

**TEACHER BELIEFS ABOUT THE  
OUTCOMES OF HIGH-STAKES TESTING AND  
MEASUREMENT-DRIVEN INSTRUCTION  
IN VIRGINIA'S PUBLIC SCHOOLS**

by

Dale E. Margheim

Dissertation Submitted to the Faculty of the  
Virginia Polytechnic Institute and State University in  
Partial Fulfillment of the Requirements for the Degree of

Doctor of Education

in

Educational Leadership and Policy Studies

D. J. Parks, Chairman  
D. N. Gillespie  
J. L. Sellers  
J. A. Sughrue  
W. M. Worner

November 8, 2001

Blacksburg, Virginia

Key Words: Teacher Beliefs, Testing, Virginia, Systemic Change, Student Achievement, Teacher Autonomy, Teacher Satisfaction, State Testing Programs, Standardized Testing

# TEACHER BELIEFS ABOUT THE OUTCOMES OF HIGH-STAKES TESTING AND MEASUREMENT-DRIVEN INSTRUCTION IN VIRGINIA'S PUBLIC SCHOOLS

Dale E. Margheim

(ABSTRACT)

The relationships between teacher opinions about Virginia's Standards of Learning testing program and five variables related to teachers' backgrounds and present working conditions were examined in the context of five dependent variables represented as discrete domains. A systematic sample of 464 members of the Virginia Education Association (VEA) was selected to participate in the study. A 52 item survey was mailed to the members of the sample to gather their opinions about the outcomes they believed were occurring as a result of Virginia's Standards of Learning testing program and the state's Regulations Establishing Standards for Accrediting Public Schools in Virginia. Three hundred fifty-two usable questionnaires were returned, for a 76% return rate. Descriptive statistics were used to portray teacher responses in five domains. Three way ANOVAs were computed to determine if any significant main effects or interactions were evident among the independent variables of tenure status, SOL test grade status, and school socio-economic level. Telephone surveys of twelve randomly selected teachers were conducted to enhance understanding of three dependent variables (domains): student outcomes, instructional outcomes, and teacher outcomes. These data were analyzed using qualitative strategies.

## Findings

Teachers' responses to the mailed survey indicated relatively unfavorable opinions about the outcomes of Virginia's Standards of Learning testing program and related regulations. Outcomes for instructional programs and for teachers themselves were rated more unfavorably than outcomes for students, outcomes for schools, or outcomes for public confidence. The ANOVA analyses indicated that teacher opinions did not vary meaningfully by any of the variables studied or by any combination of these variables. In short, even though a systematic sample of VEA members from throughout the state was surveyed, the opinions of this group of teachers were remarkably similar. Interview data confirmed that teachers had many concerns about outcomes associated with SOL testing. The interviews also indicated that teachers attributed several positive outcomes to SOL testing as well. Several rival hypotheses are presented to explain the apparent homogeneity of opinions among this systematic sample of Virginia educators.

## DEDICATION

This study is dedicated to the three great women in my life:

To my mother, Dorothy Margheim, who taught me that all people have the same inherent value, regardless of their station or circumstances in life and who helped kindle my desire to work for the success of all students.

To the late Linda Margheim, the best teacher I ever knew. Her patient guidance is largely responsible for any successes I have achieved as an educator.

To my loving wife, Lynn, who has been unbelievably interested, supportive, and encouraging throughout this study. You are simply the best.

## ACKNOWLEDGEMENTS

Several people deserve special thanks for their contributions to my research:

First, I thank Dr. David Parks who has been the most patient and knowledgeable teacher and advisor imaginable. Under his tutelage I learned a great deal about serious research and a great deal about myself as well.

To my dissertation committee: I appreciate your careful reading and recommendations about the study. I know your attention to this task took time away from other pressing responsibilities.

To the staff of Shawsville Elementary School: Thank you for being so supportive as I chased this dream.

Appreciation is also due to my fellow administrators in the Montgomery County school system. What would I have done without your assistance with all those content validity studies?

