

**THE IMPACT OF STATE-MANDATED STANDARD-BASED
HIGH-STAKES TESTING ON SELECTED TEXAS PUBLIC SECONDARY
SCHOOLS AS PERCEIVED BY SELECT ADMINISTRATORS
IN THE MEMBERSHIP OF THE TEXAS ASSOCIATION OF
SECONDARY SCHOOL PRINCIPALS**

A Record of Study

by

DAVIS M. DENNY III

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

December 2008

Major Subject: Educational Administration

**THE IMPACT OF STATE-MANDATED STANDARD-BASED
HIGH-STAKES TESTING ON SELECTED TEXAS PUBLIC SECONDARY
SCHOOLS AS PERCEIVED BY SELECT ADMINISTRATORS
IN THE MEMBERSHIP OF THE TEXAS ASSOCIATION OF
SECONDARY SCHOOL PRINCIPALS**

A Record of Study

by

DAVIS M. DENNY III

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF EDUCATION

Approved by:

Chair of Committee,	John R. Hoyle
Committee Members,	Virginia S. Collier
	Mario S. Torres
	Ronald D. Zellner
Head of Department,	Jim Scheurich

December 2008

Major Subject: Educational Administration

ABSTRACT

The Impact of State-mandated Standard-based High-stakes Testing on
Selected Texas Public Secondary Schools as Perceived by Select
Administrators in the Membership of the Texas Association of Secondary
School Principals. (December 2008)

Davis M. Denny III, B.S., Texas A&M University;

M.Ed., East Texas State University

Chair of Advisory Committee: Dr. John R. Hoyle

The purpose of this study was to assess the impact of the high-stakes standardized test movement in Texas secondary schools. The method to accomplish this task was to compare the perceptions between Texas secondary school administrators and supporters, critics, and researchers of high-stakes testing. Out of 400 potential respondents randomly selected from 2005-2006 membership list of Texas Association of Secondary School Principals, 178 administrators participated in an electronic survey to rate the extent to which 31 statements derived from supporters, critics, and the unintended consequences of high-stakes testing as reported by researchers in current literature.

Means, standard deviations, and frequencies were used to make assumptions about perceptions of secondary administrators. Independent t-tests were conducted to test for possible perception differences between groups identified in the study. Independent groups examined in this study included: Gender (Male and Female), Years of Administrative Experience (1-4 years vs. 15 or More Years), Campus Classification

(Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable). Using an alpha level of .05 to establish significance, t-tests suggest that significant differences exist between large and small school administrators on statements 5 and 7. Further, significant differences exist between male and female administrators on statements 4 and 5.

The findings of this study seem to suggest that Texas secondary principals strongly support the following statements:

1. No high-stakes decision such as grade retention or graduation should be based on the results of a single test.
2. Educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success.
3. High-stakes tests have helped focus public attention on schools with low-achieving students.
4. The public display of high-stakes test scores motivates administrators.
5. High-stakes testing has resulted in a loss of local control of curricula.
6. The implementation of high-stakes testing has been a catalyst for increased attention to students with special needs.
7. Doing poorly on high-stakes tests does not lead to increased student effort to learn.

DEDICATION

This study is dedicated to the memory of my grandmother, Juanita Chase, and my wife, Melanie Denny.

My grandmother's eternal optimism and constant support have been instrumental to the completion of this study and to my outlook on life. Her grace under duress has been inspirational to anyone that had the privilege to know her. Without her example, I could not have completed this endeavor.

My wife has been a constant source of encouragement and motivation. It was only through her extreme patience, constant encouragement, personal sacrifice, and loving understanding that this research study was finally completed.

ACKNOWLEDGMENTS

I would like to thank my committee chair, Dr. John Hoyle, and my committee members, Dr. Collier, Dr. Torres, and Dr. Zellner, for their patience, guidance, flexibility, and support throughout the course of this research.

Thanks also go to my friends and cohort members and the Educational Administration department faculty and staff for making my time at Texas A&M University a great experience. It is impossible to thank each one by name, but three friends, Dexter Upshaw, Jerome Hurt, and Marc Levesque, were especially supportive during this endeavor.

I also want to extend my gratitude to the Texas Association of Secondary School Principals, which provided their membership database vital to the completion of my study, and to all the Texas Secondary School Principals who were willing to participate in the study.

NOMENCLATURE

ESEA	Elementary and Secondary Education Act
GED	General Education Development
IQ	Intelligence Quotient
NCLB	No Child Left Behind
SAT	Scholastic Aptitude Test
SPSS	Statistical Package for Social Sciences
TAAS	Texas Assessment of Academic Skills
TAKS	Texas Assessment of Knowledge and Skills
TASSP	Texas Association of Secondary School Principals
TEKS	Texas Essential Knowledge & Skills
TEA	Texas Education Agency

TABLE OF CONTENTS

	Page
ABSTRACT	iii
DEDICATION	v
ACKNOWLEDGMENTS.....	vi
NOMENCLATURE.....	vii
TABLE OF CONTENTS	viii
LIST OF TABLES	x
 CHAPTER	
I INTRODUCTION	1
Statement of the Problem	4
Purpose of the Study	5
Research Questions	6
Operational Definitions	6
Assumptions	9
Limitations	9
Methodology	10
Significance Statement.....	13
II REVIEW OF THE LITERATURE	14
Introduction	14
Educational Evaluation	16
Standardized Testing	24
Supporters' Positions on High-stakes Testing	27
Critics' Positions on High-stakes Testing	34
Unintended Consequences of High-stakes Testing	42
Overview of <i>A Nation at Risk</i>	50
Impact of <i>A Nation at Risk</i>	56
<i>A Nation at Risk's</i> Effect on Texas	59
No Child Left Behind Act	62

CHAPTER	Page
III	METHODOLOGY66
	Population.....66
	Design of the Study70
	Instrumentation.....72
	Procedures73
	Data Analysis73
	Research Questions74
IV	RESEARCH FINDINGS76
	Research Question 180
	Research Question 2.....116
	Research Question 3.....140
	Summary of Findings.....176
V	SUMMARY AND CONCLUSIONS185
	Summary186
	Implications and Discussions190
	Conclusions195
	Recommendations201
	REFERENCES209
	APPENDIX A220
	APPENDIX B235
	VITA239

LIST OF TABLES

TABLE	Page
3.1	Summary of the First Round, Second Round, Third Round and Fourth Round Survey Response Rates.....67
3.2	Frequency Distribution of Gender as Reported by Survey Respondents.....68
3.3	Population Strata Based upon Respondents' Years of Administrative Experience68
3.4	Population Strata Based upon Respondents' Campus Classification.....69
3.5	Population Strata Based upon Respondents' Current Campus Rating.....70
4.1	Means and Standard Deviations for Statements 1 – 12.....83
4.2	Frequency Distribution of Responses to Statement Number 1: High-stakes Tests Have Helped Focus Public Attention on Schools with Low-achieving Students and, as a Result, Have Made These Students More Visible and Less Likely to Slip Between the Cracks and Fall Further Behind84
4.3	Frequency Distribution of Responses to Statement Number 2: High-stakes Tests Are Designed and Implemented to Improve Instruction by Helping Teachers Focus on What Is Most Important to Teach.....84
4.4	Frequency Distribution of Responses to Statement Number 3: High-stakes Tests Have Helped Close the Gap in Achievement Between Minority Students and Majority Students in Texas.....85
4.5	Frequency Distribution of Responses to Statement Number 4: Teachers Need to Be Held Accountable Through High-stakes Tests to Motivate Them to Teach Better, Particularly to Push the Least Motivated Ones to Perform.....85
4.6	Frequency Distribution of Responses to Statement Number 5: Doing Poorly on High-stakes Tests Will Lead to Increased Student Effort to Learn.....86
4.7	Frequency Distribution of Responses to Statement Number 6: Students Work Harder and Learn More Because They Know What Is Expected and That the High-stakes Tests Really Count86

TABLE	Page
4.8	Frequency Distribution of Responses to Statement Number 7: The Public Display of High-stakes Test Scores Motivates Administrators to Ensure That Standards (TEKS) on Which the Tests Are Based Are Part of the Curriculum and Are Being Successfully Taught.....87
4.9	Frequency Distribution of Responses to Statement Number 8: When High-stakes Tests Are Developed and Used Appropriately, They Are Among the Most Sound and Objective Knowledge and Performance Measures Available87
4.10	Frequency Distribution of Responses to Statement Number 9: Administrators Need to Be Held Accountable Through High-stakes Tests to Motivate Them to Be More Effective in Supervising Their Staffs.....88
4.11	Frequency Distribution of Responses to Statement Number 10: Increasingly, from the Classroom to the School Board Room, Educators Are Making Use of Student Performance Data Generated by High-stakes Tests to Help Them Refine Programs, Channel Funding, and Identify Roots of Success.....89
4.12	Frequency Distribution of Responses to Statement Number 11: The Implementation of High-stakes Testing Has Been a Catalyst for Increased Attention to Students with Special Needs.....89
4.13	Frequency Distribution of Responses to Statement Number 12: Driven by the Demands of High-stakes Tests, Professional Development Has Improved by Focusing on Helping Educators Hone His or Her Teaching Skills and Content Area Expertise90
4.14	Means and Standard Deviations by Campus Classification for Statements 1 – 1293
4.15	Results of t-tests by Campus Classification for Statements 1 – 12.....96
4.16	Means and Standard Deviations by Gender for Statements 1 – 12.....99
4.17	Results of t-tests by Gender for Statements 1 – 12102
4.18	Means and Standard Deviations by Years of Administrative Experience for Statements 1 – 12.....106

TABLE	Page
4.19 Results of t-tests by Years of Administrative Experience for Statements 1 – 12.....	109
4.20 Means and Standard Deviations by Current Campus Rating for Statements 1 – 12	112
4.21 Results of t-tests by Current Campus Rating for Statements 1 – 12	115
4.22 Means and Standard Deviations for Statements 13 – 19.....	118
4.23 Frequency Distribution of Responses to Statement Number 13: One Result of High-stakes Testing Is That Educators Know More About Testing Than Ever Before	119
4.24 Frequency Distribution of Responses to Statement Number 14: Prominent and Public Interest in Pupil Performance on High-stakes Tests Has Resulted in an Intensity of Effort Directed Toward Data Collection and Quality Control That Is Unparalleled	119
4.25 Frequency Distribution of Responses to Statement Number 15: High-stakes Tests Promote Greater Homogeneity of Education. A Result of Schools' Aligning Their Curricula and Instructional Focus More Closely to Outcomes Embodied in High-stakes Tests, the Experiences of and Aspirations for Children in Urban, Suburban, and Rural Districts Within a State Are More Comparable Than They Have Been in the Recent Past	120
4.26 Frequency Distribution of Responses to Statement Number 16: A Profoundly Positive Effect That the Introduction of High-stakes Consequences Has Had Lies in the Tests Themselves. High-stakes Tests Have Evolved to a State of Being: Highly Reliable; Free from Bias; Relevant and Age Appropriate; Higher Order; Tightly Related to Important Public Goals; Time and Cost Efficient; and Yielding Remarkably Consistent Decisions.....	120
4.27 Frequency Distribution of Responses to Statement Number 17: High-stakes Tests Have Exposed Educators to High-Quality Writing Prompts, Document-Based Questions, Constructed-Response Formats, and Even Challenging Multiple-Choice Items. This Has Led to Teachers Enhancing Their Own Assessment Practices	121

TABLE	Page	
4.28	Frequency Distribution of Responses to Statement Number 18: High-stakes Testing Programs Also Result in Massive Amounts of Test Preparation, Resulting in a Loss of Instructional Time.....	122
4.29	Frequency Distribution of Responses to Statement Number 19: High-stakes Testing Has Resulted in a Loss of Local Control of What Is Taught, How It Is Taught, and Who Gets High-Quality Instruction. These Decisions Are Now Greatly Impacted by Policy Makers at the State and National Levels.....	122
4.30	Means and Standard Deviations by Campus Classification for Statements 13 – 19	124
4.31	Results of t-tests by Campus Classification for Statements 13 – 19	126
4.32	Means and Standard Deviations by Gender for Statements 13 – 19	129
4.33	Results of t-tests by Gender for Statements 13 – 19	131
4.34	Means and Standard Deviations by Years of Administrative Experience for Statements 13 – 19.....	133
4.35	Results of t-tests by Years of Administrative Experience for Statements 13 – 19	135
4.36	Means and Standard Deviations by Current Campus Rating for Statements 13 – 19	138
4.37	Results of t-tests by Current Campus Rating for Statements 13 – 19	140
4.38	Means and Standard Deviations for Statements 20 – 31	143
4.39	Frequency Distribution of Responses to Statement Number 20: A Test That Has Been Validated Only for Diagnosing Strengths and Weaknesses of Individual Students Should Not Be Used to Evaluate the Educational Quality of a School or School District	144
4.40	Frequency Distribution of Responses to Statement Number 21: High-stakes Testing Compromises Educational Quality by Leading Educators to “Teach to the Test,” Which Results in a Narrowing of the Curriculum, Limiting the Scope of Tested Subjects and Shortchanging or Eliminating Subjects Not Included in the Assessments.....	144

TABLE	Page
4.41	Frequency Distribution of Responses to Statement Number 22: High-stakes Tests Are Too Expensive and Result in Diverting Scarce Resources and Attention from Serious Problems.....145
4.42	Frequency Distribution of Responses to Statement Number 23: A Focus on Standards and Accountability That Ignores the Processes of Teaching and Learning in Classrooms Will Not Provide the Direction That Teachers Need in Their Quest to Improve Instruction.....145
4.43	Frequency Distribution of Responses to Statement Number 24: Pressure Exerted from the Need to Succeed on High-stakes Tests Often Leads to Inappropriate Test Preparation Practices, Including Outright Cheating146
4.44	Frequency Distribution of Responses to Statement Number 25: High-stakes Tests Draw an Inaccurate Picture of Student Achievement and Unfairly Jeopardize Students or Schools That Are Making Genuine Efforts to Improve146
4.45	Frequency Distribution of Responses to Statement Number 26: Educational Decisions Based on High-stakes Tests Have a Disproportionate Impact on Poor and Minority Children147
4.46	Frequency Distribution of Responses to Statement Number 27: High-stakes Testing and the Accompanying Consequences of Failure Lead to Overstressed Students147
4.47	Frequency Distribution of Responses to Statement Number 28: The Pressures Inherent in Preparing Students for High-stakes Tests Are Driving out Good Teachers148
4.48	Frequency Distribution of Responses to Statement Number 29: High-stakes Tests Unfairly and Inaccurately Assess and Penalize Learners for Whom English Is Not Their First Language148
4.49	Frequency Distribution of Responses to Statement Number 30: The High-stakes Testing Movement is Resulting in a Significant Increase in Student Dropout Rates149
4.50	Frequency Distribution of Responses to Statement Number 31: No High-stakes Decision Such as Grade Retention or Graduation Should Be Based on the Results of a Single Test149

TABLE	Page
4.51 Means and Standard Deviations by Campus Classification for Statements 20 – 31	152
4.52 Results of t-tests by Campus Classification for Statements 20 – 31	155
4.53 Means and Standard Deviations by Gender for Statements 20 – 31	159
4.54 Results of t-tests by Gender for Statements 20 – 31	162
4.55 Means and Standard Deviations by Years of Administrative Experience for Statements 20 – 31	166
4.56 Results of t-tests by Years of Administrative Experience for Statements 20 – 31	169
4.57 Means and Standard Deviations by Current Campus Rating for Statements 20 – 31	172
4.58 Results of t-tests by Current Campus Rating for Statements 20 – 31	175

CHAPTER I

INTRODUCTION

According to Robert Linn (2000) the use of tests and assessments has been a key element in five waves of educational reform during the past 50 years. Linn stated that these waves include the role of tests in tracking and selection emphasized in the 1950s, the use of tests for program accountability in the 1960s, minimum competency testing programs of the 1970s, school and district accountability of the 1980s, and the standards-based accountability systems of the 1990s. In addition, Wong and Nicotera (2007) report that the standards-based accountability systems of the 1990s have evolved into data driven performance-based accountability systems currently in vogue.

As with reform efforts in many other states, testing has featured prominently in Texas. The landmark case of *Rodriquez v. San Antonio ISD* ruled the system of school finance in Texas was unconstitutional in that it discriminated against students living in poor school districts. Although the U.S. Supreme Court subsequently overturned the *Rodriquez* case in 1973, the case acted as a catalyst for the Texas legislature to try to remedy inequities in school finance. In 1979, the Texas legislature passed the Equal Educational Opportunity Act, which established the first state-mandated test called the Texas Assessment of Basic Skills (TABS). The TABS test, administered from 1980 to 1985, was a survey-type assessment without sanctions for test takers.

The style and format for this record of study follow that of the *Journal of Educational Research*.

A resounding message from the National Commission on Excellence in Education (NCEE) 1983 report, *A Nation at Risk: The Imperative for Educational Reform*, was that schools were not preparing students adequately for the workplace. This report, a catalyst for school reform, suggested that the national economy was in danger unless schools improved the quality of the education of their graduates (NCEE, 1983). In 1984, using impetus created by the *A Nation at Risk* report and on the recommendations of the Select Committee on Education and its influential chair H. Ross Perot, the Texas legislature passed a comprehensive education reform law mandating sweeping changes in education in Texas. This state reform movement mandated that all public schools follow a state-mandated curriculum called “essential elements” and mandated basic skills testing of students in odd numbered grades. The Texas Educational Assessment of Minimum Skills (TEAMS) was implemented in 1985. High school students were required to pass the “exit level” version of TEAMS, administered in the eleventh grade, in order to receive a diploma. TEAMS only tested students in math and English language arts.

In the fall of 1990, changes in state law required the implementation of a new “criterion-referenced” testing program, the Texas Assessment of Academic Skills (TAAS) and established end-of-course tests for selected high school course subjects (Haney, 2000). As compared with TEAMS, TAAS was intended to shift the focus of assessment from “minimum skills to academic skills” and to test “higher order thinking skills and problem solving ability” (Texas Education Agency (TEA), 1997, p 1). Since 1994, successful completion of TAAS tests in reading, writing, and mathematics is a prerequisite for high school graduation. In order to hold schools and school districts

accountable for student learning, the State Board of Education was mandated to rate the performance of schools and school districts according to a set of “academic excellence indicators.” The Academic Excellence Indicator System (AEIS) reported TAAS results, dropout rates, and student attendance rates disaggregated by ethnicity and socioeconomic status. Texas high schools have been rated as “exemplary,” “recognized,” “acceptable,” and “unacceptable.” In short, over the past decade TAAS has become an extremely high stakes test for students, educators and schools in the state of Texas (Haney, 2000).

The Texas State Board of Education in May of 1998 released Chapter 74 Curriculum requirements determined by Texas Essential Knowledge and Skills (TEKS). TEKS were required to be the center of the curriculum. The TEKS were blended into the current student assessment, the TAAS test.

As mandated by the 76th Texas Legislature in 1999 (TEA, 2002), the Texas Assessment of Knowledge and Skills (TAKS) was administered beginning in the 2002-2003 school year. The TAKS measures the statewide curriculum in reading at Grades 3-9; in writing at Grades 4 and 7; in English language arts at Grades 10 and 11; in mathematics at Grades 3-11; in science at Grades 5, 10, and 11; and in social studies at Grades 8, 10, and 11. The TAKS Information Booklets state that the Texas Assessment of Knowledge and Skills (TAKS) is a completely reconceived testing program. It further states the TAKS test includes more of the Texas Essential Knowledge and Skills (TEKS) than the Texas Assessment of Academic Skills (TAAS) did and attempts to ask questions in ways that are more authentic. Satisfactory performance on the TAKS at Grade 11 is prerequisite to a high school diploma.

Statement of the Problem

Schrag (2000) stated, “that as education has risen to the top of the national agenda, a great wave—some would say a frenzy—of school reform has focused on two related objectives: more-stringent academic standards and increasingly rigorous accountability for both students and schools” (p. 1). According to Haney (2000), high-stakes standardized testing as a means of reform has captured the support of many local, state, and national educational leaders including the President, members of Congress, a majority of governors, state legislatures, and boards of education. As the former Governor of Texas and current President, George W. Bush’s administrations have figured prominently in the current in this trend. On January 8, 2002, President Bush signed into law the No Child Left Behind (NCLB) Act of 2001. According to the U. S. Department of Education, the Act is the most sweeping reform of the Elementary and Secondary Education Act (ESEA) since ESEA was enacted in 1965. According to the 40th Annual Phi Delta Kappa/Gallup Poll (Bushaw & Gallup, 2008), the NCLB Act represents the greatest federal incursion into K-12 education to date.

Although testing and accountability are intended to improve achievement and motivate staff and students, concerns have been raised in both the media and the professional literature (e.g., Heubert & Hauser, 1999; Linn, 2000; Nichols & Berliner, 2007) about possible unintended consequences of these programs. The late U.S. Sen. Paul Wellstone (D-Minn.) proposed a bill that would authorize the National Research Council (NRC) of the National Academy of Sciences to study the consequences of high-stakes testing for students, teachers and schools.

Experts such as Stanley Rabinowitz, co-director of standards and assessment development at WestEd (2003), a federally funded research laboratory, complain that good research on the issue simply has not been done. He stated that the bulk of studies on the subject have been biased, coming either from overt critics of high-stakes testing or from states eager to defend their own programs.

What is clear is that our local, state and national political leaders have embraced the use of high-stakes testing as a means to improve education by holding educators accountable for student achievement. The propagation of high-stakes testing occurs without ascertaining the full consequences of these actions. What is the full impact of this unabated movement? This study provides unbiased research directly from practicing Texas educational leaders in order to address the claims of high stakes test supporters, critics, and examine the possible unintended results of Texas' high-stakes testing program.

Purpose of the Study

In 1987 the legislature established the Legislative Education Board to oversee the implementation of state-mandated education reforms and to set public education policy, which resulted in the establishment of TEA divisions including the Office of Accountability (TEA, 2003). Although testing and accountability are intended to improve achievement and motivate staff and students, concerns have been raised in both the media and the professional literature (Heubert & Hauser, 1999; Linn, 2000; Nichols & Berliner, 2007) about possible unintended consequences of these programs. The purpose of this study was to assess the impact of the high stakes standardized test

movement in Texas secondary schools. The method to accomplish this task was to compare the perceptions between Texas secondary school administrators and supporters, critics, and unintended consequences of high-stakes testing.

Research Questions

The three research questions of this study are:

1. Do perceptions of high-stakes testing supporters differ from those of Texas secondary school administrators in the membership of TASSP?
2. Do perceptions of researchers of high-stakes testing reported in current literature differ from those of Texas secondary school administrators in the membership of TASSP?
3. Do perceptions of high-stakes testing critics differ from those of Texas secondary school administrators in the membership of TASSP?

Operational Definitions

The following terms used throughout this record of study are defined as follows:

Accountability: The systematic collection, analysis, and use of information to hold schools, educators, and others responsible for student academic performance.

Criterion-referenced Tests: An approach to testing in which individual's score on a test is interpreted by comparing it to a prespecified standard of performance (Gall, Borg & Gall, 1996).

Critics: Education researchers who have published in the timeframe beginning with the call for high-stakes testing by *A Nation at Risk* through the present and categorically oppose the use of high-stakes test to improve public schools.

High-stakes Tests: Tests that carry serious consequences for students or for educators. Schools may be judged according to the school-wide average scores of their students. High school-wide scores may bring public praise or financial rewards; low scores may bring public embarrassment or heavy sanctions. For individual students, high scores may bring a special diploma attesting to exceptional academic accomplishment; low scores may result in students being held back in grade or denied a high school diploma (American Educational Research Association (AERA), 1999).

Norm-referenced Tests: An approach to testing in which individual's score on a test is interpreted by comparing it to the scores earned by a norming group (Gall, Borg & Gall, 1996).

Reliability: The consistency of your measurement, or the degree to which an instrument measures the same way each time it is used under the same condition with the same subjects.

Supporters: Policymakers and educational researchers who have published in the timeframe beginning with the call for high-stakes testing by *A Nation at Risk* through the present and categorically endorse the use of high-stakes test to improve public schools.

Texas Assessment of Knowledge and Skills (TAKS): Mandated by the 76th Texas Legislature in 1999 (TEA, 2002), the Texas Assessment of Knowledge and Skills (TAKS) was administered beginning in the 2002-2003 school year. The TAKS measures the statewide curriculum in reading at Grades 3-9; in writing at Grades 4 and

7; in English Language Arts at Grades 10 and 11; in mathematics at Grades 3-11; in science at Grades 5, 10, and 11; and in social studies at Grades 8, 10, and 11.

Texas Association of Secondary School Principals (TASSP): The Texas Association of Secondary School Principals is an association formed by and for over 5000 campus level administrators. Established in 1922, its purpose is to build an active network of educators that want to take responsibility for the quality of school leadership. TASSP focuses on the need for collaboration between all stakeholders in education while using as its foundation a very effective volunteer force that provides a statewide knowledge base and informed leadership.

Texas Essential Knowledge and Skills (TEKS): TEKS are the center of the curriculum and, as such, define the basic content of the instructional and assessment program in Texas. TEKS outline the knowledge and skills required of every student on the TAKS.

Texas Education Agency (TEA): The TEA is the administrative unit for primary and secondary public education. Under the leadership of the commissioner of education, the TEA's duties include overseeing the development of the statewide curriculum, administering the statewide assessment program, administering a data collection system on public school students, staff, and finances, and rates school districts under the statewide accountability system.

Unintended Consequences: Unforeseen results of the high-stakes accountability movement on public schools and its stakeholders as reported by educational researchers who have published in the timeframe beginning with the call for high-stakes testing by *A Nation at Risk* through the present.

Validity: Indicator of the extent to which a test truly measures what it purports to measure (Wong & Nicotera, 2007).

Assumptions

The following assumptions are assumed to be true for the purposes of this study.

1. The researcher was impartial in collecting and analyzing the questionnaire data.
2. The instrument used in this study will be able to measure the effects of high stakes tests on Texas secondary schools as perceived by secondary administrators of TASSP.
3. The respondents to the survey objectively and honestly answered questions posed to them regarding the study.
4. The interpretation of the data collected has accurately reflected that which was intended.

Limitations

The following limitations are assumed to be true for the purposes of this study.

1. The scope of the study is limited to secondary schools in Texas whose administrators are members of TASSP.
2. The scope of the study was limited to members of TASSP with email and Internet access.

3. Findings from this study may not be generalized to any other group than the secondary schools in Texas whose administrators are members of TASSP.
4. TASSP does not require that e-mail addresses are provided by its membership, thus several members were excluded from the study for the lack of a recorded e-mail address.
5. Due to the delivery method of the survey, the potential exists that the request for participation in the study may be determined to be spam and could be filtered out by school district firewalls and other security measures.

Methodology

Population

The population selected for this study was the 2005-2006 membership of Texas Association of Secondary School Principals (TASSP). The Texas Association of Secondary School Principals is an association formed by and for over 4000 campus level administrators at the middle and high school level. The 2005-2006 membership list was provided to the researcher in spreadsheet form alphabetized by the name of the administrators' school. Each member was assigned a number in order starting with 1 – 4,641. The researcher used the random number generator program developed by Scott Donato Saccenti. Once the researcher inputted the range of potential responders, 1 – 4,641, the program produced random numbers within the range. The researcher used the program to produce 600 random numbers. The generation of random number in excess of 400 was completed to handle the possible duplication of numbers and address the issue of TASSP members without recorded email addresses. The researcher designated

each potential responder by matching his or her number with the number selected by the random number generator. Thus, a random sample of 400 administrators was drawn from the TASSP membership.

Instrumentation

A questionnaire was developed following the guidelines provided in *Educational Research: An Introduction* (Gall, Borg & Gall, 1996). The questionnaire elicited information pertaining to the perceptions of secondary administrators pertaining to the impact high stakes testing has on the stakeholders of their home campus. Responses to the questionnaire were made on a five-point Likert scale. A field test using administrators from a neighboring school district was conducted. The research instrument was pre-tested by a group of five secondary administrators from other districts to assess meaning, wording, and other validity matters. Validity of the questionnaire was also addressed through review of the literature. Questions were developed based on factors of high-stakes testing identified in literature. A field test using 10 nonparticipating secondary administrators from neighboring school districts was conducted in Fall 2006 to assure clarity and content validity. Reliability was determined by calculating the alpha reliability of the questionnaire.

Procedures

The procedure for completion of the survey was an electronic submission of the instrument to each of the selected administrators. Perspective respondents received an email message one week prior to the first submission. They had an opportunity to respond to the e-mail if they wish to not participate in the survey. The electronic survey

included an introduction page explaining the proposed survey followed by the survey instrument on the next page. Participants were told that the instrument could be completed in approximately 20 minutes or less.

Follow-up occurred as warranted by the return rate. Within two weeks following the second electronic follow up, a third electronic follow was attempted. One week later, a fourth and final request was made.

Data Analysis

Quantitative data obtained was analyzed using *Statistical Package for Social Sciences* (SPSS). Results of the study were reported using numerical and graphical techniques to report such statistical such as mean, mode, standard deviations, frequencies, and percentages. In addition, t-tests were conducted on all statements for each research questions. Independent t-tests were conducted to assess whether the means of two independent groups are statistically different from each other. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years vs. 15 or More Years), Campus Classification (Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable). When comparing two independent groups, their variances must be relatively similar. Levene's Test for Equality of Variance was used to check for this. If the significance for Levene's test was 0.05 or below, then the "Equal Variances Not Assumed" test was used to determine significance. Otherwise, the use the "Equal Variances Assumed" test was used to determine significance. An alpha level of .05 was used to establish significance.

Significance Statement

Vigorous debates about the nature and role of high-stakes tests and accountability systems are healthy and needed (Cizek, 2002). The American Educational Research Association (1999) stated that because the stakes are so high for so many students, additional research should begin immediately to learn more about the intended and unintended consequences of testing in educational decision-making. Further, American Educational Research Association states that if tests are going to be used to determine which students will advance and what subjects schools will teach, it is imperative that we understand the most effective way to measure student learning and how the use of high-stakes testing will affect student drop-out rates, graduation rates, course content, levels of student anxiety, and teaching practices.

Texas provides an ideal context in which to study high-stakes testing because its accountability system has received attention from the media and the policy community, and it has been cited as possibly contributing to improved student achievement (e.g., Grissmer & Flanagan, 1998; Grissmer, Flanagan, Kawata & Williamson, 2000).

This study investigated both the intended and unintended effects of high-stakes testing on Texas secondary campuses as perceived by their campus administrators. This information has significance to all who are involved in leading learning communities through the arduous obstacles and challenges resulting from high-stakes testing movement in the secondary schools of Texas. The findings of this study should also be of interest to local, state, and national educational policymakers, legislators, educators, and fellow researchers.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

An evaluation of the success of any reform movement is determined by the degree to which student achievement is enhanced, thus student assessment is an essential part of all educational reforms (Klein & Hamilton, 2001). According to Mazzeo (2000), the role of student assessment can now be viewed as a powerful instrument of reform and change. Numerous researchers (Nichols & Berliner, 2007; Wong & Nicotera, 2007) have found that policymakers in the United States and the world have increased the use of student assessment based on the belief that these policies will motivate students, parents, teachers, administrators, and citizens and will guarantee that appropriate curriculum is taught. In the current pursuit to hold public schools accountable, policy makers are implementing ranking, rating and grading systems that primarily rely on standardized test scores. Achieve, Inc. (2002), in a review of educational reform commissioned by Texas Education Agency, stated, “Both in terms of the longevity and the results of education reform, Texas has been a leading state in what has become a national effort to raise academic standards, measure results against them, and hold schools and students accountable for those results” (p. 11). While researchers debate the impact of these state-mandated standardized high-stakes tests, the standards and accountability movement continues to gain momentum at the local, state and federal level.

This study provided an overview of the historical development of educational evaluation in the United States, chronicle the development of standardized tests, and revisit the growth of high-stakes testing. In addition, this paper will take a comprehensive look at *A Nation at Risk* and its impact. Further, this study will look at the federal incursion into local schools facilitated by politically appointed Commissions and legislation such as Elementary and Secondary Education Act (ESEA), the Secretary's Commission on Achieving Necessary Skills (SCANS) Report (SCANS, 1990) and No Child Left Behind. As the focus of this study involves Texas secondary administrators, this paper will examine the progression of the high-stakes testing reform movement in Texas. Finally, this study will examine the effects of the state-mandated high-stakes testing reform movement as perceived by both supporters and critics and the unintended effects of this movement.

Federal impetus for the standards and accountability movement has been provided by legislation such as the National Defense Education Act, the Civil Rights Act of 1964, the Elementary and Secondary Education Act (ESEA) of 1965 and No Child Left Behind Act of 2001. In addition, politicians have appointed commissions that have called for educational reform, primarily using standards-based high-stakes assessment and accountability. The 1983 *A Nation at Risk* report prepared by the National Commission on Excellence in Education, the SCANS Report (1990), and Goals 2000 (U.S. Congress Senate, 1993) are recent examples of the utilization of Presidential Commissions.

These labors represent a pervasive movement whose goal is to reform education by raising stakes for students, teachers, administrators, schools and school systems

(Mazzeo, 2000; Nichols & Berliner, 2007; Perkinson, 1995; Wong & Nicotera, 2007).

In response to No Child Left Behind (NCLB), performance standards for grade advancement and promotion have been implemented at the elementary and middle school level in most states and is now required for all states. While NCLB primarily addresses elementary and middle school levels, escalating impact of the standards-based high-stakes assessment and accountability movement at the high school level has, according to WestEd (2003), resulted in 24 states instituting exit-exams as a graduation requirement while six more states are in implementation stage. While the impact and importance of high-stakes testing currently appears to be reaching its zenith, high-stakes testing is not a new concept.

Educational Evaluation

The failure of the U. S. Constitution to mention education in conjunction with the Reserve Clause of the Tenth Amendment to the Constitution left education to the discretion of the states. Thus, individual states were left to establish, develop, supervise, and evaluate education programs. A republican ideology and a general distrust of government led our founding fathers to embrace locally controlled public schools (Tyack & Timar, 1999). The development of both education and educational evaluation has not occurred in a vacuum; rather it has been a reflection of the societal norms, values and cultures that were shaping society at that particular time (Parker, 1994). A historical look at educational evaluation reveals a shift toward high-stakes driven instruction and increased school accountability.

The dual goals of education in colonial America were to provide access to the written word and to nurture religious and moral development (Lutz & Merz, 1992; Sadker & Sadker, 2000). Education often took place at home with instruction typically provided by family members. The Massachusetts Law of 1647, also known as the Old Deluder Satan Law, was the first to establish district schools (Perkinson, 1995). According to Perkins, the concept of the Massachusetts Law, which required every town of 50 families to establish a town school, spread throughout New England and Middle Atlantic states. District schools collected tuition from all students except the very poor in order to teach reading, writing, arithmetic, and religion. In the mid 1800s, conversion of district schools to public schools supported by public funds took place slowly as policymakers in each state followed Massachusetts' lead to mandate student attendance (Sadker & Sadker, 2000). Once public monies were used to fund school, the cry for educational evaluation began.

