston, MA: Pearson.

- Texas Association of School Boards. (2008). *A guide to Texas school finance August 2008*. Austin, TX: Author.
- Texas Education Agency. (2009a). *Timeline of testing in TEXAS*. Retrieved August 10, 2009 from <http://www.ritter.tea.state.tx.us/student.assessment/resources/studies/testingtimeline.pdf>
- Texas Education Agency. (2010a). *An overview of the history of public education in Texas*. Retrieved from <http://www.tea.state.tx.us/index4.aspx?id=148>
- Texas Education Agency. (2010b). *College and career readiness standards*. Retrieved from <http://www.tea.state.tx.us/index4.aspx?id=8061>
- Texas Education Agency. (2010c). *Frequently asked questions about adequate yearly progress. (2010)*. Retrieved from <http://ritter.tea.state.tx.us/ayp/faq/faq.html>
- Texas Education Agency. (2011a). *Frequently asked questions about adequate yearly progress. (2011)*. Retrieved August 20, 2011 from <http://ritter.tea.state.tx.us/ayp/faq/faq.html>
- Texas Education Agency. (2011b). *Texas accountability system summary of ratings for 2004 through 2010*. Retrieved from <http://ritter.tea.state.tx.us/perfreport/account/summaryxys.html>
- Texas Education Agency. (n.d.) *Overview of the academic excellence indicator system*. Retrieved from <http://ritter.tea.state.tx.us/perfreport/aeis/about.aeis.html>
- Texas Education Agency Department of Assessment, Accountability and Data Quality Division of Performance Reporting. (2008). *2008 accountability manual: The 2008 accountability rating system for Texas public schools and school districts*. Retrieved from <http://ritter.tea.state.tx.us/perfreport/account/2008/manual/index.html>
- Texas Education Agency Department of Assessment, Accountability and Data Quality Division of Performance Reporting. (2009). *2009 accountability manual: The 2009 accountability rating system for Texas public schools and school districts*. Retrieved from <http://ritter.tea.state.tx.us/perfreport/account/2009/manual/index.html>
- Townsend, B. (2002). Testing while black. *Remedial & Special Education*, 23(4), 222. Retrieved from Academic Search Complete database.

- Tressler, C. (2010). *The relationship between principal ethnicity and other chosen demographics and student achievement as measured by the Texas Education Agency's accountability rating system in predominately Hispanic public high schools in Texas*. (Ph.D. dissertation, Texas A&M University, United States—Texas. Retrieved from Dissertation and Theses: Full Text. (Publication No. AAT 3416357)
- Turley, L. (2009). *Considering adequacy: Educational spending and student achievement in the Texas public school system*. (Ed.D. dissertation, Texas A&M University—Commerce, United States—Texas). Retrieved from Dissertations and Theses: Full Text. (Publication No. AAT 3361244)
- Umpstead, R. (2007). Determining adequacy: How courts are redefining responsibility for educational finance, goals, and accountability. *Brigham Young University Education & Law Journal*, 2, 281-320. Retrieved from Academic Search Complete database.
- United States Census Bureau. (2011). *State and county quick facts*. Retrieved from <http://quickfacts.census.gov/qfd/states/48000.html>
- United States Department of Education. (2000). *Digest of educational statistics*. Washington, DC: National Center for Education Statistics.
- United States Department of Education. (2006). *Elementary and Secondary Education Act of 1965*. Retrieved from <http://www.ed.gov/policy/elsec/leg/esea02/beginning.html#sec6>
- United States Department of Education. (2009). *Digest of educational statistics*. Washington, DC: National Center for Education Statistics.
- West, M., & Peterson, P. (2005). *The efficacy of choice threats within school accountability systems: Results from legislatively induced experiments*. Retrieved from Program on Education Policy and Governance website: <http://www.hks.harvard.edu/pepg/research.htm#10-17>
- Wiley, A., Wyatt, J., & Camara, W. (2010). *The development of a multidimensional college readiness index*. The College Board. New York. Retrieved from [http://professionals.collegeboard.com/profdownload/pdf/10b\\_2084\\_DevMultiDimenRR\\_WEB\\_100618.pdf](http://professionals.collegeboard.com/profdownload/pdf/10b_2084_DevMultiDimenRR_WEB_100618.pdf)
- Wößsmann, L., & West, M. (2002). *Class-size effect in school systems around the world: Evidence from between-grade variation in TIMSS*. Retrieved from Program on Education Policy and Governance website: <http://www.hks.harvard.edu/pepg/research.htm#10-17>

Yeow Meng, T. (2009). Accounting before accountability. *Estudios Sobre Educacion*, (16), 185-194. Retrieved from Academic Search Complete database.

Yeop Kim, D., Zabel, J., Stiefel, L., & Schwartz, A. (2006). School efficiency and student subgroups: Is a good school good for everyone? *Peabody Journal of Education (0161956X)*, 81(4), 95-117. doi: 10.1207/s15327930pje8104\_5

Zyskowski, G., & O'Malley, K. (2009). *Vertical scales training slides: Going vertical: Up, up, and away*. Retrieved from [http://www.tea.state.tx.us/index3.aspx?id=3818&menu\\_id3=793](http://www.tea.state.tx.us/index3.aspx?id=3818&menu_id3=793)

## BIOGRAPHICAL INFORMATION

Sarah Fitzhugh is a lifelong learner. Earning her B.S. in Applied Learning and Development from the University of Texas at Austin and her M. ed from the University of North Texas, she continues to enjoy engaging in academic conversation.