And finally, I thank my superintendent, Mr. Fred Morton, and my assistant superintendent, Dr. Kitty Rodgers, who allowed me a great deal of flexibility and autonomy to attend classes and meetings related to this study.

## TABLE OF CONTENTS

	Page
ABSTRACT .....	ii
DEDICATION.....	iii
ACKNOWLEDGEMENTS.....	iv
LIST OF TABLES.....	vii
LIST OF FIGURES.....	viii
LIST OF EXHIBITS.....	ix
CHAPTER 1 THE PROBLEM.....	1
The Quest for Accountability .....	1
Purpose of the Study .....	12
Research Question.....	12
Theoretical Model .....	12
Definitions .....	13
Personal Bias.....	15
Limitations .....	15
Summary of Chapter One and Overview of the Dissertation .....	16
CHAPTER 2 REVIEW OF RELATED LITERATURE .....	18
Outcomes for Students .....	18
Outcomes for Instructional Practices.....	21
Outcomes for Schools.....	24
Outcomes for Public Confidence in Teachers and Schools.....	25
Outcomes for Teachers.....	26
Individual Research Studies .....	28
Summary .....	60
CHAPTER 3 METHODOLOGY .....	63
Population .....	63
Selection of Sample.....	63
Instrumentation and Data Collection Procedures.....	64
Analysis of Survey Data.....	70
Construction of Teacher Interview Protocol.....	71
Selection of Participants for Interviews.....	74
Administration of Interviews.....	75
Analysis of Interview Data.....	75

CHAPTER 4 RESULTS .....	77
Data Analysis for Interviews.....	77
Summary.....	101
Data Analysis for Questionnaires .....	102
Summary .....	104
CHAPTER 5 SUMMARY, CONCLUSIONS, DISCUSSION, POLICY CONSIDERATIONS, AND RECOMMENDATIONS FOR FURTHER RESEARCH .....	106
Summary of Survey Data.....	106
Summary of Interview Data .....	107
Conclusions.....	107
Discussion of Survey Data and Interview Data.....	108
Policy Considerations .....	112
Implications for Further Research .....	114
REFERENCE LIST .....	116
APPENDICES	
Appendix A Original Survey Domains, Domain Descriptions, .....	124
and Related Items	
Appendix B Content Validation Instrument for Survey.....	127
Appendix C Content Validation Data for the Survey .....	134
Appendix D Survey Instrument .....	149
Appendix E Preliminary Interview Schedule (February, 2001).....	151
Appendix F Content Validation Instrument for Interviews.....	152
Appendix G Content Validation Data for the Interviews .....	154
Appendix H Raw Data Matrices for Interviews.....	158
Appendix I Domain Scores and Item Scores by Years of Teaching Experience, Primary Teaching Assignment, Tenure Status, SOL Test Grade Status, and Socio-Economic Level of Teachers' Schools .....	250
Appendix J Analysis of Variance Tables for the Five Domains.....	280
VITA.....	285

LIST OF TABLES

Table		Page
1	1998/99 Statewide Standards of Learning Pass Rates.....	8
2	TAAS Pass Rates by Race and Class in Texas.....	19
3	Analysis of Variance Comparison of Organizational Adjustment Scores for Pennsylvania and Maryland.....	30
4	Analysis of Variance Comparison of Culture Adjustment Scores by State.....	31
5	Strength of Accountability Systems.....	47
6	Level of Organizational Capacity .....	49
7	Student Participation in the Illinois Goal Assessment Program, 1987-1993.....	53
8	Alpha Coefficients for Scales on the Survey of Teacher Beliefs About the Outcomes of High-Stakes Testing and Measurement-Driven Instruction in Virginia’s Public Schools .....	70
9	Populations and Samples for Conducting Telephone Interviews by SOL Test Grade, Tenure Status, and School Socio-Economic Status (SES).....	78
10	Characteristics of Participants in Telephone Interviews .....	78
11	Decisions Linked to Null Hypotheses.....	105