The initial evaluative tools used in the fledgling United States were emulated from Great Britain. Much like Royal Commissions, Presidential Commissions conducted evaluations by gathering evidence and typically presented their findings in the form of testimony (Madaus, Stufflebeam & Scriven, 1983). Presidential Commission's external inspectors conducted yearly formal evaluations of school conditions and student performance. Presidential Commission's recommendations were just advisory and non-binding. The use of Presidential Commissions, such as National Commission on Excellence in Education and its report *A Nation at Risk* are still being used as a means to evaluate education and recommend change. In 1845, recommendations from an early

Presidential Commission led to the development and implementation of the Boston Survey under the direction of the Boston School Committee.

The Boston Survey tested a sample of Boston students in definitions, grammar, history, philosophy, astronomy, writing, and arithmetic (Worthen & Sanders, 1987). This resulted in the first use of printed tests for assessment of student attainment. During the same time frame of the Boston Survey evaluations, education luminary Horace Mann introduced written essay exams into the Boston grammar schools (Worthen & Sanders, 1987). The advent of written exams enabled stakeholders to gather data to base educational decisions and to facilitate inter-school comparisons. Data compiled from the written exams were used to make political decisions such as the annual appointment of headmasters (Madaus, Stufflebeam & Scriven, 1983). These two developments were the first attempts at objectively measuring student achievement to assess the quality of a large school system (Fitzpatrick, Sanders, & Worthen, 2004). Interestingly, both evaluations reported a low level of student performance and other areas of concern (Worthen & Sanders, 1987).

In 1887, Joseph Rice initiated one of the first comparative studies on the quality of instructional methods (Fitzpatrick, Sanders & Worthen, 2004). Rice used test scores to conduct multi-district comparative study of spelling instruction and arithmetic. According to Worthen and Sanders (1973), Rice's motivation was to provoke curriculum revision. Rice's findings revealed huge differences between schools' student performance, leading him to conclude that school time was used inefficiently. Thus, he proposed establishing standardized examinations (Fitzpatrick, Sanders & Worthen, 2004). In 1982, the National Education Association (NEA) developed its "national

policy for high schools” (Sadker & Sadker, 2000, p. 291). This call for increased rigor in high school curriculum is reminiscent of 1983’s *A Nation at Risk*. Efficiency in education would soon become the impetus for the next step in the evolution of evaluation.

Beginning in the 1900s, the principles of efficiency extolled by pundits of Fredrick W. Taylor’s “scientific management” philosophy began to influence education and testing (Fitzpatrick, Sanders & Worthen, 2004). The turn of the century was also witness to the rise of “muckrakers,” investigative journalists who applied pressure on school administrators and school boards, who were than motivated to evaluate their schools and programs in light of modern business methods (Callahan, 1962). Callahan details when the *Ladies’ Home Journal* and the *Saturday Evening Post* published articles criticizing schools and demanding that schools provide evidence of their contributions to society or have their budgets cut.

In response to this intense scrutiny, Charles Keyes, the president of the National Council of Education, initiated the forming of a committee on “Tests and Standards of Efficiency of Schools and School Systems” in 1911. Frank E. Spaulding, superintendent of Newton Massachusetts schools and the University of Chicago professor Franklin Bobbitt spearheaded the movement to apply scientific management to education (Callahan, 1962). Their efforts resulted in a proliferation of school surveys administered by efficiency experts that were often university professors. Further, Callahan reported that external efficiency experts used surveys and standardized tests to produce quantitative data for scales of school measurement.

Those hostile to public schools were using the results of school efficiency surveys to criticize schools and school personnel resulting in a loss of public confidence. The media's willingness to publish efficiency experts' reports had a dramatic effect on educational administration and led to the implementation of teacher-rating procedures and standardized tests (Callahan, 1962). Despite the increasing availability and use of standardized tests permitting inter-district comparisons, evaluation during this period was almost exclusively used at the local level (Walberg & Haertel, 1990).

One of the leading figures in the development and utilization of educational testing was E. J. Thorndike, regarded by many as the father of the standardized testing movement (Worthen & Sanders, 1987). Thorndike proposed the use of norm referencing to evaluate school efficiency. As a result, norm referenced tests increase the potential of standardized tests to be used to make comparisons across districts, states, and the nation (Walberg & Haertel, 1990). According to Popham (1992), the power of norm-referenced exams is their capability to "ascertain an examinee's status in relation to the performance of a group of other examinees that have completed the exam" (p. 24).

Ralph Tyler's work, especially the landmark "Eight-Year Study" (1932-1940), laid the foundation for a new approach for educational testing (Madaus, Stufflebeam & Scriven, 1983). "The Eight-Year Study" used a battery of tests, scales, inventories, questionnaires, check lists, pupil logs, and other measures to determine the degree to which curricular objectives were met in each of the 30 high schools included in the study (Worthen & Sanders, 1973). Recognized as the father of modern educational evaluation, Ralph Tyler was intent on assessing how well stated instructional objectives had been achieved. While Tyler continued to make use of the scientific method, he was

the first to make the distinction between measurement and evaluation (Parker, 1994).

Tyler's child-centered view stated that the purpose of evaluation was the appraisal of an educational program's quality, not as the appraisal of students.

Tyler believed that the success of an educational program was determined by evaluating the extent that it promoted mastery of the program's behavioral objectives (Walberg & Haertel, 1990). However, Tyler recognized that unintended effects occurred in all evaluations, thus recognition of these unintended effects must be included in order for an evaluation to be comprehensive (Wolf, 1990). Tyler's focus on outcomes rather than inputs provided a significant advantage over the comparatively costly and disruptive scientific procedures required by the scientific approach used by Rice and others (Madaus, Stufflebeam & Scriven, 1983). Thus, the goal-driven model of evaluation that used criterion-referenced tests became a viable alternative to norm-referenced tests (Parker, 1994). Criterion-referenced tests compare student's performance on a measurement relative to an external expected level of performance (Popham, 1992). Two prominent characteristics of criterion-referenced tests include a clearly defined topic or skill and a predetermined standard of acceptability (Popham, 1992). Walberg (2003) believes that the current accountability and systemic reform could be traced to the works and influence of Ralph Tyler.

E. F. Lindquist, Ralph Tyler and others combined efforts to establish the Princeton based Educational Testing Service to protect the public from poorly made tests or inappropriate use of tests (Parker, 1994). An abundance of commercially developed standardized tests were produced during the 1950s. The proliferation of standardized tests was enhanced by the advent of new technologies such as scoring machines. Testing

entities increased the uniformity of testing conditions by developing procedures that resulted in tests being administered with the same set of directions, time constraints, and uniform scoring procedures (Ebel & Frisbie, 1991). Ebel and Frisbie (1991) stated that the type of standardized tests used during this time-period included aptitude, achievement, and attitude inventories. Tyler's model of the use of curriculum, measurement, and evaluation continues to influence testing (Walberg & Haertel, 1990).

The United States saw a dramatic increase in the federal role in education from the late 1950s (Walberg & Haertel, 1990). The launching of Sputnik by the Soviet Union and the civil rights movement lent impetus to federal legislation including the National Defense Education Act, the Civil Rights Act of 1964, and the Elementary and Secondary Education Act (ESEA) of 1965. This legislation, enacted to correct perceived problems in education and society, provided massive financial support for schools throughout the nation (Lutz & Merz, 1992; Nichols & Berliner, 2007, p. 3; Sadker & Sadker, 2000; Walberg & Haertel, 1990; Wong & Nicotera, 2007, p. 7). These funds were accompanied with mandates that included specific evaluation requirements and posed new challenges for evaluators.

Fitzpatrick, Sanders and Worthen (2004) and Nichols and Berliner (2007, p. 3) identify the passage of the Elementary and Secondary Education Act (ESEA) as the one event most responsible for the emergence of contemporary program evaluation. Wolf (1990) and Nichols and Berliner (2007) stated that the prominence given to educational evaluation could be traced to the passage of ESEA. For the first time, ESEA mandated annual evaluations to verify the success of Title I and Title III programs by participating school districts (Popham, 1975; Wolf, 1990). Senator Robert F. Kennedy contended that

in order to ensure local agencies used their federal grants appropriately mandatory evaluations must be a provision of this law (Popham, 1975; Worthen & Sanders, 1973). This bill massively increased federal funding for education by providing thousands of dollars in grants to local schools and other educational entities. Thus, ESEA created a financial incentive that initiated an interest in educational evaluation that has yet to subside (Popham, 1992). To meet the requirements of ESA, evaluators were required to make professional judgments about the merit and worth of educational programs and established accountability. Thus, there was a need for an evaluation model that went beyond a model of a comparison of measured outcomes to stated goals (Walberg & Haertel, 1990). Popham (1975) and Worthen and Sanders (1973) report educators of the time were ill prepared to meet the requirements of this law. Popham (1974) was quoted, “a scene in which educational evaluation was required but the would-be evaluators were nonexistent, provided the chief stimulus for what is now a rapidly expanding field” (p. 4). One additional result of ESEA was the shift from objectives to decisions resulting in evaluation becoming a political phenomenon (Parker, 1994).

As the quest for school accountability soared, both educators and policymakers realized that no useful mechanism existed to provide nationwide data to guide public policy regarding educational spending and curriculum reform (Ebel & Frisbie, 1986). The search for such a mechanism inspired educators to devise and development standardized tests.

Standardized Testing

The distinctive characteristics of a standard test are uniform testing conditions and scoring procedures (Ebel & Frisbie, 1991). A standardized test score is typically reported by using a formula to convert a raw score into a statistical metric. The use of statistical metrics to report scores began in the 1910s (Rogers, 1995). Common statistical metrics are IQ, SAT, and ACT scale scores and percentile rankings. This quantitative approach provides an appearance of objectivity and encourages the use of multiple choice and short answer format (Bolton, 2000).

Rogers (1995) delineated two formats of standardized tests, “speed” format, which are strictly timed, and “power” format, which is generally untimed (p. 256). The Stanford, California, and Iowa achievement tests are speed tests while the Texas Assessment of Academic Skills (TAAS) and the Texas Assessment of Knowledge and Skills (TAKS) are power tests. In an interesting side note, Sacks (1999) and groups such as FairTest (2000) question why females have slightly better high school and college grades but have been routinely outscored by males on the Scholastic Aptitude Test (SAT) since 1972. They believe this evidence suggests a bias exists in this example of a speed test that seems to favor white males over females and other minority groups. Some colleges, including MIT, have adjusted rating of male and female applicants in recognition of how the SAT was and is under forecasting female college performance. The designated length of the test was not the only means of classifying standardized tests.

The standardized test pundits split into two distinct groups, achievement and aptitude (Popham, 1992). Aptitude tests, such as the Binet-Simon scale, the Army Alpha and Beta test, and the Scholastic Aptitude Test (SAT) are designed to measure an individual's intellectual potential and to predict a student's performance in future educational settings (Bolton, 2000). In contrast, Achievement tests are restricted to attempting to measure the knowledge and skills acquired by the student. The Stanford Achievement test, Iowa Every-Pupil test, and Texas' End of Course Exams are examples of achievement tests. Popham (1992) asserts that despite the distinct purposes of these two types of tests, they often overlap in a given test.

The onset of standardized test in the United States began in France. In response to a request in 1904 from the French government, Alfred Binet collaborated with a physician named Simon to develop the Binet-Simon Scale. Completed in 1905, this test was designed to diagnose and track mentally deficient children. In contrast to the future uses of his scale, Binet refused to regard IQ (Intelligence Quotient) as a general device to rank students (Gould, 1981). Gould affirmed that Binet designed his scale for the single purpose of identifying children in need of special education in order to help them. Further, he stated that Binet had concerns that his work could be misused in a manner in which IQ would be used to permanently label children. The proliferation of Binet's scale would soon lead to worldwide acceptance.

H. H. Goddard, director of research at the Vineland Training School for Feeble Minded Girls and Boys in New Jersey, brought Binet's scale to America and translated Binet's work into English (Gould, 1981). The antithesis of Binet, Goddard's hereditarian beliefs regarded tests scores as a single, innate entity resulting solely from

heredity (Gould, 1981). In his book, *The Mismeasure of Man*, Gould documented Goddard's attempt to develop a single scale suitable to rank mental deficiencies and his desire that his scale would be used to improve and protect society from those he deemed unfit to breed. His influences are attributed to helping create tighter American immigration standards. By 1982, Goddard recanted his beliefs and embraced Binet's viewpoint; however, the momentum he created would carry the hereditarian movement forward.

In 1916, a few years after Binet's death, Stanford professor Lewis W. Terman modified Binet's scale renaming it the Stanford-Binet IQ test. Terman's Stanford-Binet used a single score to represent general intelligence (Hopkins, Stanley & Hopkins, 1990). Through his efforts, Terman was the primary architect of the popularity of the Stanford-Binet test and helped it become the standard for all intelligence tests (Gould, 1981).

Lewis M. Terman, according to Gould (1981), "dreamed of a rational society that would allocate professions by IQ scores" (p. 157). Terman was not hesitant to attach exceedingly high stakes to his tests (Sacks, 1999). L. M. Terman was able to develop successful partnerships with publishing company, Houghton Mifflin, the initial publisher of Terman's Stanford-Binet test and his tract, "The Measurement of Intelligence" (Gould, 1981). In 1923, Terman combined efforts with the *World Book* to publish the Stanford Achievement Test, his latest development (Bolton, 2000). His association with these publishing companies led to the commercialization and proliferation of the mental testing industry dominated today by large corporations. Further, the Individuals with Disabilities Education Act (IDEA) has institutionalized Terman's model in the United

States (Gould, 1981). This law determined that entrance into special education requires a minimum 15 point discrepancy between potential, measured as IQ, and actual performance. Terman's success was followed by the introduction of various standardized tests developed for a myriad of diverse needs including making high-stakes decisions. Almost from their origin, high-stakes have been inextricably linked to standardized tests.

Supporters' Positions on High-stakes Testing

According to Klein, Hamilton, McCaffrey and Stecher (2000), educators should understand that testing policy represents a political solution to an educational problem. High-stakes standardized testing as a means of reform has captured the support of many local, state, and national political leaders including the President, members of Congress, a majority of governors, state legislatures, and boards of education (Haney, 2000). As a direct result of his role as Governor of Texas and as President of the United States, George W. Bush has become one of the most influential Supporters of reforming schools by using state-mandated high-stakes tests. Policymakers expect testing programs to certify a student's level of achievement, provide information about an education system's effectiveness, motivate student performance, bringing coherence to a curriculum, and hold schools and educators accountable for student performance (Hamilton, Stecher & Klein, 2002). On August 1, 2001, the President said, "Accountability is an exercise in hope. When we raise academic standards, children raise their academic sights. When children are regularly tested, teachers know where and how to improve. When scores are known to parents, parents are empowered to push

for change.” According to Dyack (1990), evidence of a growing distrust of local school control can be found in the 1986 National Governors Association’s recommendation that “states take over and run districts that fail to educate children.”

According to the Fact Sheet prepared by the House Education and Workforce Committee (2002), accountability is the centerpiece of President Bush’s plan to improve public schools and close the achievement gap that has existed between disadvantaged students and their more affluent peers. According to Wong and Nicotera (2007), the standards-based movement’s central new expectation is that all children should receive the high level of education once reserved for a fraction of our nation’s students (p. 11). This paradigm shift has radically changed expectations for the poor and previously excluded and is having a tremendous impact on educators, lawmakers, and students. Recently, expectations for special needs students have also been raised.

In spite of remarkable changes over the past quarter-century, special education reform efforts have fallen short of universally improving the achievement outcomes of all students with disabilities in a substantial way (deFur, 2002). In response to these findings, the 1997 amendments to the Individuals with Disabilities Education Act (National Information Center for Children and Youth with Disabilities, 1998) asserted that the educational progress of students with disabilities had been limited by low academic expectations that in turn narrowed student access to the general curriculum. Furthermore, the IDEA Amendments of 1997 indicated that participation in state accountability systems (assessment) was the key to increasing participation in the general curriculum and raising the academic expectations for all students with disabilities. The No Child Left Behind Act has mandated the inclusion of special

education students into the accountability movement or face having their school labeled as Not Meeting Adequate Yearly Progress and potentially losing Title I federal funds.

Supporters of the testing movement believe that high-stakes testing combined with holding teachers, administrators, and students accountable for successful performance on these tests will improve our public education system (Mazzeo, 2000; Nichols & Berliner, 2007; Perkinson, 1995; Wong & Nicotera, 2007). The explicit belief is that educator's need to be held accountable through high-stakes tests to motivate them to teach better particularly to push the least motivated ones to perform. According to Stecher and Hamilton (2002), test-based accountability can lead to educators working harder to cover more material in a given time.

Further, Supporters of testing believe that high-stakes tests have inspired educators to adopt better curricula and employ more effective teaching methods. Lawmakers are not the sole proponents of testing. A survey of 1,023 parents of school-age children from Arizona found that 83% of respondents believe tests provide important information about their students' educational progress, and 9 out of 10 sought comparative data about their children and the schools they attend. Two-thirds of the parents surveyed said they positively supported receiving standardized test results for their children in every grade; the respondents were evenly divided in giving the tests once or twice a year (Driesler, 2001). Jones, Jones and Hargrove (2003) detail a national survey in which 66% of the public consider testing is at the right amount or more testing is needed. Thus, public support has fueled the expansion of the high-stakes testing movement.

A fundamental tenet of the testing movement is that when high-stakes tests are developed and used appropriately, they are the most sound and objective knowledge and performance measures available (AERA, 2000). According to Clarke, Haney and Madaus (2000), the trend is to call them assessments rather than tests, but the issues surrounding their uses are the same. Supporters claim that the use of student performance data generated by high-stakes tests should help educators refine programs, channel funding, and identify roots of success (Wong & Nicotera, 2007). When used appropriately, high-stakes tests can help promote student learning and equal opportunity in the classroom by defining standards of student achievement and by helping school officials identify areas in which students need additional or different instruction (Heubert & Hauser, 1999). Supporters point to the fact that Texas and North Carolina have experienced gains on the NAEP far above historical averages. Grissmer and Flanagan (1998) conducted a case study to identify the characteristics of each state that have contributed to these gains. State policies in both states included state standards by grade, standard based assessments, procedures for feedback to teachers and administrators, and accountability measures. Their study reported that the most plausible reason for these gains could be attributed to the similar systemic reform policies implemented in both states in the late 1980s and 1990s. A later case study, Grissmer et al. (2000) found that after controlling for various student demographic characteristics and other factors, Texas tended to have higher NAEP scores than other states and speculation was that this was due to the high-stakes accountability system in Texas. Of paramount importance to the researcher is the effect high-stakes tests will have on teachers.

Supporters believe that the use of high-stakes tests will help teachers focus on what is important to teach (Schlechty, 1997). Wong and Nicotera (2007) stated that assessment data provides information that can be used by administrators, teachers, and support staff to influence instructional improvement. In addition, assessment data provide indications of the extent to which students have learned instructional objectives as well as providing an indication of individual student progress from year-to-year. Educators should use empirical test data to assist in instructional decision-making to revise instruction for entire classes or courses, and to develop specific intervention strategies for individual students (Mertler, 2007; Wong & Nicotera, 2007). The expectation is that high-stakes based accountability will have a positive effect on both students and educators.

Supporters believe that these tests will result in higher expectations for teachers and their students. Testing and accountability are intended to improve achievement and motivate staff and students (Heubert & Hauser, 1999; Linn, 2000). Most Supporters believe that being held accountable for high-stakes tests results will motivate teachers to improve instruction (Hamilton, Stecher & Klein, 2002; Wong & Nicotera, 2007). Following the lead of Fredrick Taylor, they believed that the natural laziness of man is a serious problem that could best be handled with external pressure (Callahan, 1962). According to Nichols and Berliner (2007), this line of thought presupposes that educators are either lazy or ineffectual and will mend their ways only after they are given guidance or external motivation replete with public consequences. Supporters insist that supervisors of teachers must be part of the high-stakes based accountability equation.

Supporters accept as true the position that school administrators need to be held accountable by high-stakes tests to motivate them to be more effective in supervising their staff. One of Ronald Edmonds' Seven Correlates of Effective Schools deals with how important instructional leadership is on the part of successful schools (Hoy & Miskel, 1996, p. 249). According to Edmonds, the role of the principal is paramount for school reform to occur. Thus, critics support the public display of high-stakes test scores to motivate administrators into ensuring that standards on which the tests are based are an integral part of the curriculum and are being taught. The high expectations for assessment were publicly debated by the National Council on Education Standards and Testing (NCEST, 1992), whose deliberations led to the enactment of the Goals 2000 and revised compensatory education legislation (U.S. Congress Senate, 1993). As a result, today's expectation is that school and district outcomes on assessments will and should be made public. Consequently, school personnel, particularly building principals, may experience consequences, potentially career threatening, because of their students' performance on high-stakes assessments.

Students are also subject to the effects of high-stakes tests. Supporters support the supposition that students work harder and learn more because they know what is expected of them and that the high-stakes tests truly count (Wong & Nicotera, 2007). A recent RAND publication (Hamilton, Stecher & Klein, 2002), found that high-stakes tests provide students with unambiguous information about their own knowledge and skill levels (p. 16). In addition, high-stakes tests send clear signals to students about what to study. Further, the study reported that these tests motivated students to work harder in school and helps students associate personal effort with success. Supporters

maintain that doing poorly on high-stakes tests will lead to increase student effort to learn. A study in Chicago found that for 102 low-achieving sixth and eighth graders who were placed in a high-stakes testing context, the majority of the students showed increased work efforts which, in turn, translated into higher gains in learning (Roderick & Engel, 2001). In addition, a study of higher education students showed that frequent testing was more effective than frequent homework for improving their retention of information particularly among low-achieving students (Tuckman, 2003).

Moreover, Supporters of testing attribute high-stakes tests for the closing of the achievement gap between minority students and majority students in Texas. Some educational experts support the position that high expectations and standards can serve as an equalizer of educational opportunities for the diverse student population served by public education (Grissmer et al., 2000). Skrla, Scheurich, Johnson and Koschoreck (2001), from the Dana Center, reported on successful Texas school districts that have improved scores and closed minority score gaps despite having clientele with high poverty levels and exhibiting demographics typically associated with low performing schools. According to this report, these school district were successful because committed educators believed in the need for educational success for all children and worked to change the beliefs of those not in line with this conviction. The No Child Left Behind Act epitomizes the principle that the greatest benefactors of the current national accountability movement will be children from low-income and ethnic minority backgrounds (Townsend, 2002). Not everyone supports these aforementioned beliefs. According to Hamilton, Stecher and Klein (2002), research suggests that large-scale

high-stakes testing has brought about both positive and negative changes within school and classrooms (p. 16).

Critics' Positions on High-stakes Testing

The accountability movement has sparked fierce debates over the reported success of the high standards movement and the means of producing mastery of them. Critics take issue with many of the canons of the Supporters of high-stakes testing. Foremost, critics take issue with the tests themselves. In many circles, standardized tests have long been considered unfair and biased against students from ethnic minority and or impoverished backgrounds because these tests are based in large measure on the experiences of middle class European Americans (Hilliard, 2000; Neill & Medina, 1989). In addition, when tests are based primarily on multiple-choice items, the response options frequently distract test-takers from conveying what they understand and they do not take into account the possible logical explanations for “incorrect” choices that test-takers make (Falk, 2002).

Policymakers expect the roles of testing programs to include certifying a student's level of achievement, providing information about an education system's effectiveness, motivating student performance, bringing coherence to a curriculum, and holding schools and educators accountable for student performance (Hamilton, Stecher & Klein, 2002; Wong & Nicotera, 2007). Critics believe that high-stakes tests are not designed for these diverse roles; therefore, it is unreasonable to expect them to do so. Further, their stance is that a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational

quality of a school or school district. According to Huebert and Hauser (1999), the content standard's goal demands criterion-referenced testing; the school or student ranking goal demands norm-referenced testing. Therefore, one test cannot adequately do both.

One of the critics' primary positions is the stance that high-stakes tests movement is using an overly simple approach to improve education. According to Kohn (2004), using catch phrases such as "raising the bar," "accountability," and "higher standards," lawmakers, without an understanding of how children learn, have mandated a test-driven version of school reform that is lowering the quality of education in this country. Further, he believes a focus on standards and accountability ignores the process of teaching and learning in classrooms and does not provide the direction that teachers need to improve instruction. Test scores shows no evidence of opening children's access to great literature, to conceptual understanding in mathematics, to fluency in writing, or to other learning experiences that seriously address previous inadequacies in their education (McNeil & Valenzuela, 2000). Critics who judge high-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve. In addition, critics question if increases in test scores reflected real improvements in student achievement (Sheppard, 2002). They point out that students trained in one test format are less able to answer the same question in another format bringing in to question the generalization of test-based learning. Furthermore, the nature and format of most standardized tests provide little opportunity for students to use higher-order thinking to solve problems or to apply their knowledge to real-world problems. Critics of the high-stakes testing movement call for

assessments that make use of formats that call on students to demonstrate their knowledge in ways that are closer to those used in real life to solve problems, explain ideas, and apply understanding and skills and systems of assessment that evaluate student progress by considering multiple forms of evidence of a students' knowledge demonstrated in a range of ways (Falk, 2002).

Some researchers believe that high-stakes tests are expensive and result in diverting scarce resources and attention from serious problems. According to Sacks (1999) expenditures on standardized tests reached \$100 million a year in 1989, more than twice the amount spent in 1960 while enrollment increased by only 15 % during the same period. In addition, the development of a quality test requires significant funds. Funds are also needed for scoring tests, interpreting test results, informing the public of test scores, and educating parents, teachers, and administrators about how to interpret test results. The costs of testing programs may extend further into rewards and recognitions, test preparation materials, testing monitors, and even free breakfasts on test days (Jones, Jones & Hargrove, 2003). Given the shortage of dollars for education and the associated lack of adequate classroom space, computers, instructional materials, and teachers, one has to question whether or not the benefits outweigh the financial costs of high-stakes testing programs (Jones, Jones & Hargrove, 2003). McNeil and Valenzuela (2000), state that the pressure of high-stakes tests has led schools to spend scarce instructional dollars for test preparation, diverting materials and activities whose only value is to increase test scores, not to produce educated children well prepared for future endeavors.

Further, critics find fault with the consequences prescribed by the high-stakes accountability movement. Research finds that students who have been held back typically do not catch up, even with remedial help and low performing students learn more if they are promoted than if they are held back (Heubert & Hauser, 1999). A significant concern is that the negative effects of holding students back are often invisible to those who make retention decisions because they occur many years later (Heubert & Hauser, 1999). Research has generally suggested that grade retention makes students more likely to dropout (Goldschmidt & Wang, 1999). Interaction with graduation test requirements may result in increased numbers of dropouts (Clarke, Haney & Madaus, 2000). Despite the substantial body of evidence that points to the harmful effects of retaining students in grade and despite the urgings of national experts and commissions to rely less on standardized testing and more on broader measures of student progress when making high-stakes decisions, the accountability movement continues on unabated (Falk, 2002). Emphatically, critics believe that high-stakes decisions such as grade retention or graduation should not be based on the results of a single test (AERA, 2000). According to Hamilton, Stecher and Klein (2002), widespread agreement exists among educators and measurement experts that high-stakes decisions about individuals should be based on factors other than test scores alone. Critics are also concerned about the impact of high-stakes tests on classrooms.

Critics believe that the pressures inherent in preparing students for high-stakes tests are having adverse effects on students and teachers; including driving good teachers out of education. Nichols and Berliner (2007) state that high-stakes testing movement is sapping the strength and vitality out of many teachers and administrators (p. 168).

Critics' belief is that high-stakes tests compromise educational quality by leading educators to "teach to the test" which results in a narrowing of the curriculum (Nevi, 2002). In addition, subjects not included in assessment are being shortchanged or even eliminated (Jones, Jones & Hargrove, 2003). Further, content not covered in some state standards may be neglected by educators (Sheppard, 2002). A frequent criticism of high-stakes testing movement is that high-stakes tests are driving curricular decisions. In fact, many school districts are accused of teaching to the test rather than providing an array of curricula inherent in a quality education. Test-driven curricula promote teacher-directed learning and preclude enrichment, student-centered learning, and higher order thinking (Jones, Jones & Hargrove, 2003; Passman, 2000). A survey of teachers in North Carolina revealed that they devoted 20% more time to the subjects or the basics (reading, writing, and mathematics) that were on the state tests (Jones, Jones & Hargrove, 2003). Nontested subjects are significantly reduced or eliminated as educators focus on tested subjects (Jones, Jones & Hargrove, 2003). Further, according to Nichols and Berliner (2007), pressure exerted from the need to succeed on high-stakes test may lead to inappropriate test practices including cheating on tests (p. 33). In addition, they found the pressure from these tests can lead to coaching students to pass the test by focusing on aspects of the test that are incidental to the curricula the test is intended to represent.

Nichols and Berliner (2007) reviewed numerous instances of adults cheating in regard to high-stakes testing. In fact, they list several types of cheating including pretest, during-the test, and post-test cheating. Pretest cheating includes providing students and or with actual exam questions, similar versions of test questions, a "peek" at the test, or

simply finding an excuse to exclude students from the test. During-the test cheating includes the use of “tip sheets,” coaching students as they take the exams, and whispering answers, and prompting children to change answers with verbal or gestured prompts. Post-test cheating involves changing answers from wrong to right and purposely using incorrect identification criteria resulting in low scores being “thrown out.” Nichols and Berliner also chronicle examples of students cheating on tests in response to the high-stakes attached to them.

Critics support the position that educational decisions based on high-stakes tests have a disproportionate negative impact on poor and minority children (McNeil, 2000b). According to Kohn (2001), narrowly defined standards do the most disservice to those student populations who have traditionally been disenfranchised. When used inappropriately, high-stakes tests can undermine the quality of education and reduce opportunities for some students especially if results are misinterpreted or misused, or students are relegated to a low quality educational experience as a result of their scores (Heubert & Hauser, 1999).

Researchers have also raised concerns about the impact high-stakes tests have on students who have a primary language other than English. Furthermore, critics believe that high-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language (Wright, 2006). In fact, Wright claims that testing limited English proficiency (LEP) students “defies logic.” He points out the lack of ability to read, write, or understand English may prevent success on state exams while schools are simultaneously punished for these students lack of success. Bolton (2000) held that “standardized tests share assumptions about language and cultural skills,”

further, “the performance on almost any test is strongly influenced by language skills” (p. 4). Hilliard (2000) and Neill and Medina (1989) suggested that in addition to standardized tests being unfair to learners for whom English is not their first language, these tests erroneously assess students and gauge their development without taking into account their English deficiency. In fact, students who speak in ethnic or regional dialects are often penalized on the tests (Townsend, 2002). In addition, the lack of context for the questions on many tests often disadvantages those from culturally and linguistically diverse backgrounds (Falk, 2002).

In addition, critics fear that high-stakes testing and the accompanying consequences of failure lead to overstressed students. In Texas, the beginning of high-stakes testing for elementary students began in the 2002-2003 school year. Haney, Fowler, Wheelock, Bebell and Malec (1999) investigated the degree to which external tests motivated students to learn by examining the self-portraits of students in testing situations. Their findings were that young students depicted themselves as anxious, angry, bored, pessimistic, and withdrawn from high-stakes tests while older students were typically more disillusioned and hostile toward tests than were younger learners. Critics fear that the high-stakes testing movement will result in a dramatic increase in student dropout rates.

A study conducted by the Center for Labor Market Studies at Northeastern University shows that the nation’s high school dropout rate may be as high as 30 %, almost three times higher than government estimates. Fassold (2000) found that the average black and Hispanic student was three times more likely to dropout, even controlling for socio-economic status, academic track, language program participation,

and school quality. His research findings suggest that because of required high-stakes testing some 40,000 of Texas' 1993 sophomores dropped out of school. The dropout rates for black, Hispanic, and white students were about 25 %, 23 %, and 13 % respectively. Research conducted by Clarke, Haney and Madaus (2000) compared the 10 states with highest dropout rate to the 10 states with the lowest dropout rate. They found a high correlation to dropout rates and attrition with states that used high-stakes tests; none of the 10 states with the lowest dropout rates conducted high-stakes testing. Data from the National Educational Longitudinal Study indicated that high-stakes testing was not associated with improved scores but was associated with higher dropout rates (Jacob, 2001). Brian Jacobs of the University of Chicago's Harris Graduate School of Public Policy Studies determined that mandatory high school graduation exams increase the probability that low-achieving students will dropout. Students in states that use these mandatory tests are 25 % more likely to drop out of high school than their peers in states that did not employ exit exams, Jacobs reported. Sean F. Reardon and Claudia Galindo of Pennsylvania State University reported even stronger evidence in a paper prepared for the April 2002 meeting of the American Educational Research Association. In the two years between eighth and 10th grade, they found the odds of dropping out of school are 39 % greater for students in schools with high-stakes tests in place than for those in schools without such assessment.

McNeil, Coppola, Radigan and Heilig (2008) state that 135,000 students dropout of Texas schools each year; with dropout rates highest for African American and Latino students. Further, these authors purport that one harmful result of the high-stakes testing accountability movement is putting poor, English language learners, African American

and Latino students in risk of being pushed out of schools. In preface of her book, *Contradictions of School Reform*, McNeil (2000b) substantiates her stance that new forms of discrimination are arising out of the high-stakes standardized testing movement. Nichols and Berliner (2007) concur with her viewpoint, predicting that students from poverty, having special learning needs, and representatives from ethnically and linguistically diverse populations will most likely be high-stakes “score suppressors” and would be most likely to be denied an education (p. 64).

In this era of accountability, the pressures on principals and teachers to improve the scores of their students are overwhelming and at times debilitating (Chafin, 2004). Policymakers attribute low-test scores to administrators’ failure to direct teachers to induce achievement in their students (McNeil, 2000b). The fear of being identified publicly as a low performing school is real and may carry stiff consequences including loss of employment. In *Contradictions of School Reform*, McNeil (2000b) goes so far as to suggest that an “extraordinary culture of intimidation” is part of high-stakes testing movement (p. 269); silencing public criticism from educators. Both critics and Supporters point to the growing body of unexpected consequences of the high-stakes testing movement to support their opposing positions.

Unintended Consequences of High-stakes Testing

In a series of articles, Dr. Gregory Cizek of the University of North Carolina (2001a; 2001b; 2002) explored the unexpected consequences of the current testing movement. In their book entitled *The Unintended Consequences of High-stakes Testing*, authors Jones, Jones and Hargrove (2003) expounded upon the unexpected results of the

high-stakes testing movement. Cizek found that one result of the high-stakes testing movement is that educators know more about tests and testing than ever before (2001a). Driven by the demands of high-stakes tests, Cizek reported that professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise. Jones, Jones and Hargrove (2003) stated that high-stakes testing help focus teachers on teaching the mechanics of essential skills in reading, writing, and mathematics. For some teachers, this has resulted in improved instruction in these areas (Jones, Jones & Hargrove, 2003). High-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items (Cizek, 2002). He believed that the highest form of praise occurs when educators rely on these exemplars to enhance their own assessment practices. These authors also found unexpected consequences of high-stakes testing for students.

Cizek (2002) and Jones, Jones and Hargrove (2003) point out that the implementation of high-stakes tests has been a catalyst for increased attention to students with special needs. However, a study of educational reform in New York (Allington & McGill-Franzen, 1992) found that there was a significant increase in students being identified as handicapped as high-stakes testing increases. In addition, he believes that high-stakes tests promote greater homogeneity of education.