## LIST OF FIGURES

Figure		Page
1	Theoretical model: Variables associated with teacher beliefs about the outcomes of measurement-driven instruction and high-stakes testing in the context of national and state standards movements.....	14



LIST OF EXHIBITS

Exhibit		Page
1	The Final Five Domains and Survey Items.....	66
2	The Final Domains and Descriptions for Interview Questions, the Interview Schedule, and the Interview Protocol.....	72

## CHAPTER 1

### THE PROBLEM

#### The Quest for Accountability

Academic standards are about accountability. They were put in place because our schools were graduating children who couldn't read or write well enough to fill out job applications. Some students were passed from grade to grade based on age, not knowledge. Parents and families began to demand more, and some began to lose faith in public education. Now parents will know whether or not their children will be equipped to work in the New Economy we are creating. Our teachers, administrators, superintendents, and schools can now be measured and held accountable by parents. (Governor Jim Gilmore, *The State of the Commonwealth Address*, January 12, 2000)

Persistent concerns about the quality of public education have led Virginia to the brink of momentous change in the focus and direction of its educational programs. Goals in Regulations Establishing Standards for Accrediting Public Schools in Virginia (Virginia Department of Education, 2000) indicate that all students should achieve to high standards, that state assessments should determine school and student progress, that student promotion and graduation should be based on these assessments, and that schools should be evaluated and accredited based on student performance on state assessments. This movement is occurring in the context of a vocal national debate about the merits of government-mandated curriculum and assessment. Yet, even in its infancy, the movement for higher educational standards could claim that some 16 states required a passing score on a state-mandated test for high school graduation (Neil, 1997).

Why do political leaders seem to place such increasing importance on mandated curricula and testing reforms? Dorn (1998) postulated that part of the reason is that "A number connotes objectivity, or at the very least legitimacy" (p. 3). Mandated testing programs can become a symbol to the public that "something is being done" to improve the problem and that the "something" is reasonable based on the public's experience (Madaus, 1988, p. 35). The implementation of high-stakes testing and measurement-driven instruction may convince much of the public that improvement is at hand, even though symbolic policy-making may be preventing the consideration of other potentially effective intervention strategies (Airasian, 1988, Airasian, 1993).

#### The National Standards Movement

The national thrust to ensure high standards in both instruction and student achievement is frequently traced to the publication of A Nation at Risk: The Imperative for Educational Reform (National Commission on Excellence in Education, 1983) in which the performance of American students was portrayed as mediocre and shameful. When compared to student achievement in other nations, American students were never first or second, and were last in 7 of 19 assessments; 23 million American adults and about 13 percent of all 17 year-olds were reported to be

functionally illiterate, and this percentage among minority youth may have reached 40 percent; the average achievement of high school students on standardized achievement tests was lower than when Sputnik was launched in 1957; and SAT scores steadily declined over the 20 years preceding the report.

Ravitch (1995) listed nine reasons for the increased popularity of initiatives to ensure higher and more precise educational standards for students and schools: (1) measured declines in student achievement in the late 1970s, (2) public discontent over poor performance by American students on international assessments in the 1980s, (3) increased public opinion in the 1980s that schools should be judged not only by their “inputs,” but also by their “outputs” or levels of student performance, (4) persistent gaps in the educational achievement of students from different ethnic and racial groups, (5) concern that low expectations were being held for the “bottom” half of the age cohort, (6) a perception that the American workplace needed more skilled workers, (7) worry that poorly educated youth might impair the nation’s competitive place in the world economy, (8) a lack of agreement among educators, parents, and the business community about what students should learn, and (9) a concern that in trying to develop programs and curricula in response to many special interest and pressure groups, schools had become “catch-all institutions, lacking in priorities or focus” (Ravitch, 1995, p. 4). The reform effort that grew from these perceived shortcomings in American education frequently came to be known as the *standards movement*. Leaders of the standards movement called for a variety of innovations and philosophies, but central to the movement was the expressed belief that all children can learn to high levels.