Cizek believes that a homogenization of education is occurring. As a direct result of schools aligning their curricula and instructional focus more closely to the standards exposed by high-stakes tests, the experiences of children in urban, suburban, and rural districts within a state are more comparable than they have been in the past.

Furthermore, Cizek discovered unexpected consequences of high-stakes testing in the assessments themselves and the data they generate (2001a; 2001b).

Cizek observed that high-stakes tests have evolved to a state of being: highly reliable, free from bias, relevant and age appropriate, higher order, tightly related to important public goals, time and cost efficient, and yielding remarkably consistent decisions (2002). This was particularly true in comparison to the tests developed by teachers for their own daily use. In addition, he found that prominent and public interest in high-stakes tests has resulted in an intensity of effort toward data collection and quality control that is unparalleled. Data is readily available for anyone with an interest in public education. Further, this high-stakes testing data is driving many education related decisions. In fact, testing program data has focused public attention to low-achieving schools and in some cases brought extra resources and staff development to schools that otherwise may have been ignored (Jones, Jones & Hargrove, 2003).

Historical precedents for high-stakes testing can be found as early as 2000 B.C. when Chinese officials conducted civil service exams and when Greek teachers used verbal evaluations as part of the learning process (Worthen & Sanders, 1973). In 1842, Henry Banard, Commissioner of Education in Connecticut, proposed what would have been the first state testing program (Mazzeo, 2000). While his proposal was ultimately unsuccessful, his fellow state school chief and friend Horace Mann of Massachusetts would play an influential role with the Boston School Committee to enact the groundbreaking Boston Survey. The Boston Survey was the first systematic assessment of academic achievement of public school students.

According to Bolton (2000), the first curriculum based high-stakes test for use in public school was produced and administered by the school superintendent of Portland Oregon. In 1874, eighth-grade students were given end-of-the-year written tests based on the superintendent's curriculum distributed to teachers prior to the start of school. True to the current trend, the test scores were printed in the paper and promotion was denied to those students who failed the test. Interestingly, an uprising of parents led to the dismissal of the superintendent and an end to these practices. With few exceptions, school-based standard testing is largely a product of the twentieth century.

Upon entering World War I in 1917, the United States had to quickly train a large number of soldiers, many who were illiterate and non-English speakers. A task ill suited for the Stanford-Binet which required one-on-one testing (Bolton, 2000). Motivated by a desire to establish his profession, psychologist Robert Yerkes brought together the major testing experts, including Goddard and Terman, to help design a suitable test (Gould, 1981). Yerkes, selecting a multiple-choice format developed by Arthur Otis (while Otis was a student of Terman), led the development of the Army Alpha and Army Beta tests to address the problems of efficiently testing large numbers of men and accurately test illiterate and non-English speakers (Hopkins, Stanley & Hopkins, 1990). The Army Alpha test's advantage over the Stanford-Binet was the ability of the test to be administered to a group, thus large numbers of men could be tested quickly (Bolton, 2000). The Army Beta test, designed for illiterate and non-English speakers, was a nonverbal test that did not require oral or written responses (Hopkins, Stanley & Hopkins; 1990). The Army Beta test was the first test to combine group and performance ideas (Hopkins, Stanley & Hopkins; 1990). These researchers

noted that the success of these tests lead to the development of similar tests for use in schools.

Perhaps no individual has been as unnoticeably involved in the escalation and maturation of standardized tests movement in the United States as Everett F. Lindquist. As a University of Iowa professor in 1928, Everett F. Lindquist, in support of scholarship competition, begins the Iowa Testing Program (Bolton, 2000). This norm-referenced test project, completed in 1931, was first published in 1935 as *The Iowa Every-Pupil Test of Basic Skills* under the direction of Lindquist. In that same year, *The Iowa Every-Pupil Tests* were extended downward to the elementary grades. These batteries were renamed *Iowa Test of Basic Skills (ITBS)* in 1955. According to the *Iowa Test of Basic Skills* website, ITBS originated the reporting of test results to elementary school parents and currently provides test for K – 8. E. F. Lindquist influence was not limited to public schools.

In 1947, Ralph Tyler, E. F. Lindquist, and others combined efforts to establish the Princeton based Educational Testing Service (ETS), dedicated to protect the public from poorly made tests or inappropriate use of tests (Parker, 1994). One of the major goals of the Center was to develop a model for statewide educational assessment. E. F. Lindquist and Ralph Tyler and other members of an army advisory committee modified the *Iowa Test of Educational Development*, also developed by Lindquist, to form the first *General Education Development (GED) Tests* (Bolton, 2000). Lindquist first administered the GED test in 1943 to veterans and soldiers on active duty. According to a government publication (National Library of Education, 1998), almost three quarters of

a million high school dropouts take the GED each year. E. F. Lindquist's also had an influence on the technology of testing.

In 1936, IBM developed a machine to score New York Regents examinations. While this was an important advance in testing technology, the machine required the use of the Markograph soft pencil electrical technology invented by Reynold Johnson. In 1956, electronic scanners developed by E. F. Lindquist and Albert Hieronymous were introduced for use (Mazzeo, 2000). The primary advancement was that these scanners allowed two-sided scoring sheets and did not require soft pencil markings. Thus, the administration and scoring of standardized exams became more efficient in terms of time, cost, and use. This use of this technology would soon be employed by all test makers, including the federal government. The proliferation of standardized test was not limited to public schools.

Universities and colleges began to use standardized test to regulate admissions. In 1925, Princeton professor Carl C. Bringham developed the Scholastic Aptitude Test (SAT) for the College Entrance Examination Board. In 1942, the College Entrance Examination Board replaced its traditional essay exams with a multiple-choice test. It was not until 1958 when the Educational Testing Service began disclosing SAT scores to test-takers. In 1959, Everett F. Lindquist and Theodore McCarrel founded the American College Test (ACT). The emergence of business ownership and their control of standardized testing industry, especially by schoolbook publishers, began early in the development of standardized tests and have continued through the present as consolidation, mergers, and acquisitions have strengthened their influence (Mazzeo, 2000).

Bolton (2000) chronicled the commercialization and consolidation of school-based standardized testing. Houghton Mifflin was the original publisher for the Stanford-Binet Test in 1916. World Book Company's entrée into the industry was their 1923 publication of the Stanford Achievement Test. In 1940, Houghton Mifflin acquired the rights to the Iowa Test of Basic Skills. 1960 was witness to Harcourt Brace and Company's acquisition of the World Book Company and its Stanford Achievement Test. In 1968, McGraw-Hill acquired the California Testing Bureau and its California Achievement Test series. In 1970, Harcourt Brace and Co. acquired The Psychological Corporation, originally founded by Edward L. Thorndike, and its Metropolitan Achievement Test series. Houghton Mifflin established its Riverside Publishing Division to publish the Iowa achievement test, Stanford-Binet test and other school-based standards tests in 1979.

According to Sacks (1999) the primary companies currently involved in the standardized tests industry include Harcourt (Brace) General Incorporated, Houghton Mifflin Company, National Computer Systems Incorporated (NCS), McGraw-Hill, and Educational Testing (ETS). Competition, relatively inexpensive tests, new testing technologies, emerging markets from state-mandated achievement testing programs, and federal legislation requiring testing have resulted in a proliferation of standardized tests in public schools at a level unseen prior to the current trend. According to Sacks (1999), Americans take nearly 400 million standardized tests yearly for educational purposes.

Over the past two decades, politicians, policy makers, educators, researchers and others have asserted that prior education reforms have not been successful in raising the academic performance of all students, particularly minority students. Teachers,

instruction methods, administrators, and universities have each been criticized for the failure of reforms to elevate student performance to acceptable levels (deFur, 2002). Increased employer demands for a more literate workforce along with public expectations for increased student achievement have fueled an ardent call for accountability in education.

The 1970s were witness to the advent of minimum standards tests, which called for a back-to-basics approach. Policymakers believed a minimum competency movement would reform our schools (Bracey, 1995; Heubert & Hauser, 1999). In order to ensure that all students would master the basics, states began to rely on tests of minimal basic skills. In the 1980s, the minimum competency test movement lost momentum.

Critics of the minimum competency test movement claim that these test promoted lower standards through the acceptance of these minimal standards as acceptable for all students (Bracey, 1995). Further, it was perceived that due to poor student performance on standardized tests the minimum competency test movement was actually lowering the content taught in schools (Lutz & Merz, 1992). *A Nation at Risk* (NCEE, 1983) signaled the end of minimum competency test movement and the beginning of high-stakes competency testing movement (Amrein & Berliner, 2003; Lutz & Merz, 1992). *A Nation at Risk* recommended the use of rigorous standards accompanied with student and school accountability through assessment to regiment and improve curricula, teachers, schools, and higher education. Thus, state-level accountability systems were developed. High-stakes test driven systems have four components: content standards that communicate essential knowledge and skills; tests

designed to measure student progress toward achieving the content standards; identifying criteria to determine if schools and students have reached the expected achievement levels; and incentives, such as sanctions and rewards for meeting performance targets (Madaus & Clarke, 2001).

Overview of *A Nation at Risk*

On August 26, 1981, based on Secretary of Education T. H. Bell's concerns that "something is remiss in our education system" (p. 4), the National Commission on Excellence in Education was formed (NCEE, 1983). According to Tyack (1990), reform periods in education typically occur when a discovery of a problem such as economic concern provide policymakers with the justification for educational reform. Led by its chairman, University of Utah President David Pierpont Gardner, the Commission's assigned task was to examine the quality of education in the United States and report their findings, including practical recommendations for educational improvement, within an 18-month period. Findings and recommendations were transmitted in a report entitled *A Nation at Risk: The Imperative for Educational Reform*. The Commission's charter directed its members to pay particular attention to teenage youth; consequently, their primary focus was on high schools.

The Commission's charter contained a litany of specific charges. The charges of interest to this study include: assessing the quality of teaching and learning in United States' public and private schools, colleges, and universities; conducting comparisons of United States schools to those of other nations; evaluating the relationship between student achievement in high school and college admissions; analyzing the impact of

major social and educational changes and their effects on student achievement; and delineating the obstacles which must be overcome in order to achieve educational excellence. The Commission would draw on a variety of sources to meet the call of their charter.

In 1983, the findings of the NCEE were released in “An Open Letter to the American People.” The premise of the report is that we have lost our preeminence in commerce, industry, science, and technological innovation to other nations. Further, it asserts that the historical accomplishments of our schools have been “eroded by a rising tide of mediocrity that threatens our very future as a Nation and a people” (p. 6). The scathing language of the report continued by alleging “If an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war.” While crises in industry and commerce were the paramount concerns raised in the Commission’s report, they also were concerned about “the intellectual, moral, and spiritual strengths of our people which knit together the very fabric of our society” (p. 7). Justifications of the preceding caustic statements were supported by a number of statements labeled as indicators of risk.

Indicators of risk listed numerous examples of document testimony which criticized the literacy rates of our population, the declines in standardized test scores such as SAT and ACT, the dismal performance of our secondary students on international achievement tests, the increasing need for remedial coursework for incoming college students, and the complaints of business and military leaders about the lack of basic skills of recent graduates. Further criticism alleges that schools are creating

a generation of scientifically and technologically illiterate students. Additional disparagement of schools were found in purported studies that claim “many schools emphasize such rudiments as reading and computation at the expense of other essential skills such as comprehension, analysis, solving problems, and drawing conclusions.” The origins of the preceding comments originate from the testimony at the Commission’s numerous fact-finding efforts.

The Commission utilized existing analysis of education while commissioning papers from experts on a variety of issues. Further, testimony from a wide range of interested parties were garnered at eight full Commission meetings, six public hearings, two panel discussions, a symposium, and a series of regional meetings hosted by the Department of Education. Testimony was received from issue experts, high school and college students, parents, educators from all levels, school board members, leaders of industry, leaders of United States armed forces, representatives of minority groups, and state officials. The culmination of the efforts of the commission resulted in numerous findings and a series of recommendations for educational reform.

A major perception communicated during these hearings was that an increasing number of high school students are graduating without the skills necessary for success in college or the workplace. Further, the Commission reported that testimony recognized a pressing need for improved instruction in mathematics, science, English, all facets of social studies, and economics. Further, the Commission said, “We have come to understand that the public will demand that educational and political leaders act forcefully and effectively on these issues” (p. 11). To meet the demands expressed in their findings, the Commissions set several goals.

The report presented its meaning of excellence in regard to individuals, schools, and our society; and set achieving “excellence in education” (p. 11) at every level of our nation as its primary goal. Another goal was to enact educational reform that will result in the development of a “Learning Society.” The cornerstone of a Learning Society is the development of lifelong learners with the ability to adapt to a rapidly changing world-based workplace. The key characteristics for a Learning Society include a set of values, high expectations, and continuing pursuit of education that enhances the quality of life as well as an economic and occupational impact. The effective utilization of the multitude of tools and abundant resources available in America is essential in order to reach the Commission’s goals and recommendations. The findings of the committee called for many changes in how we conduct education.

The first finding criticized the curriculum found at many United States schools replete with too many electives and shortage of required core coursework. Students have moved away from college and vocational tracks into a less challenging general track. The next finding calls for an increase in the expectations by and for our students.

Students, parents, and educators must demand a level of knowledge, abilities, and skills school and college graduates should possess to demonstrate mastery of core curriculum. Further, we must expect student to demonstrate traits for success such as proficiency with time, hard work, behavior, self-discipline, and motivation that are essential for high student achievement. The report found that are expectations are clearly communicated by our approach to grades, graduation requirements, presence or absence of rigorous exams, difficulty of assignments, the quality and quantity of homework, high school graduation requirements, and college entrance requirements.

Further, the report decries the use of minimum competency exams and suggests they actually lower educational standards. The next finding evaluates how effectively time is used in United States schools.

The report made derogatory remarks concerning the use of time in our schools. Their findings suggested that in comparison to other nations, American students spend much less time on schoolwork, homework, daily length of instruction, and the number of days of instruction. In addition, the report found that students are not taught disciplined and systematic study habits. The quality of United States teachers was also questioned.

The report stated that the teaching force was made up of too many teachers with limited academic capabilities. Further, teacher preparation programs were criticized for the quality of their product as well as the curriculum that is short in subject matter and an overemphasis in educational methods. The report rebuked the low pay provided by states for its teachers. The report recognized the shortage of science, math, special education, and foreign language teachers. In addition, it criticized the number of unqualified teachers found in these areas. The sum of the preceding findings was used to prepare the recommendations of the Commission.

The first Commission recommendation implored states to strengthen curriculum in the “Five New Basics” (p. 19), English, science, mathematics, social studies, and computer science. Further, two years of foreign language was recommended for college-bound students. Specific curriculum components were delineated for each of the five areas. Further, colleges and universities were implored to raise their admission requirements and corresponding levels of achievement on standardized tests in each of these five areas.

Recommendations involving the use of time included a call for more homework, seven-hour school days, 200-220 days of instruction per year and additional time to meet the needs of unconventional students. Further, the development of firm but fair code of student conduct would help teachers maintain discipline. In addition, incentives and sanctions should be employed to reduce student absenteeism and tardiness.

Recommendations for teachers included requiring high educational standards for prospective teachers, increasing salaries to competitive levels, include performance-based evaluations that include peer review, and provide incentives to attract outstanding college students to the teaching profession. The report also called for the development of career ladders for teachers who distinguish themselves. The report acknowledges the crucial leadership role that principals and superintendents must provide if reform is to occur.

The recommendation of primary interest to this study was the call for the administration of standardized achievement tests at major transition points from one level of schooling to another. Of particular interest was the call for an exit test from high school to college or work. The report recommended that the tests would be administered as part of a nationwide system of state and local standardized tests. The system should include diagnostic procedures that allow teachers, students and parents to evaluate students' progress and provide remedial intervention to those in need. The Commission called for the movement away from minimum competency exams to standardized achievement tests to check the "credentials" (p. 19) of students with consequences for those who fall short of expectations. Placement, promotion and graduation policies should be guided by academic progress.

The report produced letters to parents and students imploring them to join school and college efforts toward reform. Parents were encouraged to remain vigilant and demand the best our schools can provide. Students were beseeched to work with dedication and self-discipline in order to create and control their destiny. Finally, the report implies that the success of educational reform will determine if “America’s place in the world will be either secured or forfeited” (p. 25). Although researchers widely accepted that the report made erroneous claims concerning the decline of current student academic achievement in comparison to past generations, the impact of this report cannot be overlooked (Wong & Nicotera, 2007).

Impact of *A Nation at Risk*

In 1983, accompanied by unprecedented hype, the National Commission on Excellence in Education (NCEE) released *A Nation at Risk: The Imperative for Educational Reform*. This report was released during a period of double-digit inflation, record high inflation, and severe recession (Sacks, 2000). In a paper presented at the Annual Meeting of the American Educational Research Association, Kristen Lanier was quoted saying “*A Nation At Risk* is a unique form of lament when it was published, not only did the political maneuvering of many parties bring the report to greater attention than reports on education normally receive, but the structure, rhetorical tone, and fervor of the report, with its suggestions of a nation fallen from grace, gripped the national soul as though it were a sermon” (Lanier, 2000). In the words of T. H. Bell, Secretary of Education, *A Nation at Risk* put educational reform “on everyone’s front burner” (Perkinson, 1995, p. 190).

This thirty-two-page report is considered by many to be watershed of the current standards-based high-stakes assessment and accountability movement, stimulating state and federal educational reform of the past two decades (Bracey, 2002, Vornberg, 1991). In his book *Standardized Minds*, Sacks (1999) declared that “The impact of *A Nation at Risk*, even twenty years later, on the politics of American schools can’t be overstated” (p. 77). Further, Sacks found that this landmark report “would become a veritable New Testament for the modern-day accountability movement.” Jones, Jones and Hargrove (2003) attribute the popularity of high-stakes testing to *A Nation at Risk* and credit it with changing the public’s attitude towards testing. The dramatic influence of this report can be traced to several factors.

The prestige and political power inherent in the White House combined with an accompanying media blitzkrieg produced a catalyst for educational reform that continues to impact schools today. Endorsed by President Ronald Reagan, *A Nation at Risk* made claims about the failure of American education, chronicled of lack of success by American students on international tests, suggested we have a lack of talent and motivation among American educators, and provided evidence to back their claims (Berliner & Biddle, 1995). In an interesting paradox, *A Nation at Risk* saved the Department of Education. President Reagan had promised to abolish the Department of Education but kept the institution when he realized its potential to carry out the reforms raised by *A Nation at Risk* (Perkinson, 1995). This report “galvanized the fledgling accountability, transforming it into a national project” (Sacks, 1999, p. 77).

While *A Nation at Risk* praised the historical accomplishments of schools and colleges, it claimed the American education system has squandered prior progress and

has slipped into “a rising tide of mediocrity” (NCEE, 1983, p. 1). To back these claims, the NCEE presented evidence in the section entitled “Indicators of Risk” (1983, p. 8). Sacks (1999) found that “fully nine of thirteen indicators of the risk assembled by the NCCE pertain to a standardized test of some sort” (NCEE, 1983, p. 76). While the report’s evidence has been subsequently criticized by numerous critics, including the Sandia Report, the next decade witnessed an embracing of its findings by American Presidents, secretaries of education, federal agencies, leaders in industry, state educational leaders, and community leaders (Amrein & Berliner, 2002; Berliner & Biddle, 1995).

Influenced by *A Nation at Risk*, governors became concerned with restructuring school to address the educational problems in their states (Sadker & Sadker, 2000). Under gubernatorial leadership in many states, policies were formulated and legislated that educational accountability would extend beyond the student to include the teacher, school, district, and state. Wong and Nicotera (2007) believe that the shift to performance-based accountability recommended by *A Nation at Risk* has significantly altered the goals and functioning of the public education system by switching the focus to results driven accountability. The dramatic impact of *A Nation at Risk* is best demonstrated in a 1985 Education Week article which reported an unprecedented number of state reform laws including: an increase in 43 states raising graduation requirements; exit tests for 15 states; upgraded teacher requirements for 29 states; increase in teacher’s salaries for 18 states; and 37 instituted statewide assessment for its students (Perkinson, 1995). Hoffman (2001) pointed out that prior to this time less than a dozen states required standardized testing of their students; even fewer required high-

stakes tests for promotion or graduation. Thus, *A Nation at Risk's* legacy is that testing and accountability “became the essence of the meaning of school reform” (Sacks, 1999). Not to be outdone, the business interests seized on this momentum to bring their thoughts on our schools and the caliber of students they produce.

In May of 1990, a committee was formed to conduct a comprehensive study on how well schools prepare young people for the workforce. This effort was organized by the United States Department of Labor and initiated by the former Secretary of Labor, Lynn Martin. Titled the Secretary’s Commission on Achieving Necessary Skills (SCANS), this extensive work was momentous in that it was the first time American business was given a platform to clearly communicate to educators what students need to know in order to be successful in the workplace. The SCANS (1990) Report goes on to emphasize five core subject areas—history, geography, science, English, and mathematics, and it states they should be taught and applied within a framework of five new competencies and a three part foundation which highlights and extends basic skills. The SCANS Report stated that workers must be able to effectively use five new competencies of Resources Allocation, Interpersonal Skills, Information Acquisition and Evaluation, Systems Management, and Technology Utilization. The three-part foundation included Basic Skills, Thinking Skills, and Personal Qualities. The SCANS Report added credence and focus to the accountability movement.

***A Nation at Risk's* Effect on Texas**

Over the past two decades, the state of Texas has been at the forefront of education reform efforts to implement new standards coupled with high-stakes

assessments. According to McNeil (2000a), Texas is the model for the federal policy governing our nation's school. This trend began in earnest in 1984. Using the impetus created by the *A Nation at Risk* report, the Texas legislature passed comprehensive education reform laws mandating sweeping changes in education in Texas. These laws mandated that all public schools follow a state-mandated curriculum called "essential elements" and mandated the Texas Educational Assessment of Minimum Skills (TEAMS). Implemented in 1985, TEAMS was a basic skills math and English language arts tests of students administered in odd numbered grades. High school students were required to pass the "exit level" version of TEAMS, administered in the eleventh grade, in order to receive a diploma. High-stakes testing was only the first recommendation from *A Nation at Risk* to be adhered to by Texas policymakers. The state created career ladders, using identical terminology in *A Nation at Risk* (NCEE, 1983, p. 16) to recognize and financially reward excellence in teaching. Additionally, in response to the call for higher salaries and standards for teachers, Texas increased its educator pay but required teachers to take the Texas Examination of Current Administrators and Teachers (TECAT), a high-stakes test, which required teachers to obtain a passing score in order to keep their Texas teaching certificate. In addition, schools were required to produce and publish a student code of conduct to increase school safety. These documents, still in use today, list offenses and a range of consequences for each offense. Arguably, no other state embraced the recommendations of *A Nation at Risk* more wholeheartedly than Texas. Initially, only Texas has linked teacher evaluations to student and school results, but additional states are planning to do so in the future (Jones, Jones, & Hargrove, 2003).

Another impression on education as a direct result of *A Nation at Risk* was the augmentation to centralization of state authority. According to Tyack & Timar (1999), the impetus of *A Nation at Risk* led “educators to shed the concept of education for civic virtue and emphasized economic growth, productivity, and efficiency.” Thus, curriculum changes and corresponding curriculum standards were required to determine if schools were meeting these new goals of education. The uses of explicit curriculum standards that are enforced using high-stakes standardized tests are increasing state government’s influence and control of local schools (Klein & Hamilton, 2001). This supposition is supported by Lutz and Merz (1992), “The reforms of the 1980s increased standardization of curriculum, centralization of state authority, and a drive for accountability” (p. 29). This trend has been considerably strengthened by the No Child Left Behind Act. Further, Jones, Jones and Hargrove’s (2003) report that high-stakes testing programs also result in massive amounts of test preparation. Test preparation initiates an entire chain reaction of other negative consequences including unethical item teaching, cheating, student anxiety, and loss of instructional time.

A review of the literature reveals that state-mandated testing is explicitly a political phenomenon (Parker, 1994, Sacks, 2000). Mazzeo (2000) found that political symbolism and mechanism of control are the two primary political motivations for the utilization of high-stakes assessment. Political symbolism occurs when policymakers want to assure their electorate they are taking care of business, thus they are concerned about image and typically have little impact on the educational system. On the other hand, mechanism of control is an attempt to impact education through legislation and policy. States use state-mandated testing to control the curriculum and familiarize

teachers and administrators to performance standards and shifts authority away from local schools and school systems (Mazzeo, 2000). The mechanism of control as recommended by *A Nation at Risk* is evident in all high-stakes testing states, particularly in Texas. Consequently, accountability systems are a product of the establishment of standards of performance for individual students, schools, and school districts mandated by state governments that are often under the influence of federal mandates. Without a doubt, the impetus created by this trend has inspired legislation at both the state and national level.

No Child Left Behind Act

Improving education was one of the cornerstones of Governor Bush's platform for his initial run for the presidency. Borrowing the "leave no child behind" mantra, from Marian Wright Edelman of the Children's Defense Fund, George Bush promised to improve the nation's schools if elected. He touted the improvement of education in Texas schools resulting from the use of high-stakes tests especially the narrowing of the gap between the scores of white and minority students. After his successful election, President Bush used his influence to encourage Congress to pass educational reform legislation. A bipartisan effort resulted in The No Child Left Behind Act.

On January 8, 2002, President Bush signed into law the No Child Left Behind Act of 2001, reauthorizing the Elementary and Secondary Education Act (ESEA). Far from simply reauthorizing ESEA, the No Child Left Behind (NCLB) Act has expanded the federal role in education and set requirements in place that affect every public school in America. This act created a system of educational evaluation of all public schools.

The federal No Child Left Behind Act is having a tremendous impact on public schools, particularly schools that utilize Title I funds. Title I director Dr. Caryl Burns stated the “Key requirements of the law are closing the achievement gaps, holding schools accountable for all students performing at a high level, and having qualified teachers in every classroom” (Shuford, 2004). The 40th Annual Phi Delta Kappa/Gallup Poll (Bushaw & Gallup, 2008) found that the NCLB Act represents the greatest federal incursion into K-12 education to date. NCLB represents a major new departure from a long history of local-based control over key education decisions.

According to Nichols and Berliner (2007), NCLB is the reason for the present spread of high-stakes testing. For the first time in history, the federal government has set requirements that beginning with the 2005-06 school year, states test all students be tested in math and reading annually in grades 3 through 8 and at least once in grades 9 through 12. Schools that meet the 37 required criteria are labeled as meeting Acceptable Yearly Progress (AYP); those that do not are labeled Not Meeting AYP. AYP refers to the minimum level of improvement that states, school districts and schools must achieve each year as they progress toward the ESEA goal of having all students reaching the proficient level on state tests by 2014. Thus, the federal government is mandating high-stakes standardized tests for all United States students (Nichols & Berliner, 2007). Schools failing to meet AYP risk not receiving Title I funds and other sanctions. Before NCLB, many schools systems only concerned themselves with average scores, thus gaps in achievement between ethnic, income and disability subgroups was of limited concern. As a result of NCLB, districts must pay attention to the achievement gaps of these subgroups.

While the No Child Left Behind Act makes significant changes to raise academic standards, increase student testing and provide information to parents and communities the law also imposes new sanctions on schools based on how students perform on state tests. While this Act may provide assistance to schools that fall behind, it also levies sanctions such as allowing students to transfer to other schools, funding private tutoring programs, and shifting control of local schools to the district, state, or private contractors. The authors of this legislation are certain that setting high academic standards for students, testing students on these standards, and holding schools and educators responsible for reaching those standards will significantly improve public education in American schools.

According to Jones, Jones and Hargrove (2003), the shift in control of what is taught, of how it is taught, and who gets high-quality instruction is perhaps the most severe consequence of NCLB and the accountability movement for the education community. Whereas states once provided only curriculum frameworks and outlines, they are now dictating the content of instruction. This shift in control from local communities to policy makers at the state and national levels has quietly occurred with little discussion or recognition (Jones, Jones & Hargrove, 2003). The Supporters of educational reform using high-stakes testing, such as our current President George W. Bush and other legislators at both the state and national level, continue to shift the loci of control away from local educators and boards and towards the control of the standards to state and federal policymakers. The case for standards-based accountability has been supported by a host of powerful voices ranging from Louis Gerstner, Jr., CEO at IBM to

Ronald Wolk, editor of Education Week, to the late Albert Shanker, formerly president of the American Federation of Teachers and others.

.

CHAPTER III

METHODOLOGY

Population

The population selected for this study was the 2005-2006 membership of Texas Association of Secondary School Principals (TASSP). The Texas Association of Secondary School Principals is an association formed by and for over 4000 campus level administrators at the middle and high school level. The 2005-2006 membership list was provided to the researcher in spreadsheet form alphabetized by the name of the administrators' school. Each member was assigned a number in order starting with 1 through 4,641. The researcher used the random number generator program developed by Scott Donato Saccenti. Once the researcher inputted the range of potential responders, 1 – 4,641, the program produced random numbers within the range. The researcher used the program to produce 600 random numbers. The generation of random number in excess of 400 was completed to handle the possible duplication of numbers and address the issue of TASSP members without recorded email addresses. The researcher designated each potential responder by matching his or her number with the number selected by the random number generator. A sample size of 400 was selected to ensure that the sample was allocated proportionally to be representative of the total population.

The researcher used Survey Monkey to electronically submit the questionnaire, submit each follow-up opportunity, and gather responses. Furthermore, the researcher

sent personal email messages on the final submission informing potential responders that this was the final opportunity to participate in the research. This was an attempt to insure that spam filters had not prevented potential responders from participation as well as helping motivate procrastinators. Despite four opportunities to respond to the survey, only 178 of the targeted 400 completed the questionnaire. Respondents were given the opportunity to decline participation in the questionnaire. 17 potential respondents took that option.

Table 3.1 is a summary of response rates resulting during each round of the survey. Seventy-one secondary administrators responded to the initial opportunity to answer the survey. Forty-seven secondary administrators responded to the second opportunity to answer the survey. Forty-five secondary administrators responded to the second opportunity to answer the survey. Only 15 secondary administrators responded to the fourth and final. Thus, 39.9% of secondary administrators responded in the initial round, 26.4% during the second round, 25.3% during the third round, and 8.4% in the fourth and final round.

Table 3.1—Summary of the First Round, Second Round, Third Round and Fourth Round Survey Response Rates

Mailing	Responses	Percent	Cumulative Percent
First Round	71	39.9	39.9
Second Round	47	26.4	66.3
Third Round	45	25.3	91.6
Fourth Round	15	8.4	100.0
Total	178	100.0	

Table 3.2 shows the gender breakdown and corresponding percentages of the respondents. Of the 178 survey respondents, 110 were male (61.8%) and 68 were female (38.2%).

Table 3.2—Frequency Distribution of Gender as Reported by Survey Respondents

Gender	Frequency	Percent	Cumulative Percent
Male	110	61.8	61.8
Female	68	38.2	100.0
Total	178	100.0	

Table 3.3 shows frequencies and corresponding percentages of the years of administrative experience as reported by the respondents. Of the 178 survey respondents, 54 (30.3%) respondents have 4 or fewer years of administrative experience, 75 (54.5%) respondents have 5 through 14 years of administrative experience, and 27 (15.2%) respondents have in excess of 14 years as of experience as an administrator.

Table 3.3—Population Strata Based upon Respondents' Years of Administrative Experience

Administrative Experience	Frequency	Percent	Cumulative Percent
1 – 4 Years	54	30.3	30.3
5 – 14 Years	97	54.5	84.8
15 or More Years	27	15.2	100.00
Total	178	100.0	

According to their own website, the University Interscholastic League (UIL) was created by The University of Texas at Austin to provide leadership and guidance to public school debate and athletic teachers. UIL member schools are divided into five conferences according to enrollment. Conferences, in order of enrollment, are AAAAA, AAAA, AAA, AA and A. During the 2006-2007 school year, Class A schools had enrollments of 194 students or less, Class 2A schools had enrollments of 195 through

414 students, Class 3A schools had enrollments of 415 through 949 students. The researcher grouped campuses classified as Class A, 2A, or 3A and designated these campuses as Small. The Large school group was made up of Class 4A and 5A schools. The enrollments of Class 4A schools ranged from 950 through 1,984 students, while 5A schools have campus populations in excess of 1,984 students. Table 3.2 shows that 63 of the respondents work on campuses designated by the researcher as Small while 115 respondents work on campuses labeled as Large. Schools in the Large category have resources not available to the small. Large schools have increased role specialization resulting from multiple campus administrators, additional counselors, and potential support from larger central administrative staffs. In contrast, campuses classified as Small have much smaller campus and central administrative staffs, thus administrators in small schools tend have multiple and varied job duties. Despite the fact that Small districts are more numerous than Large districts, the pool of respondents is replete with more possible Large school administrators. Thus, it was expected that more Large school administrators would be sampled than Small school administrators. The results coincide with this premise as 115 (64.6%) of the respondents were classified in the Large classification while 63 (35.4%) were grouped in the Small classification as shown in Table 3.4.

Table 3.4—Population Strata Based upon Respondents’ Campus Classification

Classification	Frequency	Percent	Cumulative Percent
Small (A, AA, or AAA)	63	35.4	35.4
Large (AAAA & AAAAA)	115	64.6	84.8
Total	178	100.0	

In order to hold schools and school districts accountable for student learning, the Texas State Board of Education was mandated to rate the performance of schools and school districts according to a set of “academic excellence indicators.” The Academic Excellence Indicator System (AEIS) reported TAAS and TAKS results, dropout rates, and student attendance rates disaggregated by ethnicity and socioeconomic status. Texas high schools are rated as “exemplary,” “recognized,” “academically acceptable,” and “academically unacceptable.” An analysis of the population strata based upon respondent’s current campus rating found that 126 of the respondents (70.8%) work on campuses designated as Academically Acceptable. The second largest group consists of 40 (22.5%) respondents working on campuses that were Recognized. In contrast, only 6 (3.4%) of respondents working on Academically Unacceptable campuses responded. Likewise, only 3 (1.7%) respondents work on campuses receiving an Exemplary rating, the highest rating bestowed on Texas schools. Similarly, the same number and percentage of respondents work on campuses that were not rate, as shown in Table 3.5.

Table 3.5—Population Strata Based upon Respondents’ Current Campus Rating

Rating	Frequency	Percent	Cumulative Percent
Exemplary	3	1.7	1.7
Recognized	40	22.5	24.2
Academically Acceptable	126	70.8	94.9
Academically Unacceptable	6	3.4	98.3
Not Rated	3	1.7	100.0
Total	178	100.0	

Design of the Study

This investigation of the perceptions of secondary administrators on the effect state-mandated high-stakes testing has on the stakeholders of respondents home campus

is a descriptive study. This study was exploratory and was conducted during December 2006 and January of 2007.

Data were acquired from TASSP membership during the winter of 2006 through 2007 using a research instrument developed by the investigator. This study was designed in accordance to the parameters delineated by Gall, Borg and Gall (1996) in order to conduct a successful questionnaire. These steps included: defining the research objective, identifying a population, determining variables of the study, designing the instrument, pretesting and field testing the instrument, designing an introduction, and distributing the questionnaire with follow-up.

According to Gall, Borg and Gall's (1996) descriptive statistics are mathematical techniques for organizing and summarizing a set of numerical data. Mean is a measure of central tendency that is used to create a single numerical value that is used to describe the average of an entire set of scores. Standard deviation is the measure of variability and distribution in a set of numerical data. In other words, the standard deviation is a measure of the extent to which scores deviate from their mean. The mean and standard deviation, taken in conjunction, usually provide an accurate description of how members of a sample scored on a particular sample. This study used the standard deviation to make inferences because the relationship between standard deviation and the normal curve. In the normative curve 68% of the population is one standard deviation from the mean and 95% of the population is within 2 standard deviations from the mean.

Instrumentation

Educational Research: An Introduction (Gall, Borg & Gall, 1996) was the basis to developing the quantitative instrument used in this study. The research instrument was divided into four sections. The first section of the questionnaire elicited demographic information about the participants. The second section of the questionnaire elicited information concerning participants' perceptions of statements found in literature, which reflects the opinions of supporters of the use of high-stakes testing. The third section of the questionnaire elicited information concerning participants' perceptions of statements found in literature, which reflects the opinions of unbiased researchers' findings relating to the unintended consequences of the use of high-stakes testing. The fourth section of the questionnaire elicited information concerning participants' perceptions of statements found in literature, which reflects the opinions of critics of the use of high-stakes testing. Responses to all questions in the second, third and fourth sections of the questionnaire were on a five point Likert scale signifying 1) strongly agree, 2) agree, 3) uncertain, 4) disagree, 5) strongly disagree. A panel of experts will be used to establish the Content Validity. A copy of the instrument is found in Appendix A.

The research instrument was pretested by a group of five secondary administrators from other districts to assess meaning, wording, and other validity matters. Validity of the questionnaire was also addressed through review of the literature. Questions were developed based on factors of high-stakes testing as identified in a review of the literature. A field test using 10 secondary administrators from

neighboring school districts was conducted in Fall 2006 to assure clarity and content validity. Reliability was determined by calculating the alpha reliability of the questionnaire. Reliability analysis produced an alpha of .8762.

Procedures

The research instrument was electronically submitted to each of the selected administrators. The electronic survey included an introduction (Appendix B) explaining the purpose of the survey instrument and instructions for completion and return of the survey. Participants were told that the instrument could be completed in approximately 20 minutes or less.

After one week, follow-up was conducted as warranted by the return rate. Following the second electronic follow up, a third electronic follow up was sent preceded by a personal e-mail to each non-responder. At the end of January 2007, the data was compiled and a statistical analysis of the data was begun.

Data Analysis

Quantitative data was obtained and analyzed using *Statistical Package for Social Sciences* (SPSS). SPSS is a computerized statistical package that integrates data storage, retrieval and modification, and report writing. Results of the study were reported using numerical and graphical techniques to report such inferential statistical such as means, standard deviations, frequencies, and percentages. In addition, t-tests were conducted on all statements for each research questions. Independent t-tests were conducted to assess

whether the means of two independent groups are statistically different from each other. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years vs. 15 or More Years), Campus Classification (Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable). When comparing two independent groups, their variances must be relatively similar. Levene’s Test for Equality of Variance was used to check for this. If the significance for Levene’s test was 0.05 or below, then the “Equal Variances Not Assumed” test was used to determine significance. Otherwise, the use of the “Equal Variances Assumed” test was used to determine significance. An alpha level of .05 was used to establish significance.

Research Questions

Research Question 1

The question “Do perceptions of high-stakes testing supporters differ from those of Texas secondary school administrators in the membership of TASSP?” was answered by analysis of questionnaire responses to item numbers 1 through 12.

Research Question 2

The question “Do perceptions of researchers of high-stakes testing reported in current literature differ from those of Texas secondary school administrators in the membership of TASSP?” was answered by analysis of questionnaire responses to item numbers 13 through 19.

Research Question 3

The question “Do perceptions of high-stakes testing critics differ from those of Texas secondary school administrators in the membership of TASSP?” was answered by analysis of questionnaire responses to item numbers 20 through 31.

CHAPTER IV

RESEARCH FINDINGS

The purpose of this study was to assess the impact of the high stakes standardized test movement in Texas secondary schools. The assessment compares the perceptions between researchers, policy makers, and secondary school administrators. Secondary principals were asked in an electronic survey to rate the extent to which 31 statements about high-stakes tests matched their experiences on their campuses. Statements 1 through 12 match the assertions made by supporters of the high-stakes tests movement. Statement 1, high-stakes tests have helped focus public attention on schools with low-achieving students, was derived from NCLB (2002). Statement 2, high-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach, was derived from NCLB (2002). Statement 3, high-stakes tests have helped close the gap in achievement between minority students and majority students in Texas, was derived from Achieve, Inc. (2002). Statement 4, teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform, was derived from Amrein and Berliner (2002). Statement 5, doing poorly on high-stakes tests will lead to increased student effort to learn was derived from Amrein and Berliner (2002). Statement 6, students work harder and learn more because they know what is expected and that the high-stakes tests really count was derived from Amrein and Berliner (2002). Statement 7, the public display of high-stakes test scores motivates administrators to

ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught, was derived from Sacks (1999). Statement 8, when high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available, was derived from AERA (2000). Statement 9, administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs was derived from Amrein and Berliner (2002). Statement 10, educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success, was derived from Cizek (2002). Statement 11, driven by the demands of high-stakes tests, professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise, was derived from Cizek (2002). Statement 12, the implementation of high-stakes testing has been a catalyst for increased attention to students with special needs, was derived from deFur (2002).

Statements 13 through 19 are statements of unintended consequences of the current high-stakes standardized test movement. Statement 13, one result of high-stakes testing is that educators know more about testing than ever before, was derived from Cizek (2002). Statement 14, prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled, was derived from Cizek (2002). Statement 15, high-stakes tests promote greater homogeneity of education, was derived from Cizek (2002). Statement 16, high-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public

goals; time and cost efficient; and yielding remarkably consistent decisions, was derived from AERA (2000).

Statement 17, high-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has led to teachers enhancing their own assessment practices, was derived from Cizek (2002). Statement 18, high-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time, was derived from Haney (2000). Statement 19, high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels, was derived from McNeil (2000a).

Statements 20 through 31 match the assertions made by made by critics of the high-stakes standardized test movement. Statement 20, a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district, was derived from AERA (2000). Statement 21, high-stakes testing compromises educational quality by leading educators to ‘teach to the test,’ which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment, was derived from Schrag (2000). Statement 22, high-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems, was derived from Sacks (1999). For statement 23, a focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction, was derived

from McNeil (2000a). Statement 24, pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating, was derived from Cizek (2001b). Statement 25, high-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve, was derived from Haney (2000). Statement 26, Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children, was derived from Sacks (1999). Statement 27, high-stakes testing and the accompanying consequences of failure lead to overstressed students, was derived from Hancock (2001). Statement 28, the pressures inherent in preparing students for high-stakes tests are driving out good teachers, was derived from Haney (2000). Statement 29, high-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language, was derived from McNeil and Valenzuela (2002). Statement 30, the high stakes testing movement is resulting in a significant increase in student dropout rates, was derived from Clark, Haney, and Madaus (2000). Statement 31, no high stakes decision such as grade retention or graduation should be based on the results of a single test, was derived from Heubert and Hauser (1999) and the American Educational Research Association's position statement (2000).

This chapter includes the results of the electronic survey completed by a sample of secondary school administrators. The results of the survey are used to answer three specific questions:

1. Do perceptions of high-stakes testing supporters differ from those of Texas secondary school administrators in the membership of TASSP?

2. Do perceptions of researchers of high-stakes testing reported in current literature differ from those of Texas secondary school administrators in the membership of TASSP?
3. Do perceptions of high-stakes testing critics differ from those of Texas secondary school administrators in the membership of TASSP?

Research Question 1

Question 1: Do perceptions of high-stakes testing supporters differ from those of Texas secondary school administrators in the membership of TASSP?

To examine research question 1, means and standard deviations were calculated for statements 1 – 12 comparing to perceptions of high-stakes testing supporters and Texas secondary school administrators in the membership of TASSP. For all statements, the minimum was 1 (strongly agree) and the maximum was 5 (strongly disagree).

For statement 1, high-stakes tests have helped focus public attention on schools with low-achieving students, the mean was 2.47 and the standard deviation was 1.05.

For statement 2, high-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach, the mean was 2.89 and the standard deviation was 1.10.

For statement 3, high-stakes tests have helped close the gap in achievement between minority students and majority students in Texas, the mean was 3.06 and the standard deviation was 1.14.

For statement 4, teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform, the mean was 3.10 and the standard deviation was 1.21.

For statement 5, doing poorly on high-stakes tests will lead to increased student effort to learn, the mean was 3.94 and the standard deviation was .90.

For statement 6, students work harder and learn more because they know what is expected and that the high-stakes tests really count, the mean was 3.36 and the standard deviation was 1.10.

For statement 7, the public display of high-stakes test scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught, the mean was 2.16 and the standard deviation was .90.

For statement 8, when high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available, the mean was 2.88 and the standard deviation was 1.11.

For statement 9, administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs, the mean was 3.10 and the standard deviation was 1.17.

For statement 10, educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success, the mean was 2.15 and the standard deviation was .87.

For statement 11, driven by the demands of high-stakes tests, professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise, the mean was 2.60 and the standard deviation was 1.02.

For statement 12, the implementation of high-stakes testing has been a catalyst for increased attention to students with special needs, the mean was 2.63 and the standard deviation was 1.11. The means and standard deviations for statements 1 – 21 are summarized in Table 4.1.

For statement 1, high-stakes tests have helped focus public attention on schools with low-achieving students, one response, agree, yielded a simple consensus (50%). Congruent with the question, 64.6% of the respondents (115) found a level of agreement with the question. In contrast, only 40 (22.5%) disagreed and just 3 (1.7%) strongly disagreed with the question. This information is presented in Table 4.2.

On the survey statement 2, related to how well high-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach, administrators were not decisive in their opinions. A slight lean toward agreement with the statement occurred when precisely 89 (50%) agreed or strongly agreed with this position while 75 (42.1%) disagreed or strongly disagreed. This information is presented in Table 4.3.

Table 4.1—Means and Standard Deviations for Statements 1 – 12

	Statement	N	Min.	Max.	M	SD
1.	High-stakes tests have helped focus public attention on schools with low-achieving students.	178	1	5	1.47	1.05
2.	High-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach.	178	1	5	2.89	1.10
3.	High-stakes tests have helped close the gap in achievement between minority students and our majority students in Texas.	178	1	5	3.06	1.14
4.	Teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform.	178	1	5	3.10	1.21
5.	Doing poorly on high-stakes tests will lead to increased student effort to learn.	178	1	5	3.94	.90
6.	Students work harder and learn more because they know what is expected and that the high-stakes tests really count.	178	1	5	3.36	1.10
7.	The public display of high-stakes tests scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught.	178	1	5	2.16	.90
8.	When high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available.	178	1	5	2.88	1.11
9.	Administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs.	178	1	5	3.10	1.17
10.	Educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success.	178	1	5	2.15	.87
11.	Driven by the demands of high-stakes tests, Professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise.	178	1	5	2.60	1.02
12.	The implementation of high-stakes testing has been a catalyst for increased attention to students with special needs.	178	1	5	2.64	1.11

Table 4.2—Frequency Distribution of Responses to Statement Number 1: High-stakes Tests Have Helped Focus Public Attention on Schools with Low-achieving Students and, as a Result, Have Made These Students More Visible and Less Likely to Slip Between the Cracks and Fall Further Behind

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	26	14.6	14.6
Agree	89	50.0	64.6
Unsure	20	11.2	75.8
Disagree	40	22.5	98.3
Strongly Disagree	3	1.7	100.0
Total	178	100.0	

Table 4.3—Frequency Distribution of Responses to Statement Number 2: High-stakes Tests Are Designed and Implemented to Improve Instruction by Helping Teachers Focus on What Is Most Important to Teach

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	11	6.2	6.2
Agree	78	43.8	50.0
Unsure	14	7.9	57.9
Disagree	69	38.8	96.6
Strongly Disagree	6	3.4	100.0
Total	178	100.0	

Respondents reached a narrowly negative position on statement 3, high-stakes tests have helped close the gap in achievement between minority students and majority students in Texas. An indecisive outcome resulted from a failure of either position to reach a simple majority as 71 respondents (39.9%) reached a level of agreement with the statement while 81 respondents (45.5%) reached a level of disagreement with the statement. This information is presented in Table 4.4.

Table 4.4—Frequency Distribution of Responses to Statement Number 3: High-stakes Tests Have Helped Close the Gap in Achievement Between Minority Students and Majority Students in Texas

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	13	7.3	7.3
Agree	58	32.6	39.9
Unsure	26	14.6	54.5
Disagree	68	38.2	92.7
Strongly Disagree	13	7.3	100.0
Total	178	100.0	

A simple majority of the respondents (89), found a level of disagreement with statement 4, teachers need to be held accountable through high-stakes tests to motivate them to teach better. A degree of agreement with the preceding statement was reached by 74 (41.6%) respondents. This information is presented in Table 4.5.

Table 4.5—Frequency Distribution of Responses to Statement Number 4: Teachers Need to Be Held Accountable Through High-stakes Tests to Motivate Them to Teach Better, Particularly to Push the Least Motivated Ones to Perform

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	15	8.4	8.4
Agree	59	33.1	41.6
Unsure	15	8.4	50.0
Disagree	72	40.4	90.4
Strongly Disagree	17	9.6	100.0
Total	178	100.0	

Statement 5, the assertion that high-stakes test results motivate student's efforts to learn, produced a decidedly negative reaction among administrators surveyed. Surprisingly, 78.7% of the respondents (140) indicated they do not believe that doing poorly on high-stakes tests will lead to increased student effort to learn. Only 15 administrators representing just 8.4% of all respondent were in agreement with the statement. This information is presented in Table 4.6.

Table 4.6—Frequency Distribution of Responses to Statement Number 5: Doing Poorly on High-stakes Tests Will Lead to Increased Student Effort to Learn

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	3	1.7	1.7
Agree	12	6.7	8.4
Unsure	23	12.9	21.3
Disagree	95	53.4	74.7
Strongly Disagree	45	25.3	100.0
Total	178	100.0	

96 respondents (53.9%) found a level of disagreement with statement 6, the belief that students work harder and learn more because they know what is expected and that high-stakes tests count, while only 28.5% answered in the affirmative. This information is presented in Table 4.7.

Table 4.7—Frequency Distribution of Responses to Statement Number 6: Students Work Harder and Learn More Because They Know What Is Expected and That the High-stakes Tests Really Count

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	4	2.2	2.2
Agree	47	26.5	28.7
Unsure	31	17.4	46.1
Disagree	73	41.0	87.1
Strongly Disagree	23	12.9	100.0
Total	178	100.0	

Approximately 80% of respondents replied in the affirmative that public display of high-stakes test scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught. Only 12.9% disagreed to some extent on the same issue. This information is presented in Table 4.8.

Table 4.8—Frequency Distribution of Responses to Statement Number 7: The Public Display of High-stakes Test Scores Motivates Administrators to Ensure That Standards (TEKS) on Which the Tests Are Based Are Part of the Curriculum and Are Being Successfully Taught

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	32	18.0	18.0
Agree	110	61.8	79.8
Unsure	13	7.3	87.1
Disagree	21	11.8	98.9
Strongly Disagree	2	1.1	100.0
Total	178	100.0	

Neither position garnered a majority as only 48% of respondents support the position that when developed and used appropriately, high-stakes tests are among the most sound and objective knowledge and performance measures available. In contrast, only 32.6% could not support the statement, as a relatively large number of respondents, approximately 20%, were unsure about this topic. This information is presented in Table 4.9.

Table 4.9—Frequency Distribution of Responses to Statement Number 8: When High-stakes Tests Are Developed and Used Appropriately, They Are Among the Most Sound and Objective Knowledge and Performance Measures Available

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	10	5.6	5.6
Agree	75	42.1	47.8
Unsure	35	19.7	67.4
Disagree	42	23.6	91.0
Strongly Disagree	16	9.0	100.0
Total	178	100.0	

Respondents were practically evenly split since neither position was able to reach a majority position in response to the statement that administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs. A slight negative stance was found among respondents on the statement as 47.8% of the respondents found a level of disagreement with this issue. Correspondingly, 41.6% of respondents answered in the affirmative to this statement. This information is presented in Table 4.10.

Table 4.10—Frequency Distribution of Responses to Statement Number 9: Administrators Need to Be Held Accountable Through High-stakes Tests to Motivate Them to Be More Effective in Supervising Their Staffs

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	11	6.2	5.2
Agree	63	35.4	41.6
Unsure	19	10.7	52.2
Disagree	68	38.2	90.4
Strongly Disagree	17	9.6	100.0
Total	178	100.0	

An overwhelmingly affirmative response was reached in response to the position that educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success was reached by virtually 81% of the respondents. In contrast, only 12% found a level of disagreement with this position. This information is presented in Table 4.11.

Table 4.11—Frequency Distribution of Responses to Statement Number 10: Increasingly, from the Classroom to the School Board Room, Educators Are Making Use of Student Performance Data Generated by High-stakes Tests to Help Them Refine Programs, Channel Funding, and Identify Roots of Success

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	31	17.4	17.4
Agree	113	63.5	80.9
Unsure	13	7.3	88.2
Disagree	19	10.7	98.9
Strongly Disagree	2	1.1	100.0
Total	178	100.0	

Nearly two-thirds of respondents believe that the implementation of high-stakes testing has been a catalyst for increased attention to students with special needs. Only one quarter of respondents disagreed with this position. This information is presented in Table 4.12.

Table 4.12—Frequency Distribution of Responses to Statement Number 11: The Implementation of High-stakes Testing Has Been a Catalyst for Increased Attention to Students with Special Needs

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	31	14.6	14.6
Agree	89	50.0	64.6
Unsure	20	11.2	75.8
Disagree	40	22.5	98.3
Strongly Disagree	3	1.7	100.0
Total	178	100.0	

A majority of respondents, almost 62%, agreed to some degree with the position that high-stakes tests have resulted in improved professional development by focusing on helping educators hone his or her teaching skills and content area expertise. In contrast to the majority, only a quarter of the respondents disagreed with this statement. This information is presented in Table 4.13.

Table 4.13—Frequency Distribution of Responses to Statement Number 12: Driven by the Demands of High-stakes Tests, Professional Development Has Improved by Focusing on Helping Educators Hone His or Her Teaching Skills and Content Area Expertise

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	13	7.3	7.3
Agree	97	54.5	61.8
Unsure	23	12.9	74.7
Disagree	39	21.9	96.6
Strongly Disagree	6	3.4	100.0
Total	178	100.0	

To examine research question 1, four sets of independent t-tests were conducted to test for a difference between the means of two independent groups for statements 1 – 12. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years vs. 15 or More Years), Campus Classification (Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable).

Means, standard deviation, and N were calculated for campus classification. For statement 1, high-stakes tests have helped focus public attention on schools with low-achieving students, the mean for small school administrators was 2.56 and the standard deviation was 1.03 while the mean for large school administrators was 2.42 and the standard deviation was 1.06.

For statement 2, high-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach, the mean for small school administrators was 3.02 and the standard deviation was 1.04 while the mean for large school administrators was 2.83 and the standard deviation was 1.13.

For statement 3, high-stakes tests have helped close the gap in achievement between minority students and majority students in Texas, the mean for small school administrators was 2.87 and the standard deviation was 1.11 while the mean for large school administrators was 3.16 and the standard deviation was 1.14.

For statement 4, teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform, the mean for small school administrators was 3.19 and the standard deviation was 1.20 while the mean for large school administrators was 3.04 and the standard deviation was 1.21.

For statement 5, doing poorly on high-stakes tests will lead to increased student effort to learn, the mean for small school administrators was 4.11 and the standard deviation was .81 while the mean for large school administrators was 3.84 and the standard deviation was .93.

For statement 6, students work harder and learn more because they know what is expected and that the high-stakes tests really count, the mean for small school administrators was 3.54 and the standard deviation was 1.06 while the mean for large school administrators was 3.26 and the standard deviation was 1.08.

For statement 7, the public display of high-stakes test scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught, the mean for small school administrators was 2.24 and the standard deviation was .88 while the mean for large school administrators was 2.12 and the standard deviation was .91.

For statement 8, when high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available, the mean for small school administrators was 2.95 and the standard deviation was 1.16 while the mean for large school administrators was 2.84 and the standard deviation was 1.09.

For statement 9, administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs, the mean for small school administrators was 3.08 and the standard deviation was 1.17 while the mean for large school administrators was 3.10 and the standard deviation was 1.17.

For statement 10, educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success, the mean for small school administrators was 2.29 and the standard deviation was .89 while the mean for large school administrators was 2.07 and the standard deviation was .86.

For statement 11, driven by the demands of high-stakes tests, professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise, the mean for small school administrators was 2.71 and the standard deviation was .97 while the mean for large school administrators was 2.53 and the standard deviation was 1.04.

For statement 12, the implementation of high-stakes testing has been a catalyst for increased attention to students with special needs, the mean for small school administrators was 2.79 and the standard deviation was 1.15 while the mean for large

school administrators was 2.54 and the standard deviation was 1.09. The means and standard deviations for statements 1 – 12 by gender are summarized in Table 4.14.

Table 4.14—Means and Standard Deviations by Campus Classification for Statements 1 – 12

	Statement	Classification	N	M	SD
1.	High-stakes tests have helped focus public attention on schools with low-achieving students.	Small	63	2.56	1.028
		Large	115	2.42	1.059
2.	High-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach.	Small	63	3.02	1.039
		Large	115	2.83	1.126
3.	High-stakes tests have helped close the gap in achievement between minority students and our majority students in Texas.	Small	63	2.87	1.114
		Large	115	3.16	1.144
4.	Teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform.	Small	63	3.19	1.203
		Large	115	3.04	1.210
5.	Doing poorly on high-stakes tests will lead to increased student effort to learn.	Small	63	4.11	.805
		Large	115	3.84	.933
6.	Students work harder and learn more because they know what is expected and that the high-stakes tests really count.	Small	63	3.54	1.060
		Large	115	3.26	1.077
7.	The public display of high-stakes tests scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught.	Small	63	2.24	.875
		Large	115	2.12	.909
8.	When high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available.	Small	63	2.95	1.156
		Large	115	2.84	1.089
9.	Administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs.	Small	63	3.08	1.168
		Large	115	3.10	1.173
10.	Educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success.	Small	63	2.29	.888
		Large	115	2.07	.856
11.	Driven by the demands of high-stakes tests, Professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise.	Small	63	2.71	.974
		Large	115	2.53	1.037
12.	The implementation of high-stakes testing has been a catalyst for increased attention to students with special needs.	Small	63	2.79	1.152
		Large	115	2.54	1.086

To examine research question 1, four sets of independent t-tests were conducted to test for a difference between the means of two independent groups for statements 1 – 12. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years vs. 15 or More Years), Campus Classification (Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable). Independent t-tests were calculated for campus classification.

For statement 1, high-stakes tests have helped focus public attention on schools with low-achieving students, there was not a significant difference, $p = .398$.

For statement 2, high-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach, there was not a significant difference, $p = .260$.

For statement 3, high-stakes tests have helped close the gap in achievement between minority students and majority students in Texas, there was not a significant difference, $p = .110$.

For statement 4, teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform, there was not a significant difference, $p = .438$.

For statement 5, doing poorly on high-stakes tests will lead to increased student effort to learn, there was a significant difference, $p = .047$.

For statement 6, students work harder and learn more because they know what is expected and that the high-stakes tests really count, there was not a significant difference, $p = .098$.

For statement 7, the public display of high-stakes test scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught, there was a significant difference, $p = .404$.

For statement 8, when high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available, there was not a significant difference, $p = .541$.

For statement 9, administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs, there was not a significant difference, $p = .892$.

For statement 10, educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success, there was not a significant difference, $p = .118$.

For statement 11, driven by the demands of high-stakes tests, professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise, there was not a significant difference, $p = .242$.

For statement 12, the implementation of high-stakes testing has been a catalyst for increased attention to students with special needs, there was not a significant difference, $p = .153$.

Independent t-tests for statements 1 – 12 by campus classification are summarized in Table 4.15.

Table 4.15—Results of t-tests by Campus Classification for Statements 1 – 12

	Statements	Equal Variances	t	df	Sig (2-tailed)
1.	High-stakes tests have helped focus public attention on schools with low-achieving students.	Assumed	.841	176	.402
		Not Assumed	.848	130.991	.398
2.	High-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach.	Assumed	1.105	176	.271
		Not Assumed	1.131	136.625	.260
3.	High-stakes tests have helped close the gap in achievement between minority students and our majority students in Texas.	Assumed	-1.596	176	.112
		Not Assumed	-1.608	130.586	.110
4.	Teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform.	Assumed	.777	176	.438
		Not Assumed	.778	128.296	.438
5.	Doing poorly on high-stakes tests will lead to increased student effort to learn.	Assumed	1.919	176	.057
		Not Assumed	2.002	144.227	.047
6.	Students work harder and learn more because they know what is expected and that the high-stakes tests really count.	Assumed	1.661	176	.098
		Not Assumed	1.669	129.427	.098
7.	The public display of high-stakes tests scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught.	Assumed	.827	176	.409
		Not Assumed	.837	132.014	.404
8.	When high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available.	Assumed	.624	176	.533
		Not Assumed	.613	121.333	.541
9.	Administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs.	Assumed	-.136	176	.892
		Not Assumed	-.136	128.110	.892
10.	Educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success.	Assumed	1.591	176	.114
		Not Assumed	1.573	123.728	.118
11.	Driven by the demands of high-stakes tests, Professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise.	Assumed	1.155	176	.250
		Not Assumed	1.176	134.654	.242
12.	The implementation of high-stakes testing has been a catalyst for increased attention to students with special needs.	Assumed	1.463	176	.145
		Not Assumed	1.438	121.438	.153

To examine research question 1, four sets of independent t-tests were conducted to test for a difference between the means of two independent groups for statements 1 – 12. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years vs. 15 or More Years), Campus Classification (Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable). Means, standard deviation, and N were calculated for gender.

For statement 1, high-stakes tests have helped focus public attention on schools with low-achieving students, the mean for male was 2.43 and the standard deviation was 1.00 while the mean for female was 2.53 and the standard deviation was 1.13.

For statement 2, high-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach, the mean for male was 2.95 and the standard deviation was 1.10 while the mean for female was 2.81 and the standard deviation was 1.10.

For statement 3, high-stakes tests have helped close the gap in achievement between minority students and majority students in Texas, the mean for male was 2.95 and the standard deviation was 1.13 while the mean for female was 3.22 and the standard deviation was 1.14.

For statement 4, teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform, the mean for male was 3.18 and the standard deviation was 1.22 while the mean for female was 2.96 and the standard deviation was 1.18.

For statement 5, doing poorly on high-stakes tests will lead to increased student effort to learn, the mean for male was 3.99 and the standard deviation was .82 while the mean for female was 3.85 and the standard deviation was 1.01.

For statement 6, students work harder and learn more because they know what is expected and that the high-stakes tests really count, the mean for male was 3.39 and the standard deviation was 1.00 while the mean for female was 3.31 and the standard deviation was 1.19.

For statement 7, the public display of high-stakes test scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught, the mean for male was 2.28 and the standard deviation was .94 while the mean for female was 1.97 and the standard deviation was .79.

For statement 8, when high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available, the mean for male was 2.84 and the standard deviation was 1.13 while the mean for female was 2.96 and the standard deviation was 1.09.

For statement 9, administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs, the mean for male was 3.10 and the standard deviation was 1.17 while the mean for female was 3.09 and the standard deviation was 1.18.

For statement 10, educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and

identify roots of success, the mean for male was 2.15 and the standard deviation was .86 while the mean for female was 2.15 and the standard deviation was .90.

For statement 11, driven by the demands of high-stakes tests, professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise, the mean for male was 2.61 and the standard deviation was .96 while the mean for female was 2.57 and the standard deviation was 1.11.

For statement 12, the implementation of high-stakes testing has been a catalyst for increased attention to students with special needs, the mean for male was 2.71 and the standard deviation was 1.14 while the mean for female was 2.50 and the standard deviation was 1.06. The means and standard deviations for statements 1 – 12 by gender are summarized in Table 4.16.

Table 4.16—Means and Standard Deviations by Gender for Statements 1 – 12

	Statement	Gender	N	M	SD
1.	High-stakes tests have helped focus public attention on schools with low-achieving students.	Male	110	2.43	1.000
		Female	68	2.53	1.126
2.	High-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach.	Male	110	2.95	1.099
		Female	68	2.81	1.096
3.	High-stakes tests have helped close the gap in achievement between minority students and our majority students in Texas.	Male	110	2.95	1.128
		Female	68	3.22	1.144
4.	Teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform.	Male	110	3.18	1.220
		Female	68	2.96	1.177
5.	Doing poorly on high-stakes tests will lead to increased student effort to learn.	Male	110	3.99	.818
		Female	68	3.85	1.011
6.	Students work harder and learn more because they know what is expected and that the high-stakes tests really count.	Male	110	3.39	1.005
		Female	68	3.31	1.188

Table 4.16—Continued

	Statement	Gender	N	M	SD
7.	The public display of high-stakes tests scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught.	Male	110	2.28	.940
		Female	68	1.97	.791
8.	When high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available.	Male	110	2.84	1.129
		Female	68	2.96	1.085
9.	Administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs.	Male	110	3.10	1.165
		Female	68	3.09	1.181
10.	Educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success.	Male	110	2.15	.855
		Female	68	2.15	.902
11.	Driven by the demands of high-stakes tests, Professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise.	Male	110	2.61	.959
		Female	68	2.57	1.111
12.	The implementation of high-stakes testing has been a catalyst for increased attention to students with special needs.	Male	110	2.71	1.144
		Female	68	2.50	1.058

Independent t-tests were calculated for gender (male and female). For statement 1, high-stakes tests have helped focus public attention on schools with low-achieving students, there was not a significant difference, $p = .541$.

For statement 2, high-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach, there was not a significant difference, $p = .421$.

For statement 3, high-stakes tests have helped close the gap in achievement between minority students and majority students in Texas, there was not a significant difference, $p = .132$.

For statement 4, teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform, there was not a significant difference, $p = .222$.

For statement 5, doing poorly on high-stakes tests will lead to increased student effort to learn, there was not a significant difference, $p = .344$.

For statement 6, students work harder and learn more because they know what is expected and that the high-stakes tests really count, there was not a significant difference, $p = .636$.

For statement 7, the public display of high-stakes test scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught, there was a significant difference, $p = .019$.

For statement 8, when high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available, there was not a significant difference, $p = .483$.

For statement 9, administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs, there was not a significant difference, $p = .948$.

For statement 10, educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success, there was not a significant difference, $p = .991$.

For statement 11, driven by the demands of high-stakes tests, professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise, there was not a significant difference, $p = .827$.

For statement 12, the implementation of high-stakes testing has been a catalyst for increased attention to students with special needs, there was not a significant difference, $p = .216$.

Independent t-tests for statements 1 – 12 by gender (male and female) are summarized in Table 4.17.

Table 4.17—Results of t-tests by Gender for Statements 1 – 12

	Statement	Equal Variances	t	df	Sig. (2-tailed)
1.	High-stakes tests have helped focus public attention on schools with low-achieving students.	Assumed	-.631	176	.529
		Not Assumed	-.613	129.32	.541
2.	High-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach.	Assumed	.807	176	.421
		Not Assumed	.807	142.4	.421
3.	High-stakes tests have helped close the gap in achievement between minority students and our majority students in Texas.	Assumed	-1.520	176	.130
		Not Assumed	-1.515	140.6	.132
4.	Teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform.	Assumed	1.216	176	.226
		Not Assumed	1.227	146.0	.222
5.	Doing poorly on high-stakes tests will lead to increased student effort to learn.	Assumed	.997	176	.320
		Not Assumed	.949	120.1	.344
6.	Students work harder and learn more because they know what is expected and that the high-stakes tests really count.	Assumed	.493	176	.622
		Not Assumed	.474	124.5	.636
7.	The public display of high-stakes tests scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught.	Assumed	2.277	176	.024
		Not Assumed	2.371	160.0	.019

Table 4.17—Continued

	Statement	Equal Variances	t	df	Sig. (2-tailed)
8.	When high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available.	Assumed	-.696	176	.487
		Not Assumed	-.703	146.4	.483
9.	Administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs.	Assumed	.065	176	.948
		Not Assumed	.065	140.6	.948
10.	Educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success.	Assumed	-.012	176	.991
		Not Assumed	-.012	136.2	.991
11.	Driven by the demands of high-stakes tests, Professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise.	Assumed	.226	176	.821
		Not Assumed	.218	126.5	.827
12.	The implementation of high-stakes testing has been a catalyst for increased attention to students with special needs.	Assumed	1.219	176	.225
		Not Assumed	1.242	150.5	.216

Means, standard deviation, and N were calculated for years of administrative experience. For statement 1, high-stakes tests have helped focus public attention on schools with low-achieving students, the mean for 1 – 4 years of administrative experiences was 2.61 and the standard deviation was 1.04 while the mean for 15 or more years of administrative experiences was 2.61 and the standard deviation was 1.05.

For statement 2, high-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach, the mean for 1 – 4 years of administrative experiences was 3.02 and the standard deviation was 1.09 while the mean for 15 or more years of administrative experiences was 2.96 and the standard deviation was 1.06.

For statement 3, high-stakes tests have helped close the gap in achievement between minority students and majority students in Texas, the mean for 1 – 4 years of

administrative experiences was 3.24 and the standard deviation was 1.06 while the mean for 15 or more years of administrative experiences was 2.93 and the standard deviation was 1.00.

For statement 4, teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform, the mean for 1 – 4 years of administrative experiences was 3.17 and the standard deviation was 1.15 while the mean for 15 or more years of administrative experiences was 3.30 and the standard deviation was 1.24.

For statement 5, doing poorly on high-stakes tests will lead to increased student effort to learn, the mean for 1 – 4 years of administrative experiences was 3.98 and the standard deviation was .71 while the mean for 15 or more years of administrative experiences was 3.78 and the standard deviation was .89.

For statement 6, students work harder and learn more because they know what is expected and that the high-stakes tests really count, the mean for 1 – 4 years of administrative experiences was 3.31 and the standard deviation was 1.03 while the mean for 15 or more years of administrative experiences was 3.30 and the standard deviation was .95.

For statement 7, the public display of high-stakes test scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught, the mean for 1 – 4 years of administrative experiences was 1.94 and the standard deviation was .79 while the mean for 15 or more years of administrative experiences was 2.48 and the standard deviation was 1.01.

For statement 8, when high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available, the mean for 1 – 4 years of administrative experiences was 3.02 and the standard deviation was 1.06 while the mean for 15 or more years of administrative experiences was 3.19 and the standard deviation was 1.04.

For statement 9, administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs, the mean for 1 – 4 years of administrative experiences was 3.04 and the standard deviation was 1.15 while the mean for 15 or more years of administrative experiences was 3.30 and the standard deviation was 1.10.