Spurred by the ideas in *A Nation at Risk*, the standards movement gained national prominence in the form of America 2000 (U.S. Department of Education, 1991), a Bush administration initiative begun at the National Governors’ Conference in Charlottesville, Virginia, in 1989. America 2000 was not a federal education program as such, but it offered a framework of goals through which states could develop revised educational goals.

Ravitch (1993) stated that America 2000 had three fundamental objectives:

(1) to encourage every community to adopt the national goals, develop its own local strategy, and prepare an annual community report card on its progress toward these goals; (2) to stimulate the creation of thousands of “break-the-mold” schools that would approach education in totally new ways to meet the needs of today’s children and families; and, (3) to develop “world class” standards and American Achievement Tests. (p. 768)

Though never enacted into law, America 2000 continued in spirit with enactment in 1994 of the Clinton administration’s Goals 2000: Educate America Act. This act contained four elements designed to transform American public schooling (Arons, 1997): First, state governments were required to adopt content standards for public school curricula. Second, students were expected to demonstrate competency over the determined content material as a prerequisite for promotion and graduation. Third, supporters of Goals 2000 sought to create a change in public education that is systemic. That is, changes would affect all aspects of schooling,

including textbook development, teacher certification, curriculum content, and student assessment. Fourth, Goals 2000 moved the base of power from the local level to the state and national level. “In essence, Goals 2000 envisions a standardized and comprehensive scheme of schooling imposed from the top down by experts and politicians” (Arons, 1997, p.65).

Goals 2000 contained eight national education goals, six of which are directly relevant to this study:

(1) By the year 2000, all children in America will start school ready to learn. (2) By the year 2000, the high school graduation rate will increase to at least 90 percent. (3) By the year 2000, all students will leave grades 4, 8, and 12 having demonstrated competency over challenging subject matter including English, mathematics, science, foreign languages, civics and government, economics, arts, history, and geography.... (4) By the year 2000, the nation’s teaching force will have access to programs for continued improvement of their professional skills....(5) By the year 2000, U.S. students will be first in the world in mathematics and science achievement. (6) By the year 2000 every adult American will be literate and will possess the knowledge and skills necessary to compete in a global economy and exercise the rights and responsibilities of citizenship. (Arons, 1997, pp. 69-70)

Goals 2000 standards were labeled as voluntary; however, to access the millions of dollars authorized and anticipated to support the legislation, states had to comply with program components. In order to receive funding, each state was required to develop its own content standards; each state had to develop and submit an annual “state improvement plan” that includes student performance standards to the United States Department of Education; and, each state had to develop a process for ensuring that its program for teacher and administrator licensure and continuing education are aligned with its content and performance standards. In sum, the authors of Goals 2000 prescribed a blueprint for the comprehensive reform of public schools across the nation. It seems clear that the intent of the act was to invoke a government-mandated curriculum with specified performance expectations as well as staff development requirements.

Corbett and Wilson (1991) named the mindset of this movement “reform by comparison” (p. 2), and they described the latest strategy in the reform process as “school-by-school, district-by-district, even state-by-state achievement comparisons” (p. 3). The theory was that comparisons would lead to public pressure, which would lead in turn to compliance in curriculum revision and in instructional practices. The process has come to be called “high-stakes testing.” Madaus (1988) described the phenomenon as follows:

High-stakes tests include those used for the certification or recertification of teachers, promotion of students from one grade to the next, award of a high school diploma, assignment of a student to a remedial class, allocation of funds to a school or school district, award of merit pay to teachers on the basis of their students’ test performance, certification or recertification of a school or district, and placement of a school system into “educational receivership.” (p. 30)

Corbett and Wilson (1991) identified three characteristics of the standards movement: (1) increased accountability is evident on the part of educators for student performance, (2) uniform indicators of school outcomes are established and publicized to allow the measured achievement of schools and districts to be compared, and (3) the motivations behind the movement are both educationally and politically based.

This movement received a boost and direction with Popham's 1987 article, "The Merits of Measurement-Driven Instruction." Popham's major arguments were as follows: (1) The quality of public education could be improved if mediocre instructional materials were replaced with "empirically proven alternatives." (2) Education could also be improved if our teaching staff were well-paid and highly skilled. (3) Because these alternatives are costly, however, another strategy is preferred. This less expensive method is termed *measurement-driven instruction*.