For statement 10, educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success, the mean for 1 – 4 years of administrative experiences was 2.13 and the standard deviation was .87 while the mean for 15 or more years of administrative experiences was 2.30 and the standard deviation was .82.

For statement 11, driven by the demands of high-stakes tests, professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise, the mean for 1 – 4 years of administrative experiences was 2.74 and the standard deviation was 1.03 while the mean for 15 or more years of administrative experiences was 2.52 and the standard deviation was .94.

For statement 12, the implementation of high-stakes testing has been a catalyst for increased attention to students with special needs, the mean for 1 – 4 years of administrative experiences was 2.78 and the standard deviation was 1.09 while the mean

for 15 or more years of administrative experiences was 2.70 and the standard deviation was 1.10. The means and standard deviations for statements 1 – 12 by gender are summarized in Table 4.18.

Table 4.18—Means and Standard Deviations by Years of Administrative Experience for Statements 1 – 12

Statement		Administrative Experience (years)	N	M	SD
1.	High-stakes tests have helped focus public attention on schools with low-achieving students.	1 – 4	54	2.61	1.036
		15 or more	27	2.61	1.050
2.	High-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach.	1 – 4	54	3.02	1.090
		15 or more	27	2.96	1.055
3.	High-stakes tests have helped close the gap in achievement between minority students and our majority students in Texas.	1 – 4	54	3.24	1.063
		15 or more	27	2.93	.997
4.	Teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform.	1 – 4	54	3.17	1.145
		15 or more	27	3.30	1.235
5.	Doing poorly on high-stakes tests will lead to increased student effort to learn.	1 – 4	54	3.98	.714
		15 or more	27	3.78	.892
6.	Students work harder and learn more because they know what is expected and that the high-stakes tests really count.	1 – 4	54	3.31	1.025
		15 or more	27	3.30	.953
7.	The public display of high-stakes tests scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught.	1 – 4	54	1.94	.787
		15 or more	27	2.48	1.014
8.	When high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available.	1 – 4	54	3.02	1.055
		15 or more	27	3.19	1.039
9.	Administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs.	1 – 4	54	3.04	1.149
		15 or more	27	3.30	1.103
10.	Educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success.	1 – 4	54	2.13	.870
		15 or more	27	2.30	.823
11.	Driven by the demands of high-stakes tests, Professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise.	1 – 4	54	2.74	1.031
		15 or more	27	2.52	.935
12.	The implementation of high-stakes testing has been a catalyst for increased attention to students with special needs.	1 – 4	54	2.78	1.093
		15 or more	27	2.70	1.103

To examine research question 1, four sets of independent t-tests were conducted to test for a difference between the means of two independent groups for statements 1 – 12.

Independent t-tests were calculated for years of administrative experience (1 – 4 years vs. 15 or more years). For statement 1, high-stakes tests have helped focus public attention on schools with low-achieving students, there was not a significant difference, $p = .121$.

For statement 2, high-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach, there was not a significant difference, $p = .826$.

For statement 3, high-stakes tests have helped close the gap in achievement between minority students and majority students in Texas, there was not a significant difference, $p = .196$.

For statement 4, teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform, there was not a significant difference, $p = .650$.

For statement 5, doing poorly on high-stakes tests will lead to increased student effort to learn, there was not a significant difference, $p = .307$.

For statement 6, students work harder and learn more because they know what is expected and that the high-stakes tests really count, there was not a significant difference, $p = .936$.

For statement 7, the public display of high-stakes test scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of

the curriculum and are being successfully taught, there was a significant difference, $p = .020$.

For statement 8, when high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available, there was not a significant difference, $p = .501$.

For statement 9, administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs, there was not a significant difference, $p = .330$.

For statement 10, educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success, there was not a significant difference, $p = .403$.

For statement 11, driven by the demands of high-stakes tests, professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise, there was not a significant difference, $p = .334$.

For statement 12, the implementation of high-stakes testing has been a catalyst for increased attention to students with special needs, there was not a significant difference, $p = .776$. Independent t-tests for statements 1 – 12 by years of administrative experience (1 – 4 years vs. 15 or more years) are summarized in Table 4.19.

Table 4.19—Results of t-tests by Years of Administrative Experience for Statements 1 – 12

	Statement	Equal Variances	t	df	Sig. (2-tailed)
1.	High-stakes tests have helped focus public attention on schools with low-achieving students.	Assumed	1.586	79	.117
		Not Assumed	1.579	51.465	.121
2.	High-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach.	Assumed	.218	79	.828
		Not Assumed	.221	53.646	.826
3.	High-stakes tests have helped close the gap in achievement between minority students and our majority students in Texas.	Assumed	1.282	79	.204
		Not Assumed	1.310	55.192	.196
4.	Teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform.	Assumed	-.468	79	.641
		Not Assumed	-.456	48.754	.650
5.	Doing poorly on high-stakes tests will lead to increased student effort to learn.	Assumed	1.113	79	.269
		Not Assumed	1.033	43.148	.307
6.	Students work harder and learn more because they know what is expected and that the high-stakes tests really count.	Assumed	.078	79	.938
		Not Assumed	.080	55.610	.936
7.	The public display of high-stakes tests scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught.	Assumed	-2.624	79	.010
		Not Assumed	-2.412	42.144	.020
8.	When high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available.	Assumed	-.674	79	.503
		Not Assumed	-.677	52.820	.501
9.	Administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs.	Assumed	-.970	79	.335
		Not Assumed	-.984	54.047	.330
10.	Educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success.	Assumed	-.827	79	.411
		Not Assumed	-.843	54.741	.403
11.	Driven by the demands of high-stakes tests, Professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise.	Assumed	.942	79	.349
		Not Assumed	.974	56.908	.334
12.	The implementation of high-stakes testing has been a catalyst for increased attention to students with special needs.	Assumed	.287	79	.775
		Not Assumed	.286	51.694	.776

Means, standard deviation, and N were calculated for current campus rating. For statement 1, high-stakes tests have helped focus public attention on schools with low-

achieving students, the mean for exemplary and recognized campus rating was 2.30 and the standard deviation was 1.10 while the mean for academically acceptable campus rating was 2.49 and the standard deviation was 1.03.

For statement 2, high-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach, the mean for exemplary and recognized campus rating was 2.72 and the standard deviation was 1.03 while the mean for academically acceptable campus rating was 2.91 and the standard deviation was 1.11.

For statement 3, high-stakes tests have helped close the gap in achievement between minority students and majority students in Texas, the mean for exemplary and recognized campus rating was 2.77 and the standard deviation was 1.17 while the mean for academically acceptable campus rating was 3.11 and the standard deviation was 1.11.

For statement 4, teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform, the mean for exemplary and recognized campus rating was 2.91 and the standard deviation was 1.29 while the mean for academically acceptable campus rating was 3.17 and the standard deviation was 1.18.

For statement 5, doing poorly on high-stakes tests will lead to increased student effort to learn, the mean for exemplary and recognized campus rating was 3.93 and the standard deviation was .80 while the mean for academically acceptable campus rating was 3.92 and the standard deviation was .93.

For statement 6, students work harder and learn more because they know what is expected and that the high-stakes tests really count, the mean for exemplary and

recognized campus rating was 3.49 and the standard deviation was 1.03 while the mean for academically acceptable campus rating was 3.29 and the standard deviation was 1.09.

For statement 7, the public display of high-stakes test scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught, the mean for exemplary and recognized campus rating was 2.16 and the standard deviation was .92 while the mean for academically acceptable campus rating was 2.16 and the standard deviation was .92.

For statement 8, when high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available, the mean for exemplary and recognized campus rating was 2.67 and the standard deviation was 1.06 while the mean for academically acceptable campus rating was 2.90 and the standard deviation was 1.12.

For statement 9, administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs, the mean for exemplary and recognized campus rating was 2.91 and the standard deviation was 1.19 while the mean for academically acceptable campus rating was 3.13 and the standard deviation was 1.16.

For statement 10, educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success, the mean for exemplary and recognized campus rating was 2.21 and the standard deviation was .91 while the mean for academically acceptable campus rating was 2.13 and the standard deviation was .84.

For statement 11, driven by the demands of high-stakes tests, professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise, the mean for exemplary and recognized campus rating was 2.53 and the standard deviation was .93 while the mean for academically acceptable campus rating was 2.60 and the standard deviation was 1.03.

For statement 12, the implementation of high-stakes testing has been a catalyst for increased attention to students with special needs, the mean for exemplary and recognized campus rating was 2.65 and the standard deviation was 1.13 while the mean for academically acceptable campus rating was 2.60 and the standard deviation was 1.08. The means and standard deviations for statements 1 – 12 by gender are summarized in Table 4.20.

Table 4.20—Means and Standard Deviations by Current Campus Rating for Statements 1 – 12

	Statement	Rating	N	M	SD
1.	High-stakes tests have helped focus public attention on schools with low-achieving students.	Exemplary and Recognized	43	2.30	1.103
		Academically Acceptable	126	2.49	1.026
2.	High-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach.	Exemplary and Recognized	43	2.72	1.031
		Academically Acceptable	126	2.91	1.110
3.	High-stakes tests have helped close the gap in achievement between minority students and our majority students in Texas.	Exemplary and Recognized	43	2.77	1.172
		Academically Acceptable	126	3.11	1.112
4.	Teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform.	Exemplary and Recognized	43	2.91	1.288
		Academically Acceptable	126	3.17	1.178
5.	Doing poorly on high-stakes tests will lead to increased student effort to learn.	Exemplary and Recognized	43	3.93	.799
		Academically Acceptable	126	3.92	.926

Table 4.20—Continued

	Statement	Rating	N	M	SD
6.	Students work harder and learn more because they know what is expected and that the high-stakes tests really count.	Exemplary and Recognized	43	3.49	1.032
		Academically Acceptable	126	3.29	1.094
7.	The public display of high-stakes tests scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught.	Exemplary and Recognized	43	2.16	.924
		Academically Acceptable	126	2.16	.916
8.	When high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available.	Exemplary and Recognized	43	2.67	1.063
		Academically Acceptable	126	2.90	1.116
9.	Administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs.	Exemplary and Recognized	43	2.91	1.192
		Academically Acceptable	126	3.13	1.162
10.	Educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success.	Exemplary and Recognized	43	2.21	.914
		Academically Acceptable	126	2.13	.839
11.	Driven by the demands of high-stakes tests, Professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise.	Exemplary and Recognized	43	2.53	.93
		Academically Acceptable	126	2.60	1.028
12.	The implementation of high-stakes testing has been a catalyst for increased attention to students with special needs.	Exemplary and Recognized	43	2.65	1.131
		Academically Acceptable	126	2.60	1.075

Independent t-tests were calculated for current campus rating (exemplary and recognized vs. academically acceptable). For statement 1, high-stakes tests have helped focus public attention on schools with low-achieving students, there was not a significant difference, $p = .325$.

For statement 2, high-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach, there was not a significant difference, $p = .305$.

For statement 3, high-stakes tests have helped close the gap in achievement between minority students and majority students in Texas, there was not a significant difference, $p = .097$.

For statement 4, teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform, there was not a significant difference, $p = .248$.

For statement 5, doing poorly on high-stakes tests will lead to increased student effort to learn, there was not a significant difference, $p = .948$.

For statement 6, students work harder and learn more because they know what is expected and that the high-stakes tests really count, there was not a significant difference, $p = .277$.

For statement 7, the public display of high-stakes test scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught, there was not a significant difference, $p = .980$.

For statement 8, when high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available, there was not a significant difference, $p = .246$.

For statement 9, administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs, there was not a significant difference, $p = .279$.

For statement 10, educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success, there was not a significant difference, $p = .605$.

For statement 11, driven by the demands of high-stakes tests, professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise, there was not a significant difference, $p = .688$.

For statement 12, the implementation of high-stakes testing has been a catalyst for increased attention to students with special needs, there was not a significant difference, $p = .778$. Independent t-tests for statements 1 – 12 by current campus rating (exemplary and recognized vs. academically acceptable) are summarized in Table 4.21.

Table 4.21—Results of t-tests by Current Campus Rating for Statements 1 – 12

	Statement	Equal Variances	t	df	Sig. (2-tailed)
1.	High-stakes tests have helped focus public attention on schools with low-achieving students.	Assumed	-1.028	167	.306
		Not Assumed	-.991	68.456	.325
2.	High-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach.	Assumed	-.995	167	.321
		Not Assumed	-1.032	77.715	.305
3.	High-stakes tests have helped close the gap in achievement between minority students and our majority students in Texas.	Assumed	-1.727	167	.086
		Not Assumed	-1.682	69.550	.097
4.	Teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform.	Assumed	-1.219	167	.225
		Not Assumed	-1.166	67.675	.248
5.	Doing poorly on high-stakes tests will lead to increased student effort to learn.	Assumed	.061	167	.952
		Not Assumed	.065	83.481	.948
6.	Students work harder and learn more because they know what is expected and that the high-stakes tests really count.	Assumed	1.063	167	.289
		Not Assumed	1.095	76.622	.277
7.	The public display of high-stakes tests scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught.	Assumed	.025	167	.980
		Not Assumed	.025	72.151	.980
8.	When high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available.	Assumed	-1.142	167	.255
		Not Assumed	-1.170	75.936	.246
9.	Administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs.	Assumed	-1.104	167	.271
		Not Assumed	-1.090	71.156	.279

Table 4.21—Continued

	Statement	Equal Variances	t	df	Sig. (2-tailed)
10.	Educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success.	Assumed	.543	167	.588
		Not Assumed	.520	67.717	.605
11.	Driven by the demands of high-stakes tests, Professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise.	Assumed	-.385	167	.701
		Not Assumed	-.403	79.306	.688
12.	The implementation of high-stakes testing has been a catalyst for increased attention to students with special needs.	Assumed	.291	167	.772
		Not Assumed	.283	69.624	.778

Research Question 2

Question 2: Do perceptions of researchers of high-stakes testing reported in current literature differ from those of Texas secondary school administrators in the membership of TASSP?

To examine research question 2, means and standard deviations were calculated on statements 13 – 19 to determine how the perceptions of researchers of high-stakes testing reported in current literature differ from those of Texas secondary school administrators in the membership of TASSP.

For statement 13, one result of high-stakes testing is that educators know more about testing than ever before, the mean was 2.11 and the standard deviation was .89.

For statement 14, prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled, the mean was 2.24 and the standard deviation was 5.76.

For statement 15, high-stakes tests promote greater homogeneity of education, the mean was 2.56 and the standard deviation was .93.

For statement 16, high-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions, the mean was 3.64 and the standard deviation was 1.02.

For statement 17, high-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has led to teachers enhancing their own assessment practices, the mean was 2.72 and the standard deviation was 1.06.

For statement 18, high-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time, the mean was 1.96 and the standard deviation was 1.04.

For statement 19, high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels, the mean was 2.09 and the standard deviation was 1.15. The means and standard deviations for statements 13 – 19 are summarized in Table 4.22.

Table 4.22—Means and Standard Deviations for Statements 13 – 19

	Statement	N	Min.	Max.	M	SD
13.	One result of high-stakes testing is that educators know more about testing than ever before.	178	1	5	2.11	.89
14.	Prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled.	178	1	5	2.24	.76
15.	High-stakes tests promote greater homogeneity of education.	178	1	5	2.56	.93
16.	High-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions.	178	1	5	3.64	1.02
17.	High-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has lead to teachers enhancing their own assessment practices.	178	1	5	2.72	1.06
18.	High-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time.	178	1	5	1.96	1.04
19.	High-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels.	178	1	5	2.09	1.15

Almost 80% of respondents had a positive position to statement 13; one result of high-stakes testing is that educators know more about testing than ever before. Only 21 respondents out of the population of 178 administrators found any degree of disagreement with this statement. This information is presented in Table 4.23.

Table 4.23—Frequency Distribution of Responses to Statement Number 13: One Result of High-stakes Testing Is That Educators Know More About Testing Than Ever Before

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	39	21.9	21.9
Agree	103	57.9	79.8
Unsure	15	8.4	88.2
Disagree	20	11.2	99.4
Strongly Disagree	1	0.6	100.0
Total	178	100.0	

Over 74% of respondents had a positive position in regard to how prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled. Only 13 respondents out of the population of 178 administrators found any degree of disagreement with this statement. This information is presented in Table 4.24.

Table 4.24—Frequency Distribution of Responses to Statement Number 14: Prominent and Public Interest in Pupil Performance on High-stakes Tests Has Resulted in an Intensity of Effort Directed Toward Data Collection and Quality Control That Is Unparalleled

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	18	10.1	10.1
Agree	114	64.0	74.2
Unsure	33	18.5	92.7
Disagree	11	6.2	98.9
Strongly Disagree	2	1.1	100.0
Total	178	100.0	

Roughly, 58% of respondents agreed or strongly agreed with the position that High-stakes tests promote greater homogeneity of education. Surprisingly, almost 22% of administrators participating in the survey were unsure about their position on this issue. In contrast, only approximately 20% of respondents disagreed to any degree with this statement. This information is presented in Table 4.25.

Table 4.25—Frequency Distribution of Responses to Statement Number 15: High-stakes Tests Promote Greater Homogeneity of Education. A Result of Schools’ Aligning Their Curricula and Instructional Focus More Closely to Outcomes Embodied in High-stakes Tests, the Experiences of and Aspirations for Children in Urban, Suburban, and Rural Districts Within a State Are More Comparable Than They Have Been in the Recent Past

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	13	7.3	7.3
Agree	91	51.1	58.4
Unsure	39	21.9	80.3
Disagree	32	18.0	98.3
Strongly Disagree	3	1.7	100.0
Total	178	100.0	

Respondents compellingly disagreed with the statement that high-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions when 64.6% disagreed or strongly disagreed. In contrast, only 18.5% of respondents found any level of agreement with the statement. This information is presented in Table 4.26.

Table 4.26—Frequency Distribution of Responses to Statement Number 16: A Profoundly Positive Effect That the Introduction of High-stakes Consequences Has Had Lies in the Tests Themselves. High-stakes Tests Have Evolved to a State of Being: Highly Reliable; Free from Bias; Relevant and Age Appropriate; Higher Order; Tightly Related to Important Public Goals; Time and Cost Efficient; and Yielding Remarkably Consistent Decisions

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	2	1.1	1.1
Agree	31	17.4	18.5
Unsure	30	16.9	35.4
Disagree	81	45.5	80.9
Strongly Disagree	34	19.1	100.0
Total	178	100.0	

A majority of respondents, 55.1%, found a level of agreement with the position that high-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. Only 29.3% of respondents found a degree of opposition to the statement. This information is presented in Table 4.27.

Table 4.27—Frequency Distribution of Responses to Statement Number 17: High-stakes Tests Have Exposed Educators to High-Quality Writing Prompts, Document-Based Questions, Constructed-Response Formats, and Even Challenging Multiple-Choice Items. This Has Led to Teachers Enhancing Their Own Assessment Practices

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	12	6.7	6.7
Agree	86	48.3	55.1
Unsure	28	15.7	70.8
Disagree	43	24.2	94.9
Strongly Disagree	9	5.1	100.0
Total	178	100.0	

In excess of 80% of respondents forcefully supported the statement that high-stakes testing programs result in massive amounts of test preparation, resulting in a loss of instructional time. In stark contrast, only 14% of respondents disagreed with this stance. This information is presented in Table 4.28.

Almost 75% of respondents agreed at some level with the position that high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. Almost 20% of respondents disagreed with this view. A point of interest is the small number of unsure responses, a mere 4.5% generated by this statement. This information is presented in Table 4.29.

Table 4.28—Frequency Distribution of Responses to Statement Number 18: High-stakes Testing Programs Also Result in Massive Amounts of Test Preparation, Resulting in a Loss of Instructional Time

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	70	39.3	39.3
Agree	73	41.0	80.3
Unsure	10	5.6	86.0
Disagree	23	12.9	98.9
Strongly Disagree	2	1.1	100.0
Total	178	100.0	

Table 4.29—Frequency Distribution of Responses to Statement Number 19: High-stakes Testing Has Resulted in a Loss of Local Control of What Is Taught, How It Is Taught, and Who Gets High-Quality Instruction. These Decisions Are Now Greatly Impacted by Policy Makers at the State and National Levels

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	68	38.2	38.2
Agree	65	36.5	74.7
Unsure	8	4.5	79.2
Disagree	35	19.7	98.9
Strongly Disagree	2	1.1	100.0
Total	178	100.0	

To examine research question 2, four sets of independent t-tests were conducted to test for a difference between the means of two independent groups for statements 13 – 19. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years vs. 15 or More Years), Campus Classification (Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable). Means, standard deviation, and N were calculated for all statements in regard to campus classification. For statement 13, one result of high-stakes testing is that educators know more about testing than ever before, the mean for small school administrators was 2.10 and the standard deviation was .91

while the mean for large school administrators was 2.11 and the standard deviation was .87.

For statement 14, prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled, the mean for small school administrators was 2.37 and the standard deviation was .87 while the mean for large school administrators was 2.37 and the standard deviation was .87.

For statement 15, high-stakes tests promote greater homogeneity of education, the mean for small school administrators was 2.67 and the standard deviation was .86 while the mean for large school administrators was 2.50 and the standard deviation was .96.

For statement 16, high-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions, the mean for small school administrators was 3.76 and the standard deviation was .91 while the mean for large school administrators was 3.57 and the standard deviation was 1.07.

For statement 17, high-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has led to teachers enhancing their own assessment practices, the mean for small school administrators was 2.81 and the standard deviation was 1.00 while the mean for large school administrators was 2.68 and the standard deviation was 1.10.

For statement 18, high-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time, the mean for small school administrators was 1.73 and the standard deviation was .97 while the mean for large school administrators was 2.08 and the standard deviation was 1.05.

For statement 19, high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels, the mean for small school administrators was 1.86 and the standard deviation was 1.06 while the mean for large school administrators was 2.22 and the standard deviation was 1.18. The means and standard deviations for statements 13 – 19 by campus classification are summarized in Table 4.30.

Table 4.30—Means and Standard Deviations by Campus Classification for Statements 13 – 19

	Statement	Classification	N	M	SD
13.	One result of high-stakes testing is that educators know more about testing than ever before.	Small	63	2.10	.911
		Large	115	2.11	.866
14.	Prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled.	Small	63	2.37	.867
		Large	115	2.17	.692
15.	High-stakes tests promote greater homogeneity of education.	Small	63	2.67	.861
		Large	115	2.50	.959
16.	High-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions.	Small	63	3.76	.911
		Large	115	3.57	1.068

Table 4.30—Continued

	Statement	Classification	N	M	SD
17.	High-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has lead to teachers enhancing their own assessment practices.	Small	63	2.81	.998
		Large	115	2.68	1.097
18.	High-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time.	Small	63	1.73	.971
		Large	115	2.08	1.053
19.	High-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels.	Small	63	1.86	1.060
		Large	115	2.22	1.176

To examine research question 2, four sets of independent t-tests were conducted to test for a difference between the means of two independent groups for statements 13 – 19. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years vs. 15 or More Years), Campus Classification (Large and Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable). Independent t-tests were calculated for campus classification (large and small). For statement 13, one result of high-stakes testing is that educators know more about testing than ever before, there was not a significant difference, $p = .090$.

For statement 14, prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled, there was not a significant difference, $p = .135$.

For statement 15, high-stakes tests promote greater homogeneity of education, there was not a significant difference, $p = .226$.

For statement 16, high-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions, there was not a significant difference, $p = .218$.

For statement 17, high-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items, there was not a significant difference, $p = .419$.

For statement 18, high-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time, there was a significant difference, $p = .028$.

For statement 19, high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels, there was a significant difference, $p = .039$. Independent t-tests for campus classification (large and small) for statements 13 – 19 by gender are summarized in Table 4.31.

Table 4.31—Results of t-tests by Campus Classification for Statements 13 – 19

Statements	Equal Variances	t	df	Sig (2-tailed)
13. One result of high-stakes testing is that educators know more about testing than ever before.	Assumed	-.127	176	.899
	Not Assumed	-.126	124.744	.900
14. Prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled.	Assumed	1.609	176	.109
	Not Assumed	1.507	105.766	.135
15. High-stakes tests promote greater homogeneity of education.	Assumed	1.179	176	.240
	Not Assumed	1.216	139.695	.226

Table 4.31—Continued

	Statements	Equal Variances	t	df	Sig (2-tailed)
16.	High-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions.	Assumed	1.181	176	.239
		Not Assumed	1.237	145.683	.218
17.	High-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has lead to teachers enhancing their own assessment practices.	Assumed	.788	176	.432
		Not Assumed	.810	138.285	.419
18.	High-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time.	Assumed	-2.168	176	.032
		Not Assumed	-2.220	136.746	.028*
19.	High-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels.	Assumed	-2.022	176	.045
		Not Assumed	-2.085	139.269	.039*

*Significant at the .05 level.

To examine research question 2, four sets of independent t-tests were conducted to test for a difference between the means of two independent groups for statements 13 – 19. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years vs. 15 or More Years), Campus Classification (Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable). Means, standard deviation, and N were calculated for gender.

For statement 13, one result of high-stakes testing is that educators know more about testing than ever before, the mean for male was 2.17 and the standard deviation was .90 while the mean for female was 2.00 and the standard deviation was .82.

For statement 14, prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled, the mean for male was 2.24 and the standard deviation was .73 while the mean for female was 2.25 and the standard deviation was .82.

For statement 15, high-stakes tests promote greater homogeneity of education, the mean for male was 2.62 and the standard deviation was .92 while the mean for female was 2.46 and the standard deviation was .94.

For statement 16, high-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions, the mean for male was 3.71 and the standard deviation was .95 while the mean for female was 3.53 and the standard deviation was 1.11.

For statement 17, high-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has led to teachers enhancing their own assessment practices, the mean for male was 2.77 and the standard deviation was 1.06 while the mean for female was 3.53 and the standard deviation was 1.11.

For statement 18, high-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time, the mean for male was 1.89 and the standard deviation was .96 while the mean for female was 2.06 and the standard deviation was 1.15.

For statement 19, high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels, the mean for male was 1.95 and the standard deviation was 1.10 while the mean for female was 2.06 and the standard deviation was 1.15. The means and standard deviations for statements 13 – 19 by gender are summarized in Table 4.32.

Table 4.32—Means and Standard Deviations by Gender for Statements 13 – 19

	Statement	Gender	N	M	SD
13.	One result of high-stakes testing is that educators know more about testing than ever before.	Male	110	2.17	.897
		Female	68	2.00	.881
14.	Prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled.	Male	110	2.24	.729
		Female	68	2.25	.817
15.	High-stakes tests promote greater homogeneity of education.	Male	110	2.62	.919
		Female	68	2.46	.937
16.	High-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions.	Male	110	3.71	.952
		Female	68	3.53	1.113
17.	High-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has led to teachers enhancing their own assessment practices.	Male	110	2.77	1.064
		Female	68	2.65	1.062
18.	High-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time.	Male	110	1.89	.961
		Female	68	2.06	1.145
19.	High-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels.	Male	110	1.95	1.095
		Female	68	2.31	1.200

Independent t-tests were calculated for gender (male and female). For statement 13, one result of high-stakes testing is that educators know more about testing than ever before, there was not a significant difference, $p = .209$.

For statement 14, prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled, there was not a significant difference, $p = .910$.

For statement 15, high-stakes tests promote greater homogeneity of education, there was not a significant difference, $p = .260$.

For statement 16, high-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions, there was not a significant difference, $p = .271$.

For statement 17, high-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has led to teachers enhancing their own assessment practices, there was not a significant difference, $p = .445$.

For statement 18, high-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time, there was not a significant difference, $p = .315$.

For statement 19, high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels, there was a

significant difference, $p = .050$. Independent t-tests for statements 13 – 19 by gender (male and female) are summarized in Table 4.33.

Table 4.33—Results of t-tests by Gender for Statements 13 – 19

	Statement	Equal Variances	t	df	Sig. (2-tailed)
13.	One result of high-stakes testing is that educators know more about testing than ever before.	Assumed	1.257	176	.211
		Not Assumed	1.262	144.043	.209
14.	Prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled.	Assumed	-.116	176	.908
		Not Assumed	-.113	129.750	.910
15.	High-stakes tests promote greater homogeneity of education.	Assumed	1.136	176	.257
16.	High-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions.	Not Assumed	1.131	139.860	.260
		Assumed	1.147	176	.253
		Not Assumed	1.105	125.512	.271
17.	High-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has lead to teachers enhancing their own assessment practices.	Assumed	.766	176	.444
		Not Assumed	.767	142.264	.445
18.	High-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time.	Assumed	-1.052	176	.294
19.	High-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels.	Not Assumed	-1.010	123.700	.315
		Assumed	-2.021	176	.045
		Not Assumed	-1.987	132.209	.050*

*Significant at the .05 level.

For statement 13, one result of high-stakes testing is that educators know more about testing than ever before, the mean for 1 – 4 years of administrative experiences

was 2.31 and the standard deviation was 1.01 while the mean for 15 or more years of administrative experiences was 2.19 and the standard deviation was .74.

For statement 14, prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled, the mean for 1 – 4 years of administrative experiences was 2.41 and the standard deviation was .79 while the mean for 15 or more years of administrative experiences was 2.19 and the standard deviation was .68.

For statement 15, high-stakes tests promote greater homogeneity of education, the mean for 1 – 4 years of administrative experiences was 2.61 and the standard deviation was .90 while the mean for 15 or more years of administrative experiences was 2.63 and the standard deviation was .93.

For statement 16, high-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions, the mean for 1 – 4 years of administrative experiences was 3.67 and the standard deviation was .97 while the mean for 15 or more years of administrative experiences was 3.89 and the standard deviation was .80.

For statement 17, high-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has led to teachers enhancing their own assessment practices, the mean for 1 – 4 years of administrative experiences was 2.83 and the standard deviation was 1.13 while the mean for 15 or more years of administrative experiences was 3.00 and the standard deviation was .96.

For statement 18, high-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time, the mean for 1 – 4 years of administrative experiences was 2.07 and the standard deviation was 1.04 while the mean for 15 or more years of administrative experiences was 1.89 and the standard deviation was .89.

For statement 19, high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels, the mean for 1 – 4 years of administrative experiences was 2.30 and the standard deviation was 1.16 while the mean for 15 or more years of administrative experiences was 2.07 and the standard deviation was 1.04. The means and standard deviations for statements 13 – 19 administrative experiences are summarized in Table 4.34.

Table 4.34—Means and Standard Deviations by Years of Administrative Experience for Statements 13 – 19

Statement		Administrative Experience (years)			
		N	M	SD	
13.	One result of high-stakes testing is that educators know more about testing than ever before.	1 – 4	54	2.31	1.006
		15 or more	27	2.19	.736
14.	Prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled.	1 – 4	54	2.41	.790
		15 or more	27	2.19	.681
15.	High-stakes tests promote greater homogeneity of education.	1 – 4	54	2.61	.899
		15 or more	27	2.63	.926
16.	High-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions.	1 – 4	54	3.67	.971
		15 or more	27	3.89	.801

Table 4.34—Continued

Statement		Administrative	N	M	SD
		Experience (years)			
17.	High-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has lead to teachers enhancing their own assessment practices.	1 – 4	54	2.83	1.129
		15 or more	27	3.00	.961
18.	High-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time.	1 – 4	54	2.07	1.043
		15 or more	27	1.89	.892
19.	High-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels.	1 – 4	54	2.30	1.160
		15 or more	27	2.07	1.035

Independent t-tests were calculated for years of administrative experience (1 – 4 years vs. 15 or more years). For statement 13, one result of high-stakes testing is that educators know more about testing than ever before, there was not a significant difference, $p = .513$.

For statement 14, prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled, there was not a significant difference, $p = .195$.

For statement 15, high-stakes tests promote greater homogeneity of education, there was not a significant difference, $p = .932$.

For statement 16, high-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions, there was not a significant difference, $p = .278$.

For statement 17, high-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even

challenging multiple-choice items. This has led to teachers enhancing their own assessment practices, there was not a significant difference, $p = .491$.

For statement 18, high-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time, there was not a significant difference, $p = .409$.

For statement 19, high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels, there was not a significant difference, $p = .385$. Independent t-tests for statements 13 – 19 by years of administrative experience (1 – 4 years vs. 15 or more years) are summarized in Table 4.35.

Table 4.35—Results of t-tests by Years of Administrative Experience for Statements 13 – 19

	Statement	Equal Variances	t	df	Sig. (2-tailed)
13.	One result of high-stakes testing is that educators know more about testing than ever before.	Assumed	.594	79	.554
		Not Assumed	.658	68.130	.513
14.	Prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled.	Assumed	1.248	79	.215
		Not Assumed	1.311	59.470	.195
15.	High-stakes tests promote greater homogeneity of education.	Assumed	-.087	79	.931
		Not Assumed	-.086	50.748	.932
16.	High-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions.	Assumed	-1.026	79	.308
		Not Assumed	-1.095	61.900	.278

Table 4.35—Continued

	Statement	Equal Variances	t	df	Sig. (2-tailed)
17.	High-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has lead to teachers enhancing their own assessment practices.	Assumed	-.657	79	.513
		Not Assumed	-.693	60.193	.491
18.	High-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time.	Assumed	.789	79	.433
		Not Assumed	.831	60.002	.409
19.	High-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels.	Assumed	.842	79	.402
		Not Assumed	.875	57.721	.385

Means, standard deviation, and N were calculated for current campus rating. For statement 13, one result of high-stakes testing is that educators know more about testing than ever before, the mean for exemplary and recognized campus rating was 2.26 and the standard deviation was .82 while the mean for academically acceptable campus rating was 2.06 and the standard deviation was .89.

For statement 14, prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled, the mean for exemplary and recognized campus rating was 2.35 and the standard deviation was .83 while the mean for academically acceptable campus rating was 2.21 and the standard deviation was .76.

For statement 15, high-stakes tests promote greater homogeneity of education, the mean for exemplary and recognized campus rating was 2.47 and the standard deviation was .86 while the mean for academically acceptable campus rating was 2.55 and the standard deviation was .94.

For statement 16, high-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions, the mean for exemplary and recognized campus rating was 3.60 and the standard deviation was 1.00 while the mean for academically acceptable campus rating was 3.66 and the standard deviation was 1.03.

For statement 17, high-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has led to teachers enhancing their own assessment practices, the mean for exemplary and recognized campus rating was 2.93 and the standard deviation was 1.10 while the mean for academically acceptable campus rating was 2.66 and the standard deviation was 1.04.

For statement 18, high-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time, the mean for exemplary and recognized campus rating was 2.12 and the standard deviation was 1.18 while the mean for academically acceptable campus rating was 1.94 and the standard deviation was 1.00.

For statement 19, high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels, the mean for exemplary and recognized campus rating was 2.09 and the standard deviation was 1.15 while the mean for academically acceptable campus rating was 2.10 and the standard deviation was 1.13. The means and standard deviations for statements 13 – 19 by current campus rating are summarized in Table 4.36.