Popham (1987) explained that measurement-driven instruction occurs "when a high-stakes test of educational achievement, because of the important contingencies associated with the students' performance, influences the instructional program that prepares students for the test" (p. 680). Popham further described two categories of high-stakes tests. The first consists of examinations with important consequences for students such as the right to be promoted to the next grade or to receive a high school diploma. The second category is testing used to evaluate the quality of teachers, schools, and instructional programs. Popham explained that high-stakes tests serve as powerful "curricular magnets," ensuring that teachers will focus on what policy-makers have agreed should be learned.

Madaus (1991) offered a listing of advantages often attributed to high-stakes tests: (1) Such tests are a relatively objective means of ensuring that all students receive the same instruction and opportunities. (2) High-stakes tests give teachers direction and purpose in what they do. (3) They give students incentives. (4) The tests reduce the prospect of having achievement results tainted by student-teacher relationship issues. (5) And, finally, high-stakes tests are widely accepted by the public. Similarly, in a study of the effects of the Kentucky Education Reform Act (KERA), Clifford (1995) reported that the KERA had caused Kentucky educators to make substantive changes in their instructional practices and that those changes closely mirrored the state-mandated tests. However, the author stated that teachers worried about the reliability and validity of the KERA and reported that they did not have confidence that the assessment accurately portrayed schools' effectiveness.

### Virginia and the Standards Movement

Virginia's first entrance into the standards movement came in the form of the state's Literacy Passport Test (LPT) first given in 1989. The test, assessing reading, writing, and math competencies beginning at the sixth grade level, was designed to assure minimum competency. By implementing the LPT, the state was attempting to guarantee that its graduates could read, write, and compute at least at the sixth grade level. Students had to pass all three sections of the LPT to receive a regular high school diploma, but they could retake any failed section of the test year after year until they were able to pass it. School divisions commonly expended considerable

energy in tutorial and compensatory efforts to ensure that as many students as possible passed the LPT and were, therefore, eligible for regular high school diplomas. However, an analysis of 1995 LPT sixth grade pass rates showed that only 60 percent of students were able to pass all three sections of this assessment on their first try. In that year, the Virginia Department of Education indicated that "...the failure rate has not improved since the LPT was first given in 1989" (*Virginia's K-12 Education Reform: Raising Student Achievement in Virginia*, www.pen.k12.va.us).

Following the seeming dissatisfaction with the LPT as being "minimal" in nature, and given the climate encouraging standards-based reforms in the nation (Supovitz & Brennan, 1997), Virginia began a strong and steady movement toward the development of both challenging content and performance standards for its public schools. The Virginia Board of Education approved its new K-12 curriculum, Standards of Learning (SOL), in June of 1995. The SOL in English, math, science, and social studies were distributed to school divisions in July of 1995 with the expectation that these standards would become the basis of curricular offerings in every school division in the Commonwealth. Beginning in August of 1996 public hearings were held to discuss extensive revisions to the Regulations Establishing Standards for Accrediting Public Schools in Virginia (SOA). New SOA were adopted by the State Board of Education in October of 1997, and a more recent version was approved on July 28, 2000. While the SOL contained content standards for students, the SOA provided the enforcement or performance standards for schools.

The following are among the consequences for schools and students as provided in the July 2000 version of the SOA (Regulations Establishing Standards for Accrediting Public Schools in Virginia, 8VAC20-131-10 et.seq., July 28, 2000):