Table 4.36—Means and Standard Deviations by Current Campus Rating for Statements 13 – 19

	Statement	Rating	N	M	SD
13.	One result of high-stakes testing is that educators know more about testing than ever before.	Exemplary and Recognized	43	2.26	.819
		Academically Acceptable	126	2.06	.888
14.	Prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled.	Exemplary and Recognized	43	2.35	.831
		Academically Acceptable	126	2.21	.755
15.	High-stakes tests promote greater homogeneity of education.	Exemplary and Recognized	43	2.47	.855
		Academically Acceptable	126	2.55	.943
16.	High-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions.	Exemplary and Recognized	43	3.60	1.003
		Academically Acceptable	126	3.66	1.029
		Exemplary and Recognized	43	2.93	1.100
17.	High-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has led to teachers enhancing their own assessment practices.	Exemplary and Recognized	43	2.93	1.100
		Academically Acceptable	126	2.66	1.044
18.	High-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time.	Exemplary and Recognized	43	2.12	1.179
		Academically Acceptable	126	1.94	1.002
19.	High-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels.	Exemplary and Recognized	43	2.09	1.151
		Academically Acceptable	126	2.10	1.127

Independent t-tests were calculated for current campus rating (exemplary and recognized vs. academically acceptable). For statement 13, one result of high-stakes testing is that educators know more about testing than ever before, there was not a significant difference, $p = .180$.

For statement 14, prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled, there was not a significant difference, $p = .344$.

For statement 15, high-stakes tests promote greater homogeneity of education, there was not a significant difference, $p = .596$.

For statement 16, high-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions, there was not a significant difference, $p = .763$.

For statement 17, high-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has led to teachers enhancing their own assessment practices, there was not a significant difference, $p = .161$.

For statement 18, high-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time, there was not a significant difference, $p = .374$.

For statement 19, high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels, there was not a significant difference, $p = .991$. Independent t-tests for statements 13 – 19 by current campus rating (exemplary and recognized vs. academically acceptable) are summarized in Table 4.37.

Table 4.37—Results of t-tests by Current Campus Rating for Statements 13 – 19

	Statement	Equal Variances	t	df	Sig. (2-tailed)
13.	High-stakes tests have helped focus public attention on schools with low-achieving students.	Assumed	1.301	167	.195
		Not Assumed	1.354	78.228	.180
14.	High-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach.	Assumed	.990	167	.324
		Not Assumed	.954	68.349	.344
15.	High-stakes tests have helped close the gap in achievement between minority students and our majority students in Texas.	Assumed	-.507	167	.613
		Not Assumed	-.532	79.532	.596
16.	Teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform.	Assumed	-.299	167	.765
		Not Assumed	-.303	74.339	.763
17.	Doing poorly on high-stakes tests will lead to increased student effort to learn.	Assumed	1.452	167	.148
		Not Assumed	1.416	69.616	.161
18.	Students work harder and learn more because they know what is expected and that the high-stakes tests really count.	Assumed	.970	167	.333
		Not Assumed	.895	63.936	.374
19.	The public display of high-stakes tests scores motivate administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught.	Assumed	-.011	167	.991
		Not Assumed	-.011	71.436	.991

Research Question 3

Question 3: Do perceptions of high-stakes testing critics differ from those of Texas secondary school administrators in the membership of TASSP?

To examine research question 3, means and standard deviations were calculated on statements 20 – 31 to determine to what extent the perceptions of critics of high-stakes testing differ from those of Texas secondary school administrators in the membership of TASSP.

For statement 20, a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district, the mean was 2.18 and the standard deviation was 1.03.

For statement 21, high-stakes testing compromises educational quality by leading educators to 'teach to the test,' which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment, the mean was 2.28 and the standard deviation was 1.14.

For statement 22, high-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems, the mean was 2.70 and the standard deviation was 1.10.

For statement 23, a focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction, the mean was 2.29 and the standard deviation was .98.

For statement 24, pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating, the mean was 2.87 and the standard deviation was 1.03.

For statement 25, high-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve, the mean was 2.26 and the standard deviation was .98.

For statement 26, Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children, the standard deviation was 2.39 and the maximum was 1.04.

For statement 27, high-stakes testing and the accompanying consequences of failure lead to overstressed students, the mean was 2.10 and the standard deviation was 1.04.

For statement 28, the pressures inherent in preparing students for high-stakes tests are driving out good teachers, the mean was 2.25 and the standard deviation was 1.09.

For statement 29, high-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language, the mean was 2.07 and the standard deviation was .98.

For statement 30, the high stakes testing movement is resulting in a significant increase in student dropout rates, the mean was 2.58 and the standard deviation was 1.03.

For statement 31, no high stakes decision such as grade retention or graduation should be based on the results of a single test, the mean was 1 and the maximum was 5 ($M = 1.90$, $SD = 1.11$). The means and standard deviations for statements 20 – 31 are found in Table 4.38.

Virtually 70% of respondents found some degree of agreement with the statement that a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district. Over 38% of respondents strongly agreed with the preceding statement in stark contrast with the paltry 14.6% of respondents that were in disagreement with the majority. This information is presented in Table 4.39.

Table 4.38—Means and Standard Deviations for Statements 20 – 31

	Statement	N	Min.	Max.	M	SD
20.	A test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district.	178	1	5	2.18	1.03
21.	High-stakes testing compromises educational quality by leading educators to ‘teach to the test,’ which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment.	178	1	5	2.28	1.14
22.	High-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems.	178	1	5	2.70	1.10
23.	A focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction.	178	1	5	2.29	.98
24.	Pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	178	1	5	2.87	1.03
25.	High-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve.	178	1	5	2.26	.98
26.	Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children.	178	1	5	2.39	1.04
27.	High-stakes testing and the accompanying consequences of failure lead to overstressed students.	178	1	5	2.10	1.04
28.	The pressures inherent in preparing students for high-stakes tests are driving out good teachers.	178	1	5	2.25	1.09
29.	High-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language.	178	1	5	2.07	.98
30.	The high-stakes testing movement is resulting in a significant increase in student dropout rates.	178	1	5	2.58	1.03
31.	No high-stakes decision such as grade retention or graduation should be based on the results of a single test.	178	1	5	1.90	1.11

Table 4.39—Frequency Distribution of Responses to Statement Number 20: A Test That Has Been Validated Only for Diagnosing Strengths and Weaknesses of Individual Students Should Not Be Used to Evaluate the Educational Quality of a School or School District

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	50	28.1	28.1
Agree	74	41.6	69.7
Unsure	28	15.7	85.4
Disagree	24	13.6	98.9
Strongly Disagree	2	1.1	100.0
Total	178	100.0	

In excess of 71% of respondents agreed with the position that high-stakes testing leads educators to “teach to the test,” which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessments. This information is presented in Table 4.40.

Table 4.40—Frequency Distribution of Responses to Statement Number 21: High-stakes Testing Compromises Educational Quality by Leading Educators to “Teach to the Test,” Which Results in a Narrowing of the Curriculum, Limiting the Scope of Tested Subjects and Shortchanging or Eliminating Subjects Not Included in the Assessments

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	48	27.0	27.0
Agree	79	44.4	71.3
Unsure	8	4.5	75.8
Disagree	40	22.5	98.3
Strongly Disagree	3	1.7	100.0
Total	178	100.0	

The survey findings revealed that a simple majority of respondents agreed with the statement that high-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems. Surprisingly, practically 20% of administrators were unsure about this issue. This information is presented in Table 4.41.

Table 4.41—Frequency Distribution of Responses to Statement Number 22: High-stakes Tests Are Too Expensive and Result in Diverting Scarce Resources and Attention from Serious Problems

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	23	12.9	12.9
Agree	67	37.6	50.6
Unsure	35	19.7	70.2
Disagree	47	26.4	96.6
Strongly Disagree	6	3.4	100.0
Total	178	100.0	

A resounding 72% of respondents concur that a focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction. Only 17.4% opposed this position. This information is presented in Table 4.42.

Table 4.42—Frequency Distribution of Responses to Statement Number 23: A Focus on Standards and Accountability That Ignores the Processes of Teaching and Learning in Classrooms Will Not Provide the Direction That Teachers Need in Their Quest to Improve Instruction

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	32	18.0	18.0
Agree	96	53.9	71.9
Unsure	19	10.7	82.6
Disagree	29	16.3	98.9
Strongly Disagree	2	1.1	100.0
Total	178	100.0	

One of the most ambiguous results of the survey resulted from the statement of how pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating. Approximately 31% of respondents were unsure about this issue. Almost 40% were in agreement with the statement compared to only 29.3% that disagreed to some degree with the item. This information is presented in Table 4.43.

Table 4.43—Frequency Distribution of Responses to Statement Number 24: Pressure Exerted from the Need to Succeed on High-stakes Tests Often Leads to Inappropriate Test Preparation Practices, Including Outright Cheating

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	13	7.3	7.3
Agree	58	32.6	39.9
Unsure	55	30.9	70.8
Disagree	43	24.2	94.9
Strongly Disagree	9	5.1	100.0
Total	178	100.0	

In excess of 71% of respondents answered in the affirmative with regard to the statement that high-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve. In stark contrast, only 16.3% found a level of disagreement with just over 1% strongly disagreeing with this stance. This information is presented in Table 4.44.

Table 4.44—Frequency Distribution of Responses to Statement Number 25: High-stakes Tests Draw an Inaccurate Picture of Student Achievement and Unfairly Jeopardize Students or Schools That Are Making Genuine Efforts to Improve

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	35	19.7	19.7
Agree	92	51.7	71.3
Unsure	22	12.4	83.7
Disagree	27	15.2	98.9
Strongly Disagree	2	1.1	100.0
Total	178	100.0	

Virtually 62% of responders agreed to some degree with the statement that educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children. Only one respondent strongly agreed with this outlook and joined the 20% that concurred with their minority position. This information is presented in Table 4.45.

Table 4.45—Frequency Distribution of Responses to Statement Number 26: Educational Decisions Based on High-stakes Tests Have a Disproportionate Impact on Poor and Minority Children

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	36	20.2	20.2
Agree	74	41.6	61.8
Unsure	31	17.4	79.2
Disagree	36	20.2	99.4
Strongly Disagree	1	0.6	100.0
Total	178	100.0	

More than 75% of respondents agreed to some extent with the position that high-stakes testing and the accompanying consequences of failure lead to overstressed students. Almost 31% strongly agreed with this view in contrast to the mere 15% of respondents who disagreed to some degree. This information is presented in Table 4.46.

Table 4.46—Frequency Distribution of Responses to Statement Number 27: High-stakes Testing and the Accompanying Consequences of Failure Lead to Overstressed Students

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	55	30.9	30.9
Agree	80	44.9	75.8
Unsure	16	9.0	84.8
Disagree	24	13.5	98.3
Strongly Disagree	3	1.7	100.0
Total	178	100.0	

Virtually 70% of responders agreed to some extent that the pressures inherent in preparing students for high-stakes tests are driving out good teachers. Only 20% of responders held a contrary opinion. This information is presented in Table 4.47.

Table 4.47—Frequency Distribution of Responses to Statement Number 28: The Pressures Inherent in Preparing Students for High-stakes Tests Are Driving out Good Teachers

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	48	27.0	27.0
Agree	76	42.7	69.7
Unsure	19	10.7	80.3
Disagree	32	18.0	98.8
Strongly Disagree	3	1.7	100.0
Total	178	100.0	

More than 76% of respondents to some degree supported the view that high-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language. In contrast, only 10.6% of respondents disagreed with this position. This information is presented in Table 4.48.

Table 4.48—Frequency Distribution of Responses to Statement Number 29: High-stakes Tests Unfairly and Inaccurately Assess and Penalize Learners for Whom English Is Not Their First Language

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	53	29.8	29.8
Agree	83	46.6	76.4
Unsure	23	12.9	89.3
Disagree	15	8.4	97.8
Strongly Disagree	4	2.2	100.0
Total	178	100.0	

Almost a third of all respondents were unsure of their position on the statement that high-stakes testing movement is resulting in a significant increase in student dropout rates. This resulted in the highest level of ambiguity of any survey item. While virtually 49% of respondents agreed to some extent to this view, just 19.7% found a level of disagreement. This information is presented in Table 4.49.

Table 4.49—Frequency Distribution of Responses to Statement Number 30: The High-stakes Testing Movement is Resulting in a Significant Increase in Student Dropout Rates

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	27	15.2	15.2
Agree	60	33.7	48.9
Unsure	56	31.5	80.3
Disagree	30	16.9	97.2
Strongly Disagree	5	2.8	100.0
Total	178	100.0	

Overwhelmingly, 80.9% of respondents agreed to some extent to the position that no high-stakes decision such as grade retention or graduation should be based on the results of a single test. This item resulted in the highest percentage of respondents strongly agreeing to this view then any other survey item. In contrast, a paltry 11.8% found a level of disagreement with this statement. This information is presented in Table 4.50.

Table 4.50—Frequency Distribution of Responses to Statement Number 31: No High-stakes Decision Such as Grade Retention or Graduation Should Be Based on the Results of a Single Test

Response	Frequency	Percent	Cumulative Percent
Strongly Agree	81	45.5	45.5
Agree	63	35.4	80.9
Unsure	13	7.3	88.2
Disagree	13	7.3	95.5
Strongly Disagree	8	4.5	100.0
Total	178	100.0	

To examine research question 3, four sets of independent t-tests were conducted to test for a difference between the means of two independent groups for statements 20 – 31. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years vs. 15 or More Years), Campus

Classification (Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable). Means, standard deviation, and N were calculated for all statements in regard to campus classification. For statement 20, a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district, the mean for small school administrators was 1.95 and the standard deviation was .94 while the mean for large school administrators was 2.30 and the standard deviation was 1.05.

For statement 21, high-stakes testing compromises educational quality by leading educators to ‘teach to the test,’ which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment, the mean for small school administrators was 2.10 and the standard deviation was 1.10 while the mean for large school administrators was 2.37 and the standard deviation was 1.15.

For statement 22, high-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems, the mean for small school administrators was 2.51 and the standard deviation was 1.08 while the mean for large school administrators was 2.80 and the standard deviation was 1.10.

For statement 23, a focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction, the mean for small school administrators was 2.35 and the standard deviation was .99 while the mean for large school administrators was 2.25 and the standard deviation was .98.

For statement 24, pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating, the mean for small school administrators was 2.84 and the standard deviation was .92 while the mean for large school administrators was 2.89 and the standard deviation was 1.08.

For statement 25, high-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve, the mean for small school administrators was 2.22 and the standard deviation was .98 while the mean for large school administrators was 2.29 and the standard deviation was .99.

For statement 26, Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children, the mean for small school administrators was 2.43 and the standard deviation was .95 while the mean for large school administrators was 2.37 and the standard deviation was 1.10.

For statement 27, high-stakes testing and the accompanying consequences of failure lead to overstressed students, the mean for small school administrators was 1.94 and the standard deviation was .91 while the mean for large school administrators was 2.19 and the standard deviation was 1.10.

For statement 28, the pressures inherent in preparing students for high-stakes tests are driving out good teachers, the mean for small school administrators was 2.21 and the standard deviation was 1.08 while the mean for large school administrators was 2.27 and the standard deviation was 1.10.

For statement 29, high-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language, the mean for small school

administrators was 2.13 and the standard deviation was .87 while the mean for large school administrators was 2.03 and the standard deviation was 1.04.

For statement 30, the high stakes testing movement is resulting in a significant increase in student dropout rates, the mean for small school administrators was 2.68 and the standard deviation was .86 while the mean for large school administrators was 2.53 and the standard deviation was 1.11.

For statement 31, no high stakes decision such as grade retention or graduation should be based on the results of a single test, the mean for small school administrators was 1.90 and the standard deviation was 1.09 while the mean for large school administrators was 1.90 and the standard deviation was 1.12. The means and standard deviations for statements 20 – 31 by campus classification are summarized in Table 4.51.

Table 4.51—Means and Standard Deviations by Campus Classification for Statements 20 – 31

	Statement	Classification	N	M	SD
20.	A test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district.	Small	63	1.95	.941
		Large	115	2.30	1.053
21.	High-stakes testing compromises educational quality by leading educators to ‘teach to the test,’ which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment.	Small	63	2.10	1.103
		Large	115	2.37	1.151
22.	High-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems.	Small	63	2.51	1.076
		Large	115	2.80	1.102
23.	A focus on standards and accountability that ignores high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Small	63	2.35	.986
		Large	115	2.25	.981

Table 4.51—Continued

Statement	Classification	N	M	SD	
24. Pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Small		63	2.84	.919
	Large		115	2.89	1.082
25. High-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve.	Small		63	2.22	.975
	Large		115	2.29	.989
26. Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children.	Small		63	2.43	.946
	Large		115	2.37	1.096
27. High-stakes testing and the accompanying consequences of failure lead to overstressed students.	Small		63	1.94	.914
	Large		115	2.19	1.099
28. The pressures inherent in preparing students for high-stakes tests are driving out good teachers.	Small		63	2.21	1.080
	Large		115	2.27	1.103
29. High-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language.	Small		63	2.13	.871
	Large		115	2.03	1.042
30. The high-stakes testing movement is resulting in a significant increase in student dropout rates.	Small		63	2.68	.858
	Large		115	2.53	1.111
31. No high-stakes decision such as grade retention or graduation should be based on the results of a single test.	Small		63	1.90	1.088
	Large		115	1.90	1.119

To examine research question 3, four sets of independent t-tests were conducted to test for a difference between the means of two independent groups for statements 20 – 31. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years vs. 15 or More Years), Campus Classification (Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable). Independent t-tests were calculated for campus classification (large and small). For statement 20, a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district, there was a significant difference, $p = .024$.

For statement 21, high-stakes testing compromises educational quality by leading educators to 'teach to the test,' which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment, there was not a significant difference, $p = .115$.

For statement 22, high-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems, there was not a significant difference, $p = .088$.

For statement 23, a focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction, there was not a significant difference, $p = .531$.

For statement 24, pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating, there was not a significant difference, $p = .767$.

For statement 25, high-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve, there was not a significant difference, $p = .674$.

For statement 26, Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children, there was not a significant difference, $p = .728$.

For statement 27, high-stakes testing and the accompanying consequences of failure lead to overstressed students, there was not a significant difference, $p = .100$.

For statement 28, the pressures inherent in preparing students for high-stakes tests are driving out good teachers, there was not a significant difference, $p = .712$.

For statement 29, high-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language, there was not a significant difference, $p = .530$.

For statement 30, the high stakes testing movement is resulting in a significant increase in student dropout rates, there was not a significant difference, $p = .311$.

For statement 31, no high stakes decision such as grade retention or graduation should be based on the results of a single test, there was not a significant difference, $p = .958$. Independent t-tests for statements 20 – 31 by campus classification (large and small) are summarized in Table 4.52.

Table 4.52—Results of t-tests by Campus Classification for Statements 20 – 31

	Statements	Equal Variances	t	df	Sig (2-tailed)
20.	A test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district.	Assumed	-2.213	176	.028
		Not Assumed	02.287	140.345	.024*
21.	High-stakes testing compromises educational quality by leading educators to ‘teach to the test,’ which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment.	Assumed	-1.568	176	.119
		Not Assumed	-1.587	132.370	.115
22.	High-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems.	Assumed	-1.705	176	.090
		Not Assumed	-1.717	130.333	.088
23.	A focus on standards and accountability that ignores high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Assumed	.630	176	.530
		Not Assumed	.629	127.093	.531
24.	Pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Assumed	-.284	176	.777
		Not Assumed	-.297	146.073	.767

Table 4.52—Continued

	Statements	Equal Variances	t	df	Sig (2-tailed)
25.	High-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve.	Assumed	-.420	176	.675
		Not Assumed	.422	129.317	.674
26.	Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children.	Assumed	.334	176	.739
		Not Assumed	.348	144.328	.728
27.	High-stakes testing and the accompanying consequences of failure lead to overstressed students.	Assumed	-1.566	176	.119
		Not Assumed	-1.653	148.550	.100
28.	The pressures inherent in preparing students for high-stakes tests are driving out good teachers.	Assumed	-.368	176	.713
		Not Assumed	-.371	129.991	.712
29.	High-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language.	Assumed	.597	176	.551
		Not Assumed	.629	147.990	.530
30.	The high-stakes testing movement is resulting in a significant increase in student dropout rates.	Assumed	.943	176	.347
		Not Assumed	1.016	156.394	.311
31.	No high-stakes decision such as grade retention or graduation should be based on the results of a single test.	Assumed	.052	176	.958
		Not Assumed	.053	130.763	.958

*Significant at the .05 level.

Means, standard deviation, and N were calculated for gender. For statement 20, a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district, the mean for male was 2.15 and the standard deviation was 1.01 while the mean for female was 2.24 and the standard deviation was 1.05.

For statement 21, high-stakes testing compromises educational quality by leading educators to ‘teach to the test,’ which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment, the mean for male was 2.22 and the standard deviation was 1.10 while the mean for female was 2.37 and the standard deviation was 1.21.

For statement 22, high-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems, the mean for male was 2.66 and the standard deviation was 1.10 while the mean for female was 2.75 and the standard deviation was 1.10.

For statement 23, a focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction, the mean for male was 2.34 and the standard deviation was 1.01 while the mean for female was 2.21 and the standard deviation was .94.

For statement 24, pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating, the mean for male was 2.91 and the standard deviation was 1.05 while the mean for female was 2.81 and the standard deviation was .98.

For statement 25, high-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve, the mean for male was 2.22 and the standard deviation was .94 while the mean for female was 2.34 and the standard deviation was 1.05.

For statement 26, Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children, the mean for male was 2.42 and the standard deviation was .98 while the mean for female was 2.35 and the standard deviation was 1.14.

For statement 27, high-stakes testing and the accompanying consequences of failure lead to overstressed students, the mean for male was 2.03 and the standard

deviation was .99 while the mean for female was 2.22 and the standard deviation was 1.12.

For statement 28, the pressures inherent in preparing students for high-stakes tests are driving out good teachers, the mean for male was 2.24 and the standard deviation was 1.08 while the mean for female was 2.26 and the standard deviation was 1.13.

For statement 29, high-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language, the mean for male was 2.05 and the standard deviation was .97 while the mean for female was 2.09 and the standard deviation was 1.02.

For statement 30, the high stakes testing movement is resulting in a significant increase in student dropout rates, the mean for male was 2.59 and the standard deviation was 1.00 while the mean for female was 2.57 and the standard deviation was 1.08.

For statement 31, no high stakes decision such as grade retention or graduation should be based on the results of a single test, the mean for male was 1.84 and the standard deviation was 1.04 while the mean for female was 2.00 and the standard deviation was 1.21. The means and standard deviations for statements 20 – 31 by classification are summarized in Table 4.53.

Table 4.53—Means and Standard Deviations by Gender for Statements 20 – 31

	Statement	Gender	N	M	SD
20.	A test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district.	Male	110	2.15	1.012
		Female	68	2.24	1.053
21.	High-stakes testing compromises educational quality by leading educators to ‘teach to the test,’ which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment.	Male	110	2.22	1.095
		Female	68	2.37	1.208
22.	High-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems.	Male	110	2.66	1.103
		Female	68	2.75	1.098
23.	A focus on standards and accountability that ignores high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Male	110	2.34	1.007
		Female	68	2.21	.939
24.	Pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Male	110	2.91	1.054
		Female	68	2.81	.981
25.	High-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve.	Male	110	2.22	.942
		Female	68	2.34	1.045
26.	Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children.	Male	110	2.42	.980
		Female	68	2.35	1.143
27.	High-stakes testing and the accompanying consequences of failure lead to overstressed students.	Male	110	2.03	.990
		Female	68	2.22	1.118
28.	The pressures inherent in preparing students for high-stakes tests are driving out good teachers.	Male	110	2.24	1.075
		Female	68	2.26	1.128
29.	High-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language.	Male	110	2.05	.966
		Female	68	2.09	1.018
30.	The high-stakes testing movement is resulting in a significant increase in student dropout rates.	Male	110	2.59	.998
		Female	68	2.57	1.083
31.	No high-stakes decision such as grade retention or graduation should be based on the results of a single test.	Male	110	1.84	1.036
		Female	68	2.00	1.209

Independent t-tests were calculated for gender (male and female). For statement 20, a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district, there was not a significant difference, $p = .575$.

For statement 21, high-stakes testing compromises educational quality by leading educators to 'teach to the test,' which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment, there was not a significant difference, $p = .408$.

For statement 22, high-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems, there was not a significant difference, $p = .611$.

For statement 23, a focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction, there was not a significant difference, $p = .383$.

For statement 24, pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating, there was not a significant difference, $p = .521$.

For statement 25, high-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve, there was not a significant difference, $p = .441$.

For statement 26, Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children, there was not a significant difference, $p = .697$.

For statement 27, high-stakes testing and the accompanying consequences of failure lead to overstressed students, there was not a significant difference, $p = .244$.

For statement 28, the pressures inherent in preparing students for high-stakes tests are driving out good teachers, there was not a significant difference, $p = .869$.

For statement 29, high-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language, there was not a significant difference, $p = .827$.

For statement 30, the high stakes testing movement is resulting in a significant increase in student dropout rates, there was not a significant difference, $p = .915$.

For statement 31, no high stakes decision such as grade retention or graduation should be based on the results of a single test, there was not a significant difference, $p = .357$. Independent t-tests for statements 20 – 31 by gender (male and female) are summarized in Table 4.54.

Table 4.54—Results of t-tests by Gender for Statements 20 – 31

	Statement	Equal Variances	t	df	Sig. (2-tailed)
20.	A test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district.	Assumed	-.567	176	.572
		Not Assumed	-.561	137.810	.575
21.	High-stakes testing compromises educational quality by leading educators to 'teach to the test,' which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment.	Assumed	-.850	176	.396
		Not Assumed	-.831	131.504	.408
22.	High-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems.	Assumed	-.509	176	.612
		Not Assumed	-.509	142.558	.611
23.	A focus on standards and accountability that ignores high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Assumed	.861	176	.309
		Not Assumed	.876	149.614	.383
24.	Pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Assumed	.633	176	.528
		Not Assumed	.644	149.810	.521
25.	High-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve.	Assumed	-.792	176	.429
		Not Assumed	-.773	130.896	.441
26.	Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children.	Assumed	.405	176	.686
		Not Assumed	.309	125.806	.697
27.	High-stakes testing and the accompanying consequences of failure lead to overstressed students.	Assumed	-1.204	176	.230
		Not Assumed	-1.170	129.133	.244
28.	The pressures inherent in preparing students for high-stakes tests are driving out good teachers.	Assumed	-.168	176	.867
		Not Assumed	-.166	136.803	.869
29.	High-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language.	Assumed	-.221	176	.825
		Not Assumed	-.219	136.325	.827
30.	The high-stakes testing movement is resulting in a significant increase in student dropout rates.	Assumed	.109	176	.913
		Not Assumed	.107	133.218	.915
31.	No high-stakes decision such as grade retention or graduation should be based on the results of a single test.	Assumed	-.960	176	.339
		Not Assumed	-.925	125.703	.357

Means, standard deviation, and N were calculated for years of administrative experience. For statement 20, a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district, the mean for 1 – 4 years of administrative experiences was 2.19 and the standard deviation was 1.03 while the mean for 15 or more years of administrative experiences was 2.11 and the standard deviation was .93.

For statement 21, high-stakes testing compromises educational quality by leading educators to ‘teach to the test,’ which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment, the mean for 1 – 4 years of administrative experiences was 2.20 and the standard deviation was 1.12 while the mean for 15 or more years of administrative experiences was 2.37 and the standard deviation was .97.

For statement 22, high-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems, the mean for 1 – 4 years of administrative experiences was 2.70 and the standard deviation was 1.08 while the mean for 15 or more years of administrative experiences was 2.74 and the standard deviation was .98.

For statement 23, a focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction, the mean for 1 – 4 years of administrative experiences was 2.43 and the standard deviation was 1.00 while the mean

for 15 or more years of administrative experiences was 2.33 and the standard deviation was 1.04.

For statement 24, pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating, the mean for 1 – 4 years of administrative experiences was 2.83 and the standard deviation was .91 while the mean for 15 or more years of administrative experiences was 2.74 and the standard deviation was 1.06.

For statement 25, high-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve, the mean for 1 – 4 years of administrative experiences was 2.31 and the standard deviation was .97 while the mean for 15 or more years of administrative experiences was 2.30 and the standard deviation was .91.

For statement 26, Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children, the mean for 1 – 4 years of administrative experiences was 2.41 and the standard deviation was 1.07 while the mean for 15 or more years of administrative experiences was 2.56 and the standard deviation was 1.01.

For statement 27, high-stakes testing and the accompanying consequences of failure lead to overstressed students, the mean for 1 – 4 years of administrative experiences was 2.17 and the standard deviation was .97 while the mean for 15 or more years of administrative experiences was 2.22 and the standard deviation was 1.12.

For statement 28, the pressures inherent in preparing students for high-stakes tests are driving out good teachers, the mean for 1 – 4 years of administrative

experiences was 2.20 and the standard deviation was 1.02 while the mean for 15 or more years of administrative experiences was 1.96 and the standard deviation was 1.13.

For statement 29, high-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language, the mean for 1 – 4 years of administrative experiences was 2.11 and the standard deviation was 1.02 while the mean for 15 or more years of administrative experiences was 2.07 and the standard deviation was 1.07.

For statement 30, the high stakes testing movement is resulting in a significant increase in student dropout rates, the mean for 1 – 4 years of administrative experiences was 2.59 and the standard deviation was 1.07 while the mean for 15 or more years of administrative experiences was 2.56 and the standard deviation was 1.05.

For statement 31, no high stakes decision such as grade retention or graduation should be based on the results of a single test, the mean for 1 – 4 years of administrative experiences was 1.94 and the standard deviation was .98 while the mean for 15 or more years of administrative experiences was 1.89 and the standard deviation was 1.05. The means and standard deviations for statements 20 – 31 by gender are summarized in Table 4.55.

Table 4.55—Means and Standard Deviations by Years of Administrative Experience for Statements 20 – 31

		Administrative Experience (years)	N	M	SD
20.	A test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district.	1 – 4	54	2.19	1.029
		15 or more	27	2.11	.934
21.	High-stakes testing compromises educational quality by leading educators to ‘teach to the test,’ which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment.	1 – 4	54	2.20	1.122
		15 or more	27	2.37	.967
22.	High-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems.	1 – 4	54	2.70	1.075
		15 or more	27	2.74	.984
23.	A focus on standards and accountability that ignores high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	1 – 4	54	2.43	1.002
		15 or more	27	2.33	1.038
24.	Pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	1 – 4	54	2.83	.906
		15 or more	27	2.74	1.059
25.	High-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve.	1 – 4	54	2.31	.968
		15 or more	27	2.30	.912
26.	Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children.	1 – 4	54	2.41	1.073
		15 or more	27	2.56	1.013
27.	High-stakes testing and the accompanying consequences of failure lead to overstressed students.	1 – 4	54	2.17	.966
		15 or more	27	2.22	1.121
28.	The pressures inherent in preparing students for high-stakes tests are driving out good teachers.	1 – 4	54	2.20	1.016
		15 or more	27	1.96	1.126
29.	High-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language.	1 – 4	54	2.11	1.022
		15 or more	27	2.07	1.072
30.	The high-stakes testing movement is resulting in a significant increase in student dropout rates.	1 – 4	54	2.59	1.073
		15 or more	27	2.56	1.050
31.	No high-stakes decision such as grade retention or graduation should be based on the results of a single test.	1 – 4	54	1.94	.979
		15 or more	27	1.89	1.050

Independent t-tests were calculated for years of administrative experience (1 – 4 years vs. 15 or more years). For statement 20, a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district, there was not a significant difference, $p = .746$.

For statement 21, high-stakes testing compromises educational quality by leading educators to ‘teach to the test,’ which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment, there was not a significant difference, $p = .491$.

For statement 22, high-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems, there was not a significant difference, $p = .878$.

For statement 23, a focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction, there was not a significant difference, $p = .703$.

For statement 24, pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating, there was not a significant difference, $p = .699$.

For statement 25, high-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve, there was not a significant difference, $p = .933$.

For statement 26, Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children, there was not a significant difference, $p = .545$.

For statement 27, high-stakes testing and the accompanying consequences of failure lead to overstressed students, there was not a significant difference, $p = .827$.

For statement 28, the pressures inherent in preparing students for high-stakes tests are driving out good teachers, there was not a significant difference, $p = .354$.

For statement 29, high-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language, there was not a significant difference, $p = .882$.

For statement 30, the high stakes testing movement is resulting in a significant increase in student dropout rates, there was not a significant difference, $p = .882$.

For statement 31, no high stakes decision such as grade retention or graduation should be based on the results of a single test, there was not a significant difference, $p = .819$. Independent t-tests for statements 20 – 31 by years of administrative experience (1 – 4 years vs. 15 or more years) are summarized in Table 4.56.

Table 4.56—Results of t-tests by Years of Administrative Experience for Statements 20 – 31

	Statement	Equal Variances	t	df	Sig. (2-tailed)
20.	A test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district.	Assumed	.315	79	.754
		Not Assumed	.325	56.890	.746
21.	High-stakes testing compromises educational quality by leading educators to ‘teach to the test,’ which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment.	Assumed	-.659	79	.512
		Not Assumed	-.692	59.580	.491
22.	High-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems.	Assumed	-.150	79	.881
		Not Assumed	-.155	56.426	.878
23.	A focus on standards and accountability that ignores high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Assumed	.387	79	.699
		Not Assumed	.383	50.502	.703
24.	Pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Assumed	.410	79	.683
		Not Assumed	.389	45.502	.699
25.	High-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve.	Assumed	.083	79	.934
		Not Assumed	.084	54.970	.933
26.	Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children.	Assumed	-.597	79	.553
		Not Assumed	-.608	54.900	.545
27.	High-stakes testing and the accompanying consequences of failure lead to overstressed students.	Assumed	-.231	79	.818
		Not Assumed	-.220	45.814	.827
28.	The pressures inherent in preparing students for high-stakes tests are driving out good teachers.	Assumed	.969	79	.335
		Not Assumed	.936	47.626	.354
29.	High-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language.	Assumed	.151	79	.880
		Not Assumed	.149	49.951	.882
30.	The high-stakes testing movement is resulting in a significant increase in student dropout rates.	Assumed	.147	79	.883
		Not Assumed	.149	53.139	.882
31.	No high-stakes decision such as grade retention or graduation should be based on the results of a single test.	Assumed	.235	79	.815
		Not Assumed	.230	48.990	.819

Means, standard deviation, and N were calculated for current campus rating. For statement 20, a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district, the mean for exemplary and recognized campus rating was 2.14 and the standard deviation was 1.08 while the mean for academically acceptable campus rating was 2.21 and the standard deviation was 1.03.

For statement 21, high-stakes testing compromises educational quality by leading educators to 'teach to the test,' which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment, the mean for exemplary and recognized campus rating was 2.33 and the standard deviation was 1.21 while the mean for academically acceptable campus rating was 2.28 and the standard deviation was 1.14.

For statement 22, high-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems, the mean for exemplary and recognized campus rating was 2.91 and the standard deviation was 1.11 while the mean for academically acceptable campus rating was 2.63 and the standard deviation was 1.09.

For statement 23, a focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction, the mean for exemplary and recognized campus rating was 2.30 and the standard deviation was .91 while the mean for academically acceptable campus rating was 2.31 and the standard deviation was 1.02.

For statement 24, pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating, the

mean for exemplary and recognized campus rating was 2.84 and the standard deviation was 1.05 while the mean for academically acceptable campus rating was 2.89 and the standard deviation was 1.02.

For statement 25, high-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve, the mean for exemplary and recognized campus rating was 2.53 and the standard deviation was 1.03 while the mean for academically acceptable campus rating was 2.21 and the standard deviation was .97.

For statement 26, Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children, the mean for exemplary and recognized campus rating was 2.51 and the standard deviation was 1.01 while the mean for academically acceptable campus rating was 2.40 and the standard deviation was 1.07.

For statement 27, high-stakes testing and the accompanying consequences of failure lead to overstressed students, the mean for exemplary and recognized campus rating was 2.19 and the standard deviation was .88 while the mean for academically acceptable campus rating was 2.10 and the standard deviation was 1.11.

For statement 28, the pressures inherent in preparing students for high-stakes tests are driving out good teachers, the mean for exemplary and recognized campus rating was 2.33 and the standard deviation was 1.15 while the mean for academically acceptable campus rating was 2.25 and the standard deviation was 1.09.

For statement 29, high-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language, the mean for exemplary and

recognized campus rating was 2.05 and the standard deviation was 1.00 while the mean for academically acceptable campus rating was 2.11 and the standard deviation was 1.00.

For statement 30, the high stakes testing movement is resulting in a significant increase in student dropout rates, the mean for exemplary and recognized campus rating was 2.88 and the standard deviation was .91 while the mean for academically acceptable campus rating was 2.55 and the standard deviation was 1.04.

For statement 31, no high stakes decision such as grade retention or graduation should be based on the results of a single test, the mean for exemplary and recognized campus rating was 2.09 and the standard deviation was 1.17 while the mean for academically acceptable campus rating was 1.87 and the standard deviation was 1.10.

The means and standard deviations for statements 20 – 31 by gender are summarized in Table 4.57.

Table 4.57—Means and Standard Deviations by Current Campus Rating for Statements 20 – 31

	Statement	Rating	N	M	SD
20.	A test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district.	Exemplary and Recognized	43	2.14	1.082
		Academically Acceptable	126	2.21	1.032
21.	High-stakes testing compromises educational quality by leading educators to 'teach to the test,' which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment.	Exemplary and Recognized	43	2.33	1.210
		Academically Acceptable	126	2.28	1.136
22.	High-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems.	Exemplary and Recognized	43	2.91	1.109
		Academically Acceptable	126	2.63	1.093
23.	A focus on standards and accountability that ignores high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Exemplary and Recognized	43	2.30	.914
		Academically Acceptable	126	2.31	1.023

Table 4.57—Continued

	Statement	Rating	N	M	SD
24.	Pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Exemplary and Recognized	43	2.84	1.045
		Academically Acceptable	126	2.89	1.022
25.	High-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve.	Exemplary and Recognized	43	2.53	1.032
		Academically Acceptable	126	2.21	.974
26.	Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children.	Exemplary and Recognized	43	2.51	1.009
		Academically Acceptable	126	2.40	1.066
27.	High-stakes testing and the accompanying consequences of failure lead to overstressed students.	Exemplary and Recognized	43	2.19	.880
		Academically Acceptable	126	2.10	1.113
28.	The pressures inherent in preparing students for high-stakes tests are driving out good teachers.	Exemplary and Recognized	43	2.33	1.149
		Academically Acceptable	126	2.25	1.086
29.	High-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language.	Exemplary and Recognized	43	2.05	.999
		Academically Acceptable	126	2.11	.998
30.	The high-stakes testing movement is resulting in a significant increase in student dropout rates.	Exemplary and Recognized	43	2.88	.905
		Academically Acceptable	126	2.55	1.040
31.	No high-stakes decision such as grade retention or graduation should be based on the results of a single test.	Exemplary and Recognized	43	2.09	1.171
		Academically Acceptable	126	1.87	1.103

Independent t-tests were calculated for current campus rating (exemplary and recognized vs. academically acceptable). For statement 20, a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district, there was not a significant difference, $p = .694$.

For statement 21, high-stakes testing compromises educational quality by leading educators to ‘teach to the test,’ which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment, there was not a significant difference, $p = .821$.

For statement 22, high-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems, there was not a significant difference, $p = .167$.

For statement 23, a focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction, there was not a significant difference, $p = .966$.

For statement 24, pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating, there was not a significant difference, $p = .779$.

For statement 25, high-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve, there was not a significant difference, $p = .072$.

For statement 26, Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children, there was not a significant difference, $p = .527$.

For statement 27, high-stakes testing and the accompanying consequences of failure lead to overstressed students, there was not a significant difference, $p = .588$.

For statement 28, the pressures inherent in preparing students for high-stakes tests are driving out good teachers, there was not a significant difference, $p = .692$.

For statement 29, high-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language, there was not a significant difference, $p = .715$.

For statement 30, the high stakes testing movement is resulting in a significant increase in student dropout rates, there was not a significant difference, $p = .046$.

For statement 31, no high stakes decision such as grade retention or graduation should be based on the results of a single test, there was not a significant difference, $p = .284$. Independent t-tests for statements 20 – 31 by current campus rating (exemplary and recognized vs. academically acceptable) are summarized in Table 4.58.

Table 4.58—Results of t-tests by Current Campus Rating for Statements 20 – 31

	Statement	Equal Variances	t	df	Sig. (2-tailed)
20.	A test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district.	Assumed	-.405	167	.686
		Not Assumed	-.396	69.887	.694
21.	High-stakes testing compromises educational quality by leading educators to ‘teach to the test,’ which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessment.	Assumed	.234	167	.815
		Not Assumed	.227	68.988	.821
22.	High-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems.	Assumed	1.405	167	.162
		Not Assumed	1.395	71.800	.167
23.	A focus on standards and accountability that ignores high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Assumed	-.041	167	.967
		Not Assumed	-.043	80.688	.966
24.	Pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.	Assumed	-.285	167	.776
		Not Assumed	-.282	71.316	.779
25.	High-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve.	Assumed	1.881	167	.062
		Not Assumed	1.829	69.298	.072
26.	Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children.	Assumed	.618	167	.538
		Not Assumed	.635	76.409	.527
27.	High-stakes testing and the accompanying consequences of failure lead to overstressed students.	Assumed	.485	167	.628
		Not Assumed	.544	91.260	.588
28.	The pressures inherent in preparing students for high-stakes tests are driving out good teachers.	Assumed	.409	167	.683
		Not Assumed	.397	69.341	.692

Table 4.58—Continued

	Statement	Equal Variances	t	df	Sig. (2-tailed)
29.	High-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language.	Assumed	-.366	167	.714
30.	The high-stakes testing movement is resulting in a significant increase in student dropout rates.	Not Assumed	-.366	72.642	.715
		Assumed	1.888	167	.061
31.	No high-stakes decision such as grade retention or graduation should be based on the results of a single test.	Not Assumed	2.022	82.722	.046*
		Assumed	1.112	167	.268
		Not Assumed	1.079	69.117	.284

*Significant at the .05 level.

Summary of Findings

Summary of Findings for Research Question 1 (Statements 1 – 21)

The researcher used means, standard deviations, and frequencies to make assumptions about perceptions of the population of secondary administrators differing from supporters' statements. Data from all three sources were combined to make an eyeball determination. The researcher determined that an agreement rate would represent the percentage of respondents that either strongly agreed or agreed to a survey statement. Conversely, a disagreement rate would represent the percentage of respondents that either strongly disagreed or disagreed to a survey statement.

The findings of this study suggest that perceptions of Texas secondary school administrators do match several of the statements derived from supporters' perceptions. With factors such as an 81% agreement rate, a mean of 2.15, and a standard deviation of .87, secondary administrators' perceptions appear to closely align with supporters' perceptions in regard to statement 10, educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success. Similarly, with an agreement rate of 80%, a mean of 2.16, and

a standard deviation of .90, administrators also appear to strongly agree with the supporters in regard to statement 7, the public display of high-stakes test scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught.

An agreement rate of 65%, a mean of 1.47, and a standard deviation of 1.05, suggests that secondary administrators' perceptions seem to complement statement 1, high-stakes tests have helped focus public attention on schools with low-achieving students. Correspondingly, an agreement rate of 65%, a mean of 2.64, and a standard deviation of 1.11, suggests that secondary administrators' perceptions also match up with statement 12, the implementation of high-stakes testing has been a catalyst for increased attention to students with special needs. Similarly, an agreement rate of 62%, a mean of 2.60, and a standard deviation of 1.02, suggests that secondary administrators' perceptions are in relative alignment with statement 11, driven by the demands of high-stakes tests, professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise.

The findings of this study suggest that perceptions of Texas secondary school administrators may strongly differ from a statement derived from supporters' perceptions. A disagreement rate of 79%, a mean of 3.94, and a standard deviation of .90, suggests that secondary administrators' perceptions may be at odds with statement 5, doing poorly on high-stakes tests will lead to increased student effort to learn.

Some of the findings of this study appear to be inconclusive when attempting to determine if perceptions of Texas secondary school administrators differ from several supporter statements. Factors such as marginal agreement or disagreement rates, centric

means, and larger standard deviations made arriving at a conclusion problematic for statements 2, 3, 4, 6, 8, and 9.

Independent t-tests were conducted to test for possible perception differences between groups identified in the study. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years vs. 15 or More Years), Campus Classification (Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable).

On statement 5, doing poorly on high-stakes tests will lead to increased student effort to learn, a significant difference in means ($p = .047$) was found between large and small school responders. With a mean of 4.11 and a standard deviation of .81, small school responders found a higher level of disagreement than large school responders with the statement.

On statement 7, the public display of high-stakes tests scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught, a significant difference in means ($p = .019$) was found between male and female responders. With a mean of 1.97 and a standard deviation of .80, small school responders found a higher level of agreement with the statement than large school responders.

On statement 7, the public display of high-stakes tests scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught, a significant difference in means ($p = .020$) was found between responders with 1 – 4 years of experience and responders with 15 years or more of experience. With a mean of 1.94 and a standard deviation of .79,

responders with 1 – 4 years of experience found a higher level of agreement with the statement than responders with 15 years or more of experience.

Summary of Findings for Research Question 2 (Statements 13 – 19)

The researcher used means, standard deviations, and frequencies to conjecture about perceptions of researchers of high-stakes testing reported in current literature and the corresponding perceptions of Texas secondary school administrators. Data from all three sources were combined to make an eyeball determination. The researcher determined that an agreement rate would represent the percentage of respondents that either strongly agreed or agreed to a survey statement. Conversely, a disagreement rate would represent the percentage of respondents that either strongly disagreed or disagreed to a survey statement.

The findings of this study suggest that perceptions of Texas secondary school administrators may strongly support several of the statements derived from unintended effects of high-stakes testing reported in current literature. An agreement rate of 80%, a mean of 2.11, and a standard deviation of .89, suggests that secondary administrators' perceptions seem highly aligned with statement 13, one result of high-stakes testing is that educators know more about testing than ever before. Similarly, an agreement rate of 80%, a mean of 1.96, and a standard deviation of 1.04, implies that secondary administrators perceptions seem closely associated with statement 18, high-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time. Likewise, an agreement rate of 75%, a mean of 2.09, and a standard deviation of 1.15, intimates that secondary administrators' perceptions seem closely

coupled with statement 19, high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels. Equally, an agreement rate of 74%, a mean of 2.24, and a standard deviation of .89, advocates that secondary administrators' perceptions seem closely tied with statement 14, prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled.

The findings of this study suggest that perceptions of Texas secondary school administrators may differ from a statement derived from unintended effects of high-stakes testing as reported in current literature. A disagreement rate of 65%, a mean of 3.64, and a standard deviation of 1.02, suggests that secondary administrators' perceptions may be at odds with statement 16, high-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions.

Some of the findings of this study appear to be inconclusive when attempting to determine if perceptions of Texas secondary school administrators differ from a unintended effects of high-stakes testing reported in current literature. Factors such as marginal agreement or disagreement rates, centric means, and larger standard deviations made arriving at a conclusion problematic for statements 15 and 17.

Independent t-tests were conducted to test for possible perception differences between groups identified in the study. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years

vs. 15 or More Years), Campus Classification (Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable).

On statement 18, high-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time, a significant difference in means ($p = .028$) was found between large and small school responders. With a mean of 1.73 and a standard deviation of .97, small school responders found a higher level of agreement with the statement than large school responders.

On statement 19, high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels, a significant difference in means ($p = .039$) was found between large and small school responders. With a mean of 1.86 and a standard deviation of 1.06, small school responders found a higher level of agreement with the statement than large school responders.

On statement 19, high-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels, a significant difference in means ($p = .050$) was found between male and female responders. With a mean of 1.95 and a standard deviation of 1.10, male responders found a higher level of agreement with the statement than female responders.

Summary of Findings for Research Question 3 (Statements 20 – 31)

The researcher used means, standard deviations, and frequencies to make assumptions about perceptions of the population of secondary administrators differing

from critics' perceptions. Data from of all three sources were combined to make an eyeball determination. The researcher determined that an agreement rate would represent the percentage of respondents that either strongly agreed or agreed to a survey statement. Conversely, a disagreement rate would represent the percentage of respondents that either strongly disagreed or disagreed to a survey statement.

The findings of this study suggest that perceptions of Texas secondary school administrators may solidly support several of the statements derived from perceptions of high-stakes testing critics. An agreement rate of 81%, a mean of 1.90, and a standard deviation of 1.11, suggests that secondary administrators' perceptions may be highly aligned with statement 31, no high stakes decision such as grade retention or graduation should be based on the results of a single test. Similarly, an agreement rate of 76%, a mean of 2.07, and a standard deviation of .98, implies that secondary administrators' perceptions seem closely associated with statement 29, high-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language. Likewise, an agreement rate of 76%, a mean of 2.10, and a standard deviation of 1.04, intimates that secondary administrators' perceptions seem closely coupled with statement 27, high-stakes testing and the accompanying consequences of failure lead to overstressed students. Equally, an agreement rate of 72%, a mean of 2.24, and a standard deviation of .89, advocates that secondary administrators' perceptions seem closely tied with statement 23, a focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction.

The findings of this study suggest that perceptions of Texas secondary school administrators may principally support several of the statements derived from perceptions of high-stakes testing critics. An agreement rate of 71%, a mean of 2.26, and a standard deviation of .98, suggests that secondary administrators' perceptions may be aligned with statement 25, high-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve. Similarly, an agreement rate of 71%, a mean of 2.28, and a standard deviation of 1.14, implies that secondary administrators' perceptions seem associated with statement 21, high-stakes testing compromises educational quality by leading educators to 'teach to the test,' which results in a narrowing of the curriculum. Likewise, an agreement rate of 70%, a mean of 2.18, and a standard deviation of 1.0, intimates that secondary administrators' perceptions may be coupled with statement 20, a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district. Equally, an agreement rate of 70%, a mean of 2.10, and a standard deviation of 1.04, advocates that secondary administrators' perceptions may be aligned with statement 28, the pressures inherent in preparing students for high-stakes tests are driving out good teachers.

Some of the findings of this study appear to be inconclusive when attempting to determine if perceptions of Texas secondary school administrators differ from critics' perceptions. Factors such as marginal agreement or disagreement rates, centric means, and larger standard deviations made arriving at a conclusion problematic for statements 24, 30, and 26.

Independent t-tests were conducted to test for possible perception differences between groups identified in the study. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years vs. 15 or More Years), Campus Classification (Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable).

On statement 20, a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district, a significant difference in means ($p = .024$) was found between large and small school responders. With a mean of 1.95 and a standard deviation of .94, small school responders found a higher level of agreement with the statement than large school responders.

On statement 30, the high-stakes testing movement is resulting in a significant increase in student dropout rates, a significant difference in means ($p = .046$) was found between responders from exemplary or recognized schools and responders from academically acceptable schools. With a mean of 2.88 and a standard deviation of .91, responders from exemplary or recognized schools found a higher level of disagreement with the statement than responders from academically acceptable schools.

CHAPTER V

SUMMARY AND CONCLUSIONS

This chapter presents a summary of the research findings and offers recommendations for future study, research, and implications for the current testing environment. The purpose of this study was to determine the impact of the high-stakes standardized test movement in Texas secondary schools. The method used to accomplish this goal was to compare and contrast what researchers believe the impact of the high stakes standardized test movement is on secondary campuses with what Texas secondary administrators' perceive the impact of the high stakes standardized test movement is on their own campuses. Secondary principals were asked in an electronic survey to rate the extent to which 31 statements matched their experiences on their campuses in regard to the impact of high-stakes tests. Statements 1 through 12 were derived from assertions made by supporters of the high-stakes testing movement. Statements 13 through 19 were derived from researcher's findings of unintended consequences of the current high-stakes standardized test movement. Statements 20 through 31 were derived from assertions made by critics of the high-stakes standardized test movement.

To examine all three research questions, the researcher used means, standard deviations, and frequencies to make assumptions about perceptions of the population of secondary administrators differing from statements derived from supporters, the unintended effects of high-stakes testing as reported in current literature, and critics.

Data from of all three sources were used to make an eyeball determination on how individual statements match the perceptions of Texas secondary school administrators.

In addition, Independent t-tests were conducted for all three research questions to detect if a possible perception difference exists between groups identified in the study. Independent groups examined in this study included: gender (Male and Female), Years of Administrative Experience (1 – 4 years vs. 15 or More Years), Campus Classification (Large vs. Small), and Current Campus Rating (Exemplary and Recognized vs. Academically Acceptable).

The three research questions of this study were:

1. Do perceptions of high-stakes testing supporters differ from those of Texas secondary school administrators in the membership of TASSP?
2. Do the unintended effects of high-stakes testing reported in current literature differ from those of Texas secondary school administrators in the membership of TASSP?
3. Do perceptions of high-stakes testing critics differ from those of Texas secondary school administrators in the membership of TASSP?

Summary

Research Question 1

Do perceptions of high-stakes testing supporters differ from those of Texas secondary school administrators in the membership of TASSP?

The findings of this study suggest that perceptions of Texas secondary school administrators match up to some degree with several statements derived from supporters' perceptions but seem to differ with one statement, too. As discussed previously in chapter IV, the findings of this study suggest that statements 10 (the use of data) and 7 (motivation resulting from the public display of test scores) appear to be closely aligned with supporters' perceptions. A more moderate alignment is perceived to exist with statements 1 (focusing attention on low-achieving schools) 12 (increased attention to special education students), and 11 (improvement of professional development).

Conversely, findings of this study appear to firmly differ with statement 5, doing poorly on tests increases student effort. In addition, a combination of factors including marginal agreement or disagreement rates, centric means, and larger standard deviations made arriving at a conclusion problematic for statements 2, 3, 4, 6, 8, and 9.

Independent t-tests suggest that several perception differences exist between groups in this study. Small school responders may have a significantly higher level of disagreement with statement 5, doing poorly on tests increases student effort, than large school responders. Also, statement 7, public display test scores motivates administrators, appears to have perception differences between two different groups.

This study suggests that small school responders found a significantly higher level of agreement with statement 7 than large school responders. Furthermore, responders with 1 – 4 years of experience found significantly a higher level of agreement with statement 7 than administrators with 15 years or more of experience did.

Research Question 2

Do perceptions of researchers of high-stakes testing reported in current literature differ from those of Texas secondary school administrators in the membership of TASSP?

The findings of this study suggest that perceptions of Texas secondary school administrators match up to some degree with several statements derived from perceptions of researchers of high-stakes testing reported in current literature, but seem to differ with one statement as well. As discussed previously in chapter IV, the findings of this study suggests a close alignment with statements 13 (increased knowledge of testing), 18 (excessive time spent on test preparation), 19 (loss of local control of instructional decisions), and 14 (intense data collection and use).

In opposition, this study suggests that administrators seem to adamantly differ with statement 16, current tests have positively evolved in most respects. In addition, a combination of factors including marginal agreement or disagreement rates, centric means, and larger standard deviations made arriving at a conclusion difficult for statements 15 (the promotion of educational homogeneity) and 17 (the enhancement of assessment practices).

Independent t-tests suggest that possible perception differences exist between groups in this study. This study advises that small school responders found a higher level of agreement with statement 18, excessive time spent on test preparation, than large school responders. Also, statement 19, loss of local control of instructional decisions, appears to have perception differences between two different groups.

This study alludes that small school responders found a higher level of agreement with statement 19 than large school responders. In addition, male responders found a higher level of agreement with this same statement than female responders did.

Research Question 3

Do perceptions of high-stakes testing critics differ from those of Texas secondary school administrators in the membership of TASSP?

The findings of this study suggest that perceptions of Texas secondary school administrators correspond to some degree with several statements derived from critics perceptions of high-stakes testing. The findings of this study suggest that perceptions of Texas secondary school administrators may solidly support statements 31 (high-stakes decisions should not be based on a single test) 29 (the unfair effect on students with limited English proficiency), 27 (resulting overstressed students), and 23 (lack of direction to improve instruction). Similarly, the findings of this study suggest that perceptions of Texas secondary school administrators may principally support statements 25 (jeopardizing improvement efforts), 21 (resulting narrowing of the curriculum), 20 (the validity of the multiple uses of high-stakes tests), and 28 (the pressures driving teacher out of education). Additionally, a combination of factors including marginal agreement or disagreement rates, centric means, and larger standard deviations made arriving at a conclusion problematic for statements 24, 26, and 30.

Independent t-tests suggest that possible perception differences exist between groups in this study for two statements. This study suggests that small school responders found a significantly higher level of agreement with statement 20, the validity of the

multiple uses of high-stakes tests, than large school responders did. In addition, responders from exemplary or recognized schools seem to exhibit a significantly higher level of disagreement with statement 30, significant increase in student dropout rates, than responders from academically acceptable schools did.

Implications and Discussions

The survey results of the impact of high-stakes testing movement resulted in numerous conclusions for Texas secondary teachers and administrators. Primarily, high-stakes testing was implemented by its pundits to point out deficiencies in schools and guide improvement for their students (Klein & Hamilton, 2001; Mazzeo, 2000; Perkinson, 1995). A extensive partnership of constituencies have championed standards-based high-stakes tests as a means of improving public schools' accountability, establishing a world class workforce, and decreasing the achievement gap among various racial, socioeconomic, and ethnic groups (Wong & Nicotera, 2007).

Perhaps the most profound issue of the accountability movement is the impact the use of data generated by the high-stakes movement is having on most facets of educational institutions; thus educators' decisions have become increasingly data driven (Wong & Nicotera, 2007). According to Schlechty (1997), principals are now expected to manage by results rather than manage by programs or rules. In his book *The Fifth Discipline*, Senge (1994) concluded that ultimately, the learning organization must be judged by results. In accordance with the supporters' position, this study confirms that the collection and use of copious data produced by high-stakes testing has positively impacted secondary campuses in that educators are currently making use of data to make

important decisions, refine programs, channel funding, identify roots of success, and allocate resources (Mertler, 2007, p. 28). Moreover, the use of data is providing students with personalized information about their individual knowledge and skills and can lead to differentiated instruction for individual students. In addition, this study confirms the position that data produced by high-stakes tests has helped to improve professional development by focusing on helping educators hone teaching skills and content expertise.

Supporters of the testing movement believe that high-stakes testing combined with holding teachers, administrators, and students accountable for successful performance on these tests will improve our public education system (Mazzeo, 2000; Nichols & Berliner, 2007). The implication is that coercion is necessary to motivate administrators and teachers to teach to the state standards. This study confirms the supporters' position that, as intended, public display of high stakes scores motivates administrators to use their position to ensure that the recommended standards are a part of the curriculum. The possibility of public humiliation, sanctions, and career endangerment are tangible possible consequences for administrators and teachers as a consequence of a lack of student success on high-stakes tests. In addition, this study affirmed that high-stakes tests are focusing teachers' instruction. However, this study was less than decisive on the how necessary the accompanying accountability of high-stakes testing is in motivating teachers and administrators to do their respective jobs.

According to Wong and Nicotera (2007), the standards-based movement's central new expectation is that all children should receive the high level of education once reserved for a fraction of our nation's students. This paradigm shift has radically

changed expectations for the poor and previously excluded and is having a tremendous impact on educators, lawmakers, and students (Wong & Nicotera, 2007). This study validates the supporter's belief that high-stakes tests have helped identify and focus attention on low-achieving students, particularly in designated minority subpopulations or those attending low performing schools. Accordingly, this study supports the position that increased attention to special education students and the corresponding increase of access to on-grade-level instruction is a direct result of high-stakes testing. The third implication derived from the supporter's position is that inclusion, not exclusion, is the new mantra of education. All students, including poor, minority, and special education students, must be successfully provided with a post secondary preparatory based education.

Testing and accountability are intended to improve achievement and motivate staff and students (Heubert & Hauser, 1999; Linn, 2000; Nichols & Berliner, 2007). This study brings into question the supporters' tenet that high-stakes testing is motivating students to apply more effort into their work and to study harder. Further, this study finds strong opposition to the position that failure on a test will increase student effort to learn. This study suggests that while the students recognize that high-stakes tests are obstacles they will need to surmount, this knowledge is not motivating them on a daily basis. Further, failure on high-stakes test often leads to negative reactions on the part of unsuccessful testers and may prompt some students to give up. Research is clear that students who repeat a grade are significantly more likely to drop out of school (Goldschmidt & Wang, 1999). Research finds that students who have been held back typically do not catch up, even with remedial help and low performing

students learn more if they are promoted than if they are held back (Heubert & Hauser, 1999). In addition, this study casts doubt on how successfully the gap of achievement between minority and majority students has been closed. The clear implication is that policy makers should re-examine policies and procedures intended to help students and ensure that students are not harmed.

This study confirmed that the increased impact by policy makers at the state and national levels has resulted in a loss of local control of educational decisions. Efforts to establish national standards and tests grew out of several key developments, such as *A Nation at Risk: The Imperative for Educational Reform* (NCEE, 1983), the adaptation of President Bush's and the nation's governors' six national education goals (Swanson, 1989), the establishment of the National Council on Education Standards and Testing (NCEST, 1992), and the No Child Left Behind Act. Although the Clinton Administration tried to encourage the development of "voluntary" national tests in 1997, most experts agree that the national-standards debate is now over (Doyle & Pimentel, 1999). The No Child Left Behind Act has mandated that states create tests and set standards. Whereas states once provided only curriculum frameworks and outlines, they are now dictating the content of instruction (Jones, Jones & Hargrove, 2003). Therefore, this study suggests that high-stakes testing has created de facto statewide curricula in Texas, totally usurping the traditional local control over such matters. Further, the increasing federal incursion will eventually lead to national curricula for Texas and the other states of the union.

Critics purport that despite the focus on low performing students, doubt has been generated on how effective high-stakes testing have been in closing the gap between

minority and majority students. In fact, this study supports the critics' position and goes on to suggest that educational decisions based on high-stakes tests have a disproportionate impact on poor and minority students. This is viewed as particularly true of students for whom English is not their first language. In many circles, standardized tests have long been considered unfair and biased against students from ethnic minority and or impoverished backgrounds because these tests are based in large measure on the experiences of middle class European Americans (Hilliard, 2000; Neill & Medina, 1989). Thus, policy makers and psychometricians must evaluate current assessment to ensure that these tests are free from bias, relevant, and age appropriate and demonstrate a high degree of reliability and validity, a position currently not substantiated by this study.

Further, this study indicates those high-stakes tests are having a negative impact on the students they were designed to help. Strong sentiment exists that fear of failure on high-stakes tests is resulting in overstressed students. Research by Hancock (2001) showed that all students, regardless of their tendencies toward test anxiety, achieve more poorly under conditions of high evaluative threat. This study also suggests that high-stakes testing is perceived as increasing student dropout rates and resulting in a corresponding drop in graduation and completion rates. An obvious implication of this study is that the impact of high-stakes tests on student dropout rates, graduation rates, and levels of student anxiety must be addressed. It is imperative that a series of strategies must be employed to identify potential dropout students and provide effective prevention programs to diminish this problem. In addition, coping mechanisms and stress management must be incorporated into test preparation.

Policymakers expect testing programs to certifying a student's level of achievement, provide information about an education system's effectiveness, motivate student performance, bringing coherence to a curriculum, and hold schools and educators accountable for student performance (Hamilton, Stecher & Klein, 2002). This study finds fault with the current usage of high-stakes tests. The study questions the appropriateness of using a test that has validated only for diagnosing strengths and weaknesses of individual students to evaluate the educational quality of a school or school district. According to Heubert and Hauser (1999), the content standard's check demands criterion-referenced testing; the school or student ranking goal demands norm-referenced testing. Therefore, one test cannot adequately do both. Policy makers should evaluate the current usage of high-stakes tests and design separate means to evaluate students and schools.

Conclusions

This study suggests that the impact of high-stakes testing as perceived by Texas secondary school administrators differs little from many of the perceptions of supporters, critics, and researchers alike. In contrast, the findings of this study imply that only one of the perceptions of critics and the unintended effects of high-stakes testing reported in current literature contrasts with the perceptions of Texas secondary school administrators.

Perhaps the most positive impact of the accountability movement is the use of copious data generated by the high-stakes movement. This study suggests that, in accordance with the findings on statement 1, the interest in data collection and use in

public education is certain. Accordingly, secondary school administrators' decisions appear to have become increasingly data driven. The appropriate use of data has potential to focus instruction, guide staff development, and help improve students' academic achievement.

Perceptions of secondary school administrators found in this survey appear to strongly support statement 31, no high stakes decision such as grade retention or graduation should be based on the results of a single test. While alternative means of promotion exist in most high-stakes grade levels, this is not the case for the exit exams required for high school graduation. The researcher concludes that other appropriate methods to ascertain completion of high-stakes criteria should be developed.

The perceptions of secondary school administrators apparently support the primary tenets of the accountability system that high-stakes tests have helped focus public attention on schools with low-achieving students. This study seems to substantiate these findings as discussed previously for statement 1. Further, perceptions of secondary school administrators and review of the literature affirm that high-stakes tests may be a catalyst for increased attention to students with special needs.

In accordance with a review of the literature and seemingly corroborated by findings of this study, the current accountability system attempts to use coercion to motivate students, teachers, and administrators alike. This study, particularly in regard to statement 7, seems to substantiate the perception that possible negative consequences do influence administrators and to a lesser degree, teachers. Interestingly, t-test findings suggest that the perceptions of a small school secondary administrators' acknowledge a greater impact in the public display of high-stakes scores than their large school cohorts.

The researcher suggests that the increased visibility inherent in the role of small school administrator may enhance this effect, thus explaining this finding. Conversely, the relative anonymity of large school administrator may explain these results. In addition, the constraints of small staffs and the limited scope of central administration inherent in small school districts may place more of the responsibility of high-stakes results squarely on the shoulders of small school administrators. In contrast, the dynamic of larger staffs may allow the perception that responsibility is shared amongst other staff members and central office personnel.

Further, an additional t-test finding implies that a significant difference exists between the perceptions of secondary school administrators with 1 – 4 years of experience and responders with 15 years or more of experience in regard to the impact of the public display of high-stakes results. Unexpectedly, these findings seem to suggest that responders with 1 – 4 years of experience give a higher level of agreement with the statement than responders with 15 years or more of experience. Perhaps this can best be explained by the diverse curricula of college administrative training programs these two groups may have experienced in search of administrative certification. While accountability has long been a issue addressed in certification training, the emphasis of preparing secondary administrators to address the new realities of accountability certainly have increased in recent years. Unquestionably, the possibility exists that many of the secondary administrators with 15 years or more of experience may have obtained certification in educational administrative training programs that predate the current high-stakes accountability movement. Discussions in educational administration

classes taken over a decade ago often centered around the appropriateness of high-stakes testing and not in developing the skills needed to deal with this reality.

This study brings in to question the effectiveness of high-stakes tests in motivating students. Perceptions of secondary school administrators appear to differ strongly with this primary tenet of the accountability movement. Current literature and results from statement 5 suggest that failure on high-stakes tests leading to grade retention and or a lack of promotion may result in negative consequences for students, including but not limited to giving up and becoming a dropout. Interestingly, t-test findings suggest that a significant difference exists in the perceptions between small school secondary administrators and their large school peers. The t-test findings indicate that the perceptions of small school responders find a higher level of disagreement than the perceptions of large school responders in regard to the effectiveness of motivation resulting from poor student performance on high-stakes test.

The review of the literature suggests that the use of public monies inspired taxpayers, politicians, and journalists to call for educational evaluation to insure their funds were used effectively. The influx of federal monies into public schools is thought to have exacerbated this phenomenon culminating in the current high-stakes standard-based accountability movement. The perceptions of secondary school administrators in this study seem to indicate that there has been a shift in control from local communities to policy makers at the state and national levels. The perceptions of secondary school administrators is that local control of what is taught, how it is taught, and who gets high-quality instruction has been eroded. In this regard, t-test findings indicate that a significant difference exists between the perceptions of small and school responders.

Small school administrators found a higher level of agreement with the position that local control is eroding. This finding may result from the autonomy often experienced by small school administrators in contrast to the shared authority typically found in large schools with more administrators at every level. Further, another t-test suggests that a significant difference exists between male and female responders in regard to this same issue. Perhaps this may be due to the evolution of the paradigm, as secondary administrators move away from management by rules and policies toward management by results (Schlechty, 1997).

The researcher suggests that the implementation of state-mandated Texas Essential Knowledge and Skills (TEKS) has created a de facto state curriculum. Further, a review of the literature seemingly demonstrates that the No Child Left Behind Act has mandated high-stakes tests for state public schools, increasing the federal influence of public schools to unforeseen levels. While Tyack (1990) has established a long history of federal interest in public schools, the findings of the study suggest that the No Child Left Behind Act has dramatically expanded the federal role in education. NCLB has mandated a system of educational evaluation of all public schools. The researcher speculates that if this trend continues, national curricula may soon become a reality.

This study suggests that perceptions of secondary school administrators may support the position that high-stakes testing is having a disproportionate impact on poor students, minority students, and students for whom English is not their first language.

While the perceptions of secondary school administrators appear inconclusive when looking at this group as a whole, a significant difference in means was found between responders from exemplary or recognized schools and responders from

academically acceptable schools. Perceptions of secondary school administrators from exemplary or recognized schools found a higher level of disagreement with this statement than responders from academically acceptable schools. This finding seems intuitively logical, as dropouts would be a limited problem in most exemplary or recognized schools. Conversely, the researcher expects that academically acceptable schools would experience higher dropout rates that are often associated with academic struggles.

Perceptions of secondary school administrators appear to differ with the position that high-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions. Thus, the researcher concludes that policy makers and psychometricians must re-evaluate current assessment and improve test instruments to eliminate bias. Similarly, the findings of the study questions the appropriateness of using a test that was validated only for diagnosing strengths and weaknesses of individual students to evaluate the educational quality of a school or school district. As discussed in the review of literature, criterion-referenced testing and norm-referenced testing have distinct capabilities. Perceptions of secondary school administrators as found in this study suggest that one test cannot effectively cover the diverse purposes it is currently asked to cover in Texas. In this regard, t-test findings indicate that a significant difference exists between the perceptions of found between large and small school responders to statement 20, a test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district. Small school responders

were found to have a higher level of agreement with this statement than large school responders. As discussed earlier in the conclusion, the constraints of small staffs and the limited scope of central administration inherent in small school districts may place more of the responsibility of high-stakes results squarely on the shoulders of small school administrators. The researcher suggests that the ownership inherent in this arrangement may explain this finding.