(1) Schools are directed to consider SOL test performance in grades 3, 5, and 8 when making promotion-retention decisions. (2) Diplomas are to be differentiated so that students must pass six state competency tests at the high school level to earn *verified units of credit* and a *standard diploma*. Students with documented disabilities may earn a *modified standard diploma* by passing all course work and state-mandated "literacy and numeracy competency assessments." Students who cannot pass competency tests can earn only *certificates of program completion*. (3) School accreditation will continue to be based on student performance on SOL tests. For schools to be accredited, students will have to reach increasingly higher SOL benchmark scores each year until finally by 2003-04 a 70% pass rate will be required in the core academic areas except in grades 3 and 5 English where a 75% pass rate will be required. (4) Beginning in 2000-01, schools rated as *Accredited with Warning* due to low scores in English and math must undergo an "academic review" and must adopt instructional methods with "...a proven track record of success" as specified by the Virginia Department of Education. (5) The SOA also require the preparation and dissemination of an annual Performance Report Card to be sent to all parents for each public school. The card provides information about the school's test results, accreditation status, attendance and other factors. The Report Cards are distributed and made available through the Internet and are widely reported in the local and state media. Additionally, each parent is sent an individualized report on how their child achieved *on each item on the test* as compared to passing

percentages for the item in the school and in the division. And, schools themselves are required to disseminate the same information annually through a school-generated report.

By any definition, the stakes for the accountability system in Virginia are high for students, parents, communities, and educators. Parents face the potential that their children may not be able to graduate from high school or may need unknown amounts of remediation to pass the state's verified units of credit tests. Communities face the potential of having schools that are not accredited with the likely perception that their quality of education is low. Non-accredited schools can mean a loss of community prestige and lowered property values. Educators face public humiliation and criticism if their school scores on SOL tests do not meet the required standard or if they are significantly lower than surrounding or seemingly comparable schools. And, teachers and administrators may be transferred, receive lower compensation, or lose their jobs if their schools do not score high enough on the state-mandated tests.

### Results of First and Second Year SOL Tests

Table 1 contains the results of 1998 and 1999 statewide SOL tests. Three conclusions may be drawn from these data: (1) Students generally do not pass the tests in very high numbers, (2) African-American students do considerably worse than the overall population tested. At any level, it is common to see a pass rate that is 20 percent or more lower for African-American students than for the overall population. Therefore, one might reasonably expect that disproportionately fewer African-American students will receive regular high school diplomas than will the general population. (3) If these scores are illustrative of the levels of student achievement to be expected by the graduating class of 2004, many students will be unable to complete requirements to allow them to receive a regular high school diploma. (Overall pass rates for high school courses in Virginia for 1999 were as follows: English - 75.69%, Math - 49.5%, History - 44.82%, and Science - 69.34%). One must keep in mind that students must pass 22 standard units of credit (Carnegie units) and all end-of-course state tests to graduate. Passing scores in five of the six required tests and one test failure will preclude the awarding of a regular high school diploma (Regulations Establishing Standards for Accrediting Public Schools in Virginia, 2000, pp. 5-6).

### Opposition to the Standards Movement

Though the notion of "high standards for all students" has much appeal, consequences associated with the standards movement have produced more than a few critics.

National standards and assessments have been promoted as a means for upgrading curriculum and student performance in schools. They are unlikely to have that effect for several reasons: First top-down specifications of content linked to tests cannot take into account the many pathways to learning that will be appropriate for different students in schools across the country. Second, school communities must undertake their own hard work on standard-setting and consensus development if they are to become committed to and knowledgeable about change. And, finally, large inequalities in opportunities to learn are more responsible for learning gaps than a paucity of tests. Standards and tests have already proved

themselves to be an ineffectual means for leveraging resource equalization.  
(Darling-Hammond, 1994b, p. 478)



Table 1

1998/99 Statewide Standards of Learning Pass Rates

Standards of learning tests	1998 passing rate %		1999 passing rate %	
	Overall pass %	African-American pass %	Overall pass %	African-American pass %
<b>Grade 3</b>				
English	53	33	61	42
Mathematics	63	40	68	45
History & Social Science	49	25	62	39
Science	63	37	68	43
<b>Grade 5</b>				
English: Reading, Literature & Research	68	47	69	48
English: Writing	65	45	81	67
Mathematics	47	24	51	27
History & Social Science	33	13	46	24
Science	59	33	67	41
<b>Grade 8</b>				
English: Reading, Literature & Research	64	45	67	45
English: Writing	67	48	70	52
Mathematics	53	27	60	36
History & Social Science	35	14	40	19
Science	71	47	78	56
<b>High School</b>				
English: Reading, Literature & Research	72	55	75	59
English: Writing	71	54	81	68
Algebra I	40	20	56	36
Algebra II	31	13	51	29
Geometry	52	25	62	34
Earth Science	58	31	65	40
Biology	72	50	81	64
Chemistry	54	31	64	41
World History to 1000 A.D. & World Geography	62	38	68	46
World History from 1000 A.D. to the Present & World Geography	41	17	47	21
U.S. History	30	12	32	13

Adapted from [www.pen.k12.va.us/VDOE/NewHome/pressreleases/aug599.html](http://www.pen.k12.va.us/VDOE/NewHome/pressreleases/aug599.html) August 5, 1999, Virginia Department of Education press release and July 1999 Virginia Standards of Learning Assessments, Statewide Passing Rates, Spring 1998 and Spring 1999: [www.pen.k12.va.us](http://www.pen.k12.va.us).

Berliner (1995) argued that the standards movement has, in fact, been manufactured by special interest groups:

The Manufactured Crisis was not an accidental event. Rather, it appeared within a specific historical context and was led by identifiable critics whose political goals could be furthered by scapegoating educators. It was also supported from its inception by an assortment of questionable techniques including misleading methods for analyzing data, distorting reports of findings, and suppressing contradictory evidence. Moreover, it was tied to misguided schemes for “reforming” education schemes that would, if adopted, seriously damage American schools. ( p. 4)

While the National Commission on Excellence in Education (1983) stated that American students do not fare well on international tests of achievement, Rotberg (1998) refuted this notion. She noted that sampling and population differences account for most or all of the performance variance among countries. Rotberg concluded that low levels of participation and high levels of exclusion in the testing practices of other countries led to the perception of inferior performance by U.S. students. Bracey (1991) opined other positive indicators of the quality of American education, including dramatic increases in U.S. high school graduation rates, a steady increase in the percentage of the population earning bachelors degrees, and the unparalleled representation of Americans in professional journal publications.

The notion of viewing all students as possessing the same potential to achieve at high levels on a standardized test has drawn substantial criticism. McDonnell, McLaughlin, and Morison (1997) concluded that the prospect of conveying a challenging curriculum effectively to the entire student body is unreasonable, stating, “The expectations of those advocating standards-based reforms currently exceed the limits of existing professional practice and expert knowledge” (p. 40). Biddle (1997) reported that levels of school funding and child poverty in the U.S. are strongly associated with differences in math achievement among both districts and states. In a study specific to Virginia, Cross (1999) was able to predict two-thirds of the variance in SOL scores across Virginia school divisions on the basis of only three indicators of socio-economic status. He questioned why the state would want to test students and then sanction students and schools when most of the variance in student achievement among schools can be predicted by socio-economic factors over which schools have little control. Similarly, in an Arizona study, Loughrin (1998) determined that student achievement on both norm-referenced and performance-based tests was significantly affected by students’ socio-economic level, minority status, and mobility.

McGill-Franzen and Allington (1993) found that in an effort to raise state-mandated test scores students suffered many injustices, including being retained before the benchmark “test grade” for state-mandated tests, being assigned to transitional classes to postpone their being tested for a year or more, and being classified as handicapped before high-stakes testing began.

Madaus and Kellaghan (1993) described what they term “the tradition of past exams” in which teacher knowledge of mandated exams determines future curriculum. Smith and Rottenberg (1991) described a host of negative effects from high-stakes testing, including a reduction in the time available for ordinary instruction (as opposed to preparing for tests), a decline in teacher morale, and more homogeneous grouping and tracking. Rodgers, Paredes, and Mangingo (1991) found that after four years of emphasis on minimum competency testing in Texas, “both basic and high level skills as measured by the Tests of Achievement and Proficiency went down” (p.7).