Recommendations

Recommendations to Policymakers

Based upon this study's findings and conclusions, several recommendations to policy makers and administrators can be made. While high-stakes tests have demonstrated that they are useful monitors of student progress, the impact of high-stakes testing should continually be evaluated to insure that the unintended negative effects of high-stakes accountability do not outweigh the intended positive effects. Since high-stakes tests are used to determine which students will advance and influence which subjects schools will teach, the imperative is that educators understand how the use of high-stakes testing will impact student drop-out rates, graduation rates, course curriculum, levels of student anxiety, and teaching practices. Policy makers should establish a system for evaluating both the intended positive effects and the harmful, unintended negative effects of the system (AERA, 2000). Accordingly, policies and procedures must be adjusted or enacted to mitigate the negative side effects of high-stakes testing.

This study makes the case that Texas secondary administrators question the position that current high-stakes tests are highly reliable; free from bias; relevant and age appropriate; and yield remarkably consistent decisions. Thus, policymakers should require that new high-quality assessments be developed. These psychometric creations must take into account the impact on diverse targeted populations, particularly on racial and ethnic-minority students, students of lower socioeconomic status, or students who do not have English as a primary language. With regard to the changing demographics of Texas, particular consideration should be taken when a student lacks proficiency in English; the test could become a measure of their ability to communicate in English rather than a measure of their knowledge or skills in a subject matter.

The strongest recommendation generated from this study is that policy makers desist from making crucial educational decisions such as graduation based on the results of a single test. In fact, the most virulent support of any survey statement was reached on statement 31, no high stakes decision such as grade retention or graduation should be based on the results of a single test. This statement was derived from the American Educational Research Association (AERA) position statement on high-stakes testing in pre-K – 12 education and was adopted July 2000. Instead, multiple indicators of educational attainment should be developed. Doing so should increase the validity of inferences based upon observable gains in student achievement. When evidence exists that a test score may not adequately reflect a student's true proficiency, policy makers should provide alternative acceptable means by which to demonstrate attainment of the tested standard.

Based upon this study, policy makers should cease using a test validated for diagnosing strengths and weaknesses of individual students to evaluate the educational quality of a school or a district. This study found that respondents overwhelmingly believe that the current system of high-stakes tests draws an inaccurate picture of student achievement and unfairly jeopardizes students and schools that are making genuine efforts to improve. Instead, more emphasis should be placed on student performance from year to year, thus allowing for differences in students' current level while maintaining an expectation of improvement from that level. Thus, consideration of a value-added system is supported by this study. Value-added approach provides schools with students who begin behind grade level expectation a reasonable chance to show improvement while concurrently promoting high expectations for those schools and students.

Recommendations to Administrators

Based upon this study's findings, administrators should examine policies and procedures that impact curriculum and instruction. Care must be exercised in the allocation of resources and selection of course offerings to insure that an unintentional narrowing of curriculum or elimination of courses does not ensue. The administrative focus should be on increased learning, not just increased test scores. Thus, administrators should minimize wasting resources on simple test preparation and support integration of test standards into the everyday curriculum. In order to prevent the elimination Support of programs with curricula not included in high-stakes test is

essential to enable students to receive a well-rounded education. A holistic approach in allocation of all resources is essential on the part of administrators to avoid this pitfall.

An additional recommendation for administrators and teachers generated by this study is to ensure that the district curriculum alignment occurs between what is taught and what is tested. This can be best achieved when administrators support and expect a rich curriculum, imbedded with skills and knowledge needed on high-stakes tests, but replete with rigor and relevance for students. A systemic approach to curriculum alignment, that clearly delineates when skills and knowledge will be taught, can eliminate redundancy and insure that all test standards are addressed. In addition, this should reduce circumstances that might lead to inappropriate test practices, including outright cheating.

Based upon this study's findings, administrators should enhance the use of the profuse data produced by high-stakes testing to guide decisions on resources, programs, and personnel. This data should be used to guide program implementation, development and evaluation. Additionally, the data should help guide staff development selection and implementation, too. Administrators should assist the other stakeholders, particularly teachers and students, to understand and effectively utilize the copious data produced by high-stakes testing. This data could be used to provide teachers with personalized information about the individual knowledge and skills level of each of his or her students and drive both curriculum and lesson plan development. Further, diagnosis of individual student needs should result in differentiated instruction decisions guided by longitudinal data. While the aforementioned summative data is important, formative data is crucial to identify content not mastered by students, evaluate the effectiveness of lessons, and

redirect ongoing instruction. Thus, active use of classroom data must be a cornerstone of an efficient and effective teaching regimen.

An additional finding from this study is that the advent of high-stakes testing has led to improved staff development. Administrators must insure that district staff development is personalized to meet the specific problems and needs facing their staffs. Subject specific staff development is extremely useful at the secondary level. Providing opportunities for staff members to provide input into staff development often increase the appropriateness of the training and a corresponding buy-in to the offerings. One critical piece of staff development that is typically ignored is follow-up. The curriculum coach model holds serious promise by helping teachers put into practice the skills acquired during staff development, and provides feedback as to the effectiveness of the implementation.

The final recommendation for administrators is that systems must be in place to help students and teachers cope with the stress of high-stakes testing. While a certain amount of anxiety is natural and expected, some students experience extreme reactions to the stress generated by the consequences of high-stakes tests. Thus, coping technique must be taught as part of a comprehensive test preparation program. Intuitively, the best means to reduce test anxiety is to have students well prepared for the test. Knowing the material reduces stress by building confidence. Secondly, the use of practice tests can help gauge readiness while increasing the familiarity with the organization of the test, the type of assessment questions, the time constraints, and the development of testing strategies. Test strategies include encouraging students to skip questions they find difficult, returning to them after building momentum and reducing anxiety after

answering easier questions. Third, test administrator's demeanor and execution of duties can be beneficial toward anxiety control. Reading instructions is the first step of the standardized test process. A relaxed, calm, and controlled reading of directions can set the tone for the test. A hurried or rushed approach to test administration may increase anxiety. In addition, students who do not fully understand any part of the process must feel confident enough to ask questions. This will only take place if the student is confident that he or she will not be embarrassed for asking questions. The test administrator's response to the student question is of paramount importance, maintaining the dignity of the student is necessary to build a climate of trust. Finally, physical methods to reduce tension should be taught and their implementation encouraged during tests. Taking slow deep breaths, stretching exercises, neck rolls, and frequent breaks may help reduce tension and promote stress relief. In addition, sucking on mints or chewing gum may help some students handle anxiety as well.

Teachers who feel sole responsibility for test results face high anxiety, a decrease in job satisfaction, and this may result in driving him or her out of education. While administrators and teachers must confront the brutal facts, each is responsible for only a part of high-stakes test results. Collaboration between administrators and teachers toward planning viable solutions to shortcomings, working the plan together, and using formative evaluation along the way could empower teachers and help alleviate excessive stress. Departmental collaboration and division of duties can help teachers work smarter, not just harder. In addition, administrators can enhance the prowess of their staffs by providing staff development, supportive services, and encouragement to their staffs. To promote teacher retention, administrators should create a supportive climate

and culture by developing a system of recognizing excellence in teaching. The power of sincere praise cannot be overemphasized. Praise could include honors recognition, meals out, compensatory periods off, and notices on evaluations.

Recommendations for Further Study

Several directions for further research are evident from the present study. As the target population was secondary Texas administrators, generalization of the results to elementary schools is in question. Replication of this study with a target population of Texas elementary administrators would be valuable for verification of results and provide opportunities for comparison. In addition, since the target population was restricted to Texas secondary administrators, generalization of the results to other states is in question. Replication of this study in other states would be valuable for verification of results and provide opportunities for comparison. As statements in the study were used to evaluate administrators' opinions on the impact of high-stakes testing on teachers, it may be more effective and efficient to replicate part of this study to directly ascertain teachers' positions on these areas.

These replication studies should extend the present study in several ways. First, more demographic data should be obtained from respondents. For example, the grade level taught or supervised should be delineated to determine if the intensity of the impact of high-stakes tests differs by these parameters. Teachers involved with grades taking high-stakes test may have a differing opinion of the impact of high-stakes tests than teachers with non-tested grades.

One procedural change, which may enhance future replication studies by reducing perceived bias to the survey, is to randomly arrange the statements in the survey. In the present study, statements are grouped into three cohorts, those generated by supporters (1 – 21), those recognized by researchers of unintended consequences (13 – 19), and those generated by critics. Despite an attempt by the researcher to maintain a balance of positions, one respondent believed that the survey was biased toward the critics' position. This may be explained by the arrangements of the statements, with the critics' statements coming at the end of the survey. In addition, random arrangement of statements would reduce the possible effect of test fatigue.

Finally, the findings of the study may be enhanced if particular school districts were selected. A stratified random sample of districts, using factors such as district and campus ratings, could be selected to ascertain the systemic impact of high-stakes tests. This would necessitate an agreement with the superintendents of each of these districts to allow access to each of his or her administrators. Assurances of protection of identity would be a paramount concern. This benefit would allow the researcher to determine the state of curriculum alignment, amount of campus control over curriculum, and strategies used for test preparation. Administrators at every level could be interviewed to discover if the recommendations in this study are feasible and practical.

REFERENCES

- Achieve, Inc. (2002). *Aiming higher: Meeting the challenges of educational reform in Texas*. Retrieved December 22, 2003 from <http://ritter.tea.state.tx.us/curriculum/aimhitexas.pdf>.
- Allington, R. L., & McGill-Franzen, A. (1992). Unintended effects of educational reform in New York. *Educational Policy*, 6(4), 397-398.
- American Educational Research Association (AERA). (1999). *Appropriate use of high-stakes testing in our nation's schools*. Washington, DC: Author.
- American Educational Research Association (AERA). (2000). *AERA position statement concerning high-stakes testing in pre-K-12 education*. Washington, DC: Author.
- Amrein, A. L., & Berliner, D. C. (2002). *High-stakes testing & student learning*. Retrieved December 22, 2003 from <http://epaa.asu.edu/epaa/v10n18/>.
- Amrein, A. L., & Berliner, D. C. (2003). The testing divided: New research on the intended and unintended impact of high-stakes testing. *Peer Review*, 5(2), 31-32.
- Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools*. Reading, MA: Addison-Wesley.
- Bolton, D. G. (2000). Critical issues in school governance, finance, and management. *School Business Affairs*, 66(9), 4-6.
- Bracey, G. W. (1995). Debunking the myths about money for schools. *Educational Leadership*, 53(3), 65-67.
- Bracey, G. W. (2002). April foolishness: The 20th anniversary of "A Nation at Risk." *Phi Delta Kappan*, 84(8), 616-621.

- Bushaw, W. J., & Gallup, A. M. (2008). *The 40th annual Phi Delta Kappa/Gallup Poll of the public's attitudes toward the public schools*. Retrieved September 1, 2008 from http://www.pdkmembers.org/members_online/publications/e-GALLUP/kpoll_pdfs/pdkpoll40_2008.pdf.
- Callahan, R. E. (1962). *Education and the cult of efficiency*. Chicago: University of Chicago Press.
- Chafin, C. K. (2004). *Using student performance data humanely: The danger of losing perspective on teaching and learning and the value of test scores*. Retrieved December 29, 2005 from <http://www.aasa.org/publications/saarticledetail.cfm?ItemNumber=1067>.
- Cizek, G. J. (2001a). More unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 20(4), 19-27.
- Cizek, G. J. (2001b). Cheating to the test. *Education Matters*, 1(1), 40-47.
- Cizek, G. J. (2002). Unintended consequences of high-stakes testing. *Educational Measurement: Issues and Practice*, 22(1), 36-39.
- Clarke, M. M., Haney, W., & Madaus, G. (2000). *High stakes testing and high school completion*. Retrieved March 13, 2002 from <http://www.bc.edu/research/nbetpp/publications/v1n3.html>.
- deFur, S. H. (2002). Education reform, high-stakes assessment, and students with disabilities: One state's approach. *Remedial and Special Education*, 23(4), 203-205.
- Doyle, D. P., & Pimentel, S. (1999). *Raising the standard: An eight-step action guide for schools and communities* (2nd ed.). Thousand Oaks, CA: Corwin Press.

- Driesler, S. D. (2001). Whiplash about backlash: The truth about public support for testing. *National Council on Measurement in Education Newsletter*, 9(3), 2-5.
- Ebel, R. L., & Frisbie, D. A. (1986). *Essentials of educational measurement* (4th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, NJ: Prentice-Hall.
- FairTest. (2000). *MCAS: Making the Massachusetts dropout crisis worse*. Retrieved September 29, 2002 from <http://www.fairtest.org/mcas-alert-september-2000>.
- Falk, B. (2002). Standards-based reforms: Problems and possibilities. *Phi Delta Kappan*, 83, 612-620.
- Fassold, M. A. J. (2000). Disparate impact analyses of TAAS scores and school quality. *Hispanic Journal of Behavioral Sciences*, 22(4), 460.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2004). *Program evaluation: Alternative approaches and practical guidelines* (3rd ed.). Boston: Pearson/Allyn & Bacon.
- Gall, M., Borg, W., & Gall, J. (1996). *Educational research: An introduction* (6th ed.). New York: Longman Publisher USA.
- Goldschmidt, P., & Wang, J. (1999). When can schools affect dropout behavior? A longitudinal multilevel analysis. *American Educational Research Journal*, 36(4), 715-716.
- Gould, S. J. (1981). *The mismeasure of man* (1st ed.). New York: W. W. Norton & Company.

- Grissmer, D., & Flanagan, A. (1998). *Exploring rapid achievement gains in North Carolina and Texas*. Retrieved February 13, 2003 from <http://govinfo.library.unt.edu/negp/reports/grissmer.pdf>.
- Grissmer, D., Flanagan, A., Kawata, J., & Williamson, S. (2000). *Improving student achievement: What state NAEP test scores tell us*. Santa Monica, CA: Rand Corporation.
- Hamilton, L. S., Stecher, B. M., & Klein, S. P. (2002). *Making sense of test-based accountability in education*. Santa Monica, CA: Rand Corporation.
- Hancock, D. R. (2001). Effects of test anxiety and evaluative threat on students' achievement and motivation. *Journal of Educational Research*, 94(5), 284-290.
- Haney, W. (2000). *The myth of the Texas miracle in education*. Retrieved January 26, 2001 from <http://epaa.asu.edu/epaa/v8n41>.
- Haney, W., Fowler, C., Wheelock, A., Bebell, D., & Malec, N. (1999). *Less truth than error? An independent study of the Massachusetts Teacher Tests*. Retrieved February 2, 2001 from <http://epaa.asu.edu/epaa/v7n4/>.
- Heubert, J. P., & Hauser, R. M.. (1999). *High-stakes: Testing for tracking, promotion, and graduation*. Washington, DC: National Academy Press.
- Hilliard, A. G., III. (2000). Excellence in education versus high-stakes standardized testing. *Journal of Teacher Education*, 51(4), 293-296.
- Hoffman, C. M. (2001). *Federal support for education: Fiscal years 1980 to 2001*. Retrieved September 29, 2002 from <http://nces.ed.gov/pubs2002/2002129.pdf>.

- Hopkins, K. D., Stanley, J. C., & Hopkins, B. R. (1990). *Educational and psychological measurement and evaluation*. Englewood Cliffs, NJ: Prentice Hall.
- House Education and Workforce Committee. (2002). *"No Child Left Behind" emphasizes results, expands options for children with special needs*. Washington, DC: Author.
- Hoy, W. K., & Miskel, C. G. (1996). *Educational administration: Theory, research, and practice* (5th ed.). Boston, MA: McGraw-Hill.
- Jacob, B. A. (2001). Getting tough? The impact of high school graduation exams. *Educational Evaluation and Policy Analysis*, 23(2), 99-121.
- Jones, M. G., Jones, B. D., & Hargrove, T. Y. (2003). *The unintended consequences of high-stakes testing*. Lanham, MD: Rowman & Littlefield.
- Klein, S. P., & Hamilton, L. (2001). *Large-scale testing: Current practices and new directions*. Santa Monica, CA: Rand Corporation.
- Klein, S. P., Hamilton, L. S., McCaffrey, D. F., & Stecher, B. M. (2000). *What do test scores in Texas tell us?* Retrieved March 13, 2002 from http://www.rand.org/pubs/issue_papers/IP202/index2.html.
- Kohn, A. (2001). Fighting the tests: A practical guide to rescuing our schools. *Phi Delta Kappan*, 82(5), 349-357.
- Kohn, A. (2004). *Rescuing our schools from tougher standards*. Retrieved January 17, 2005 from <http://www.alfiekohn.org/standards/standards.htm>.
- Lanier, K. (2000, April). *Hearken to the sound of the trumpet*. Paper presented to the American Educational Research Association, New Orleans, LA.

- Linn, R. L. (2000). Assessments and accountability. *Educational Researcher*, 29(2), 4-16.
- Lutz, F. W., & Merz, C. (1992). *The politics of school/community relations*. New York: Teachers College Press.
- Madaus, G. F., & Clarke, M. (2001). *The adverse impact of high stakes testing on minority students: Evidence from 100 years of test data*. New York: Century Foundation Press.
- Madaus, G. F., Stufflebeam, D. L., & Scriven, M. S. (1983). *Evaluation models: Viewpoints on educational and human services evaluation*. Boston: Kluwer-Nijhoff Publishing.
- Mazzeo, C. J. (2000). *Examining examinations: Assessment policy in historical perspective, 1865-1935*. Unpublished dissertation, Stanford University, Stanford, CA
- McNeil, L. M. (2000a). Sameness, bureaucracy, and the myth of educational equity: The TAAS system of testing in Texas public schools. *Hispanic Journal of Behavioral Sciences*, 22(4), 508-510.
- McNeil, L. M. (2000b). *Contradictions of school reform: Educational costs of standardized testing*. London: Routledge.
- McNeil, L., Coppola, E., Radigan, J., & Heilig, J. V. (2008). *Avoidable losses: High-stakes accountability and the dropout crisis*. Retrieved August 2, 2008 from <http://epaa.asu.edu/epaa/v16n3/>.

- McNeil, L., & Valenzuela, A. (2000). *The harmful impact of the TAAS system of testing in Texas: Beneath the accountability rhetoric*. Retrieved March 13, 2002 from <http://www.edb.utexas.edu/latino/McNeil%20&%20Valenzuela.pdf>.
- Mertler, C. A. (2007). *Interpreting standardized test scores: Strategies for data-driven instructional decision making*. Los Angeles: Sage Publications.
- National Commission on Excellence in Education (NCEE). (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: United States Department of Education.
- National Council on Education Standards and Testing (NCEST). (1992). *Raising standards for American education. A report to congress, the secretary of education, the national education goals panel, and the American people*. Washington, DC: U.S. Government Printing Office.
- National Information Center for Children and Youth with Disabilities. (1998). *The IDEA amendments of 1997*. Retrieved November 10, 2002 from <http://old.nichcy.org/pubs/newsdig/nd26pdf.pdf>.
- National Library of Education. (1998) *Educational and labor market performance of GED recipients*. Retrieved January 5, 2004 from <http://www.ed.gov/pubs/GED/backgrd.html>.
- Neill, D. M., & Medina, N. J. (1989). Standardized testing: Harmful to educational health. *Phi Delta Kappan*, 70(9), 688-697.
- Nevi, C. (2002). *The abuse of accountability*. Retrieved January 17, 2002 from http://findarticles.com/p/articles/mi_m0JSD/is_6_59/ai_86876002.

- Nichols, S. L., & Berliner, D. C. (2007). *Collateral damage: The effects of high-stakes testing on America's schools*. Cambridge, MA: Harvard Education Press Group.
- Parker, L. C. (1994). *A historical study of the use of program evaluation in education*. Tulsa, OK: The University of Tulsa.
- Passman, R. (2000, January). *Pressure cooker: Experiences with student-centered teaching and learning in high-stakes assessment environments*. Paper presented at the Annual Meeting of the Southwest Educational Research Association, Dallas, TX.
- Perkinson, H. J. (1995). *Imperfect panacea: American faith in education* (4th ed.). New York: McGraw Hill.
- Popham, W. J. (1974). *Evaluation in education: Current applications*. Berkeley, CA: McCutchan Publishing Corporation.
- Popham, W. J. (1975). *Education evaluation*. Englewood Cliffs, NJ: Prentice-Hall.
- Popham, W. J. (1992). *Educational evaluation* (3rd ed.). Needham Heights, MA: Allyn & Bacon.
- Roderick, M., & Engel, M. (2001). The grasshopper and the ant: Motivational responses of low-achieving students to high-stakes testing. *Educational Evaluation and Policy Analysis*, 23(3), 197-227.
- Rogers, T. B. (1995). *The psychological testing enterprise: An introduction*. Belmont, CA: Thomson Brooks/Cole Publishing Co.
- Sacks, P. (1999). *Standardized minds: The high price of America's testing culture and what we can do to change it*. Cambridge, MA: Perseus Publishing.

- Sacks, P. (2000). Predictable losers in testing schemes. *School Administrator*, 57(11), 6, 8-9.
- Sadker, M. P., & Sadker, D. M. (2000). *Teachers, schools, and society* (5th ed.). Boston: McGraw Hill.
- Slechty, P. (1997). *Inventing better schools: An action plan for educational reform*. San Francisco: Jossey-Bass.
- Schrag, P. (2000, August). High stakes are for tomatoes. *The Atlantic Monthly*, 286(2), 19-21.
- Secretary's Commission on Achieving Necessary Skills (SCANS). (1990). *Teaching the SCANS competencies*. Washington, DC: Superintendent of Documents.
- Senge, P. (1994). *The fifth discipline: The art and practice of the learning organization*. New York: Doubleday.
- Sheppard, L. A. (2002). *The hazards of high-stakes testing: Hyped by many as the key to improving the quality of education*. Retrieved December 22, 2003 from <http://www.issues.org/19.2/shepard.htm>.
- Shuford, T. (2004). *Unintended consequences of no child left behind*. Retrieved December 29, 2005 from <http://ednews.org/articles/26/1/Unintended-Consequences-of-No-Child-Left-Behind/Page1.html>.
- Skrla, L., Scheurich, J. J., Johnson, J. F., Jr., & Koschoreck, J. W. (2001). Accountability for equity: Can state policy leverage social justice? *International Journal of Leadership in Education*, 4(3), 237-260.
- Stecher, B. M., & Hamilton, L. S. (2002). Putting theory to the test. *Rand Review*, 26(1), 16-17.

- Swanson, B. (1989). *An overview of the six national education goals*. Retrieved February 13, 2001 from http://findarticles.com/p/articles/mi_pric/is_199105/ai_3171609006.
- Texas Education Agency (TEA). (1997). *Timeline of testing in Texas*. Retrieved February 13, 2001 from <http://www.state.tx.us/student.assessmentet/resources/studies/testingtimeline.pdf>.
- Texas Education Agency (TEA). (2002). *TAKS information booklets*. Austin, TX: Author.
- Texas Education Agency (TEA). (2003). *The handbook of Texas online*. Retrieved January 3, 2004 from <http://www.tsha.utexas.edu/handbook/online/articles/view/TT/met2.html>.
- Townsend, B. L. (2002). Testing while black: Standards-based school reform and African American learners. *Remedial and Special Education, 23*(4), 222-230.
- Tuckman, B. W. (2003). The effect of learning and motivation strategies training on college students' achievement. *Journal of College Student Development, 44*(3), 430-437.
- Tyack, D. (1990). "Restructuring" in historical perspective: Tinkering toward Utopia. *Teachers College Record, 92*(2), 170-191.
- Tyack, D., & Timar, T. (1999). *The invisible hand of ideology: Perspectives from the history of school governance*. Denver, CO: ECS Distribution Center.
- U.S. Congress Senate. (1993). *Goal 2000: Educate America Act*. Hearings 103rd Congress, 1st session, Committee on Labor and Human Resources. Washington, DC: Government Printing Office.

- Vornberg, J. A. (1991). *Texas public school organization and administration* (2nd ed.). Dubuque, IA: Kendall Hunt Publishing Co.
- Walberg, H. J. (2003). Accountability unplugged. *Education Next*, 3(2), 76-79.
- Walberg, H. J., & Haertel, G. D. (Eds.). (1990). *The international encyclopedia of educational evaluation* (1st ed.). New York: Pergamon Press.
- WestEd. (2003). *Making sure exit exams get a passing grade*. Retrieved January 3, 2004 from http://eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/1b/66/4b.pdf.
- Wolf, R. M. (1990). *Evaluation in education: Foundations of competency assessment and program review* (3rd ed.). New York: Praeger.
- Wong, K. K., & Nicotera, A. W. (2007). *Successful schools and educational accountability: Concepts and skills to meet leadership challenges*. Boston, MA: Allyn & Bacon.
- Worthen, B. R., & Sanders, J. R. (1973). *Educational evaluation: Theory and practice*. Belmont, CA: Wadsworth Publishing.
- Worthen, B. R., & Sanders, J. R. (1987). *Education evaluation: Alternative approaches and practical guidelines*. New York: Longman.
- Wright, W. E. (2006). A catch-22 for language learners. *Educational Leadership*, 64(3), 22-27.

APPENDIX A

SURVEY INSTRUMENT

1. Introduction

My name is Dave Denny. Like you, I'm a Texas Principal and I need your help. A few of you may know me as the Principal of Thorndale High School or as the Region 13 Representative for TASSP.

I am completing my requirements for a doctorate in Educational Administration through Texas A&M University. The final project towards this milestone is the completion of a Record of Study. My Record of Study is entitled, "The Impact of State-Mandated Standard-based High Stakes Testing on Selected Texas Public Secondary Schools as Perceived by Select Administrators in the Membership of the Texas Association of Secondary School Principals." To complete this study I am distributing this survey to group of 400 randomly assigned high school principals throughout Texas.

You have been asked to participate in a research study about the impact of state-mandated, standard based, high-stakes tests on Texas secondary campuses. This survey examines issues raised in current literature about the impact of high stakes testing. As a framework for the survey questions, I am measuring the extent to which Texas high school principals' opinions match the opinions of authors of current research. In addition, the study will evaluate if there are any significant differences between the responses of principals segregated into demographic groups.

You were randomly selected to be a possible participant because of your membership in the Texas Association of Secondary School Principals. The purpose of this study is to provide research data directly from practicing educational leaders in order to compare the claims of proponents, critics, and examine possible unintended results of Texas' testing movement.

If you agree to be in the study, you will be asked to answer seven demographic questions and thirty-one survey question. The study will take about ten minutes. As this study is confidential, there is little likelihood that you will face any risk by participating in the study. There are no benefits or compensation for participation.

This study is confidential. The records of the study will be kept private. No identifiers linking you to the study will be included in any sort of report that might be published. Research records will be stored securely and only I (Dave Denny) will have access to the records. Your decision whether or not to participate will not affect your current or future relations with Texas A&M University. You can withdraw at anytime without affect. You can contact Dave Denny, Thorndale High School Principal, at (512) 898-2321 or my advisor Dr. John Hoyle at (979) 845-2748 with any questions about this study.

This research study has been reviewed by the Institutional Review Board – Human Subjects in Research, Texas A&M University. For research-related problems or questions regarding subjects' rights, you can contact the Institutional Review Board through Ms. Melissa McIlhaney, IRB Program Coordinator, Office of Research Compliance, (979) 458-4067, mcilhaney@tanu.edu.

Please be sure you have read the above information, asked questions, and received answers to your satisfaction.

2. Demographic Information

The information in this section will be used to make comparisons by demographic groups.

D1. Gender

- Male Female

D2. Campus Classification

- Small School (A, AA, or AAA) Large School (AAAA or AAAAA)

D3. Title I Campus

- Yes No

D4. Years of Administrative Experience

- 1 - 4 years
 5 - 14 years
 15 or more years

D5. Total Years as an Educator

- 1-7
 8-17
 18-22
 23 or more years

D6. Current Administrative Position

- Principal
 Associate/Assistant/Vice Principal
 Other

D7. Current Campus Rating

- Exemplary
- Recognized
- Academically Acceptable
- Academically Unacceptable
- Not Rated

3. Survey

Please answer the following question in regard to your campus.

1. High-stakes tests have helped focus public attention on schools with low-achieving students and, as a result, have made these students more visible and less likely to slip between the cracks and fall further behind.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

2. High-stakes tests are designed and implemented to improve instruction by helping teachers focus on what is most important to teach.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

3. High-stakes tests have helped close the gap in achievement between minority students and majority students in Texas.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

4. *Teachers need to be held accountable through high-stakes tests to motivate them to teach better, particularly to push the least motivated ones to perform.*

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

5. *Doing poorly on high-stakes tests will lead to increased student effort to learn.*

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

6. *Students work harder and learn more because they know what is expected and that the high-stakes tests really count.*

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

7. *The public display of high-stakes test scores motivates administrators to ensure that standards (TEKS) on which the tests are based are part of the curriculum and are being successfully taught.*

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

8. When high-stakes tests are developed and used appropriately, they are among the most sound and objective knowledge and performance measures available.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

9. Administrators need to be held accountable through high-stakes tests to motivate them to be more effective in supervising their staffs.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

10. Increasingly, from the classroom to the school board room, educators are making use of student performance data generated by high-stakes tests to help them refine programs, channel funding, and identify roots of success.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

11. Driven by the demands of high-stakes tests, professional development has improved by focusing on helping educators hone his or her teaching skills and content area expertise.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

12. The implementation of high-stakes testing has been a catalyst for increased attention to students with special needs.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

13. One result of high-stakes testing is that educators know more about testing than ever before.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

14. Prominent and public interest in pupil performance on high-stakes tests has resulted in an intensity of effort directed toward data collection and quality control that is unparalleled.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

15. High-stakes tests promote greater homogeneity of education. A result of schools' aligning their curricula and instructional focus more closely to outcomes embodied in high-stakes tests, the experiences of and aspirations for children in urban, suburban, and rural districts within a state are more comparable than they have been in the recent past.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

16. A profoundly positive effect that the introduction of high-stakes consequences has had lies in the tests themselves. High-stakes tests have evolved to a state of being: highly reliable; free from bias; relevant and age appropriate; higher order; tightly related to important public goals; time and cost efficient; and yielding remarkably consistent decisions.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

17. High-stakes tests have exposed educators to high-quality writing prompts, document-based questions, constructed-response formats, and even challenging multiple-choice items. This has led to teachers enhancing their own assessment practices.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

18. High-stakes testing programs also result in massive amounts of test preparation, resulting in a loss of instructional time.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

19. High-stakes testing has resulted in a loss of local control of what is taught, how it is taught, and who gets high-quality instruction. These decisions are now greatly impacted by policy makers at the state and national levels.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

20. A test that has been validated only for diagnosing strengths and weaknesses of individual students should not be used to evaluate the educational quality of a school or school district.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

21. High-stakes testing compromises educational quality by leading educators to "teach to the test," which results in a narrowing of the curriculum, limiting the scope of tested subjects and shortchanging or eliminating subjects not included in the assessments.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

22. High-stakes tests are too expensive and result in diverting scarce resources and attention from serious problems.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

23. A focus on standards and accountability that ignores the processes of teaching and learning in classrooms will not provide the direction that teachers need in their quest to improve instruction.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

24. Pressure exerted from the need to succeed on high-stakes tests often leads to inappropriate test preparation practices, including outright cheating.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

25. High-stakes tests draw an inaccurate picture of student achievement and unfairly jeopardize students or schools that are making genuine efforts to improve.

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

26. *Educational decisions based on high-stakes tests have a disproportionate impact on poor and minority children.*

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

27. *High-stakes testing and the accompanying consequences of failure lead to overstressed students.*

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

28. *The pressures inherent in preparing students for high-stakes tests are driving out good teachers.*

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

29. *High-stakes tests unfairly and inaccurately assess and penalize learners for whom English is not their first language.*

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

30. *The High stakes testing movement is resulting in a significant increase in student drop out rates.*

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

31. *No high-stakes decision such as grade retention or graduation should be based on the results of a single test.*

- Strongly Agree
- Agree
- Unsure
- Disagree
- Strongly Disagree

32. *Do you have any suggestion or comments about this survey?*

APPENDIX B

OPEN ENDED QUESTION REPLIES

Do you have any suggestion or comments about this survey?

1. Great survey!
2. Good survey, hope to see the results.
3. One thing you need to consider is that in high stakes testing there is never a comparison of how a certain group of students improve from year to year, not different groups.
4. So, what are you thinking?
5. Some of the questions are misleading such as #13 - hard work does not necessarily equate to more learning.
6. Shorten the number of questions per screen to about 5-7
7. Children can handle high stakes tests. Teachers can handle teaching beyond high stakes tests. Administrators can handle making sure both happen.
8. Thank you for asking for an educator's input. The decisions regarding this testing are made by people who are not in the trenches each and every day.
9. As I read these questions, it appears to me that you are trying to get answers that are against the testing in Texas without giving any credence to the great future this will mean to our workforce.
10. Teachers and staff are now to the point that they are teaching around the parameters of the high stakes test. Limits the ability of the teacher to develop each different individual learner and ready.
11. We are in the people business. It's not all about a test score.
12. Very interesting.

13. Does not address the NCLB goal of 100% passage.
14. Great questions! Good luck.
15. Could High-stakes testing be a prominent factor in the shortage of school administrators across the state?
16. Can I get a copy of results?
17. I appreciate the work you're doing. Please share the results of your survey with me.
18. Some of the questions are lengthy.
19. Question # 34 - leads to overstressed teachers.
20. Interesting. Sometimes it would be nice to have a chance to explain our responses.
Sometimes I answered 'agree' or 'disagree' but maybe not for the reasons you would think.
21. Good Luck, I'll be interested in your find.
22. Ignores the testing business and financial ties to it. The bureaucratic costs of testing.
The level of impact on genuine teaching. Levels of dis-satisfaction in tea are eluded to in one of the questions
23. We qualify as a Title 1 School, but no secondary schools in our district access the funds. At the secondary level, two tracks are needed- one for college bound students and the other for vocational.
24. Fire the whole lot in the Legislature, followed closely by the Commissioner of Ed. and the Governor.
25. Would like to have a copy of your results.

26. On many questions, I wanted to put the word most in front of the word students. As I answered the questions, I mentally did that anyway. Some of the questions I felt the need to justify my answers.
27. I think this is a great study. Good luck and if there is anything else I can do please let me know.
28. Sorry I'm late, hope you can use it.
29. Some questions, especially in the first section, are really two questions, resulting in ambiguous responses.

VITA

Name: Davis M. Denny III

Address: 4490 E. University Ave.
Georgetown, Texas 78626

Email Address: dmd3txag@farm-market.net

Education Doctor of Education – 2008
Major: Educational Administration
Texas A&M University
College Station, Texas

Master of Education – 1992
Major: Educational Administration
East Texas State University
Commerce, Texas

Bachelor of Science – 1980
Major: Physical Educational
Texas A&M University
College Station, Texas