Young (1996) described vast differences in perceptions among teachers, principals, superintendents, and state legislators about the effectiveness and fairness of the Tennessee Value-Added Assessment System (TVAAS). Principals, teachers, and superintendents held similar beliefs that the TVAAS was not fair to teachers, did not improve instruction, and did not have a positive effect on teacher morale. These issues were all seen more positively by state legislators and state department of education officials, indicating a strong variance in perception between educators and policymakers.

The following position statement was shared at a meeting to accept public comments on proposed revisions to Regulations Establishing Standards for Accrediting Public Schools in Virginia. The meeting, sponsored by the Virginia Department of Education, was held in Marion, Virginia on November 30, 1999. The document was signed by representatives of (1) the Virginia Association of Elementary School Principals, (2) the Virginia Association of School Superintendents, (3) the Virginia Association of Secondary School Principals, (4) the Virginia Congress of Parents and Teachers, (5) the Virginia Education Association, and (6) the Virginia School Counselors Association:

Evaluations of students for purposes of promotion, retention, and graduation should be based on multiple criteria, including, but not limited to, classroom performance, teacher-developed assessments, and assessments of the Standards of Learning.

Accreditation of schools should be based on multiple criteria, including, but not limited to, assessments of the Standards of Learning.

Standardized test scores should not be the sole criterion, or even the primary criterion, for determining promotion and retention, graduation, and school accreditation. Standardized tests, when related to the curricula being taught and used in conjunction with other criteria, can be one measure of student and school achievement but should not be the sole or primary criterion. A considerable body of research has shown that numerous factors must be included for true evaluation and that a single test cannot accomplish this task.

#### Potential Benefits Associated With the Standards Movement

Though proportionally fewer research studies document the effectiveness of high-stakes testing and measurement-driven instruction in improving student achievement, those in the standards movement can claim a body of research to support their position.

Waters, Burger, and Burger (1995) described the positive outcomes of standards-based school reform in a school division in Colorado. The system required students to demonstrate mastery of various concepts before being allowed to progress to the next level, “for example, middle school” (p. 35). The authors reported that the initiative resulted in a decreased gap between Hispanic and Anglo achievement scores along with increased performance for both groups. Davison, Scheisman, Koeppen, Wu, and Kwak (2001) documented the potential of using high-stakes tests to attract students needing basic skills remediation to summer school. The researchers described students’ academic gains as a result of Minnesota’s high school performance standards tests and state-funded summer school.

In a 1995 dissertation, New surveyed 3,500 Arkansas educators concerning the outcomes of Arkansas’ high-stakes testing requirements. While some of her findings indicated negative outcomes, she also determined that teachers supported some aspects of the testing requirements. Approximately 60% of respondents indicated that the tests provided teachers with valuable diagnostic information about students; two-thirds of responding teachers reported that their districts used test results to modify the curriculum and instructional priorities; and, over 80% of teachers surveyed said that eighth graders who had failed the tests were evaluated by an assessment team as required by the enabling state legislation, thus enhancing the probability of focused academic intervention for these students.

Noll (1999) conducted a qualitative study in Minnesota to determine how various stakeholders viewed the effectiveness of Minnesota’s high-stakes graduation standards. A large majority, 76% of those interviewed, felt that the standards would prepare students for the world of work and 80% of the interviewees thought the state’s emphases on basic skills and high standards were strengths.

Schleisman (2000) completed a study that assessed school districts’ responses to students who initially failed in their first attempt to pass the state’s high-stakes graduation tests. The author reported that school districts were, in fact, identifying students who may otherwise have “slipped through the cracks,” and that these students were being offered a range of compensatory programs and opportunities. She determined that greater curricular consistency was evident throughout the school divisions in the state.

Ravitch (1995) argued that the missing ingredient in improving schools was “agreement on the value of a challenging and rigorous education for everyone” (p. 35). Ravitch made the case for a revised curriculum that is not textbook-based but comes rather from what our citizens think children should learn. Ravitch further supported a national examination system in which the progress of American students could be meaningfully compared. “If such a test were widely recognized for its quality, students could submit their results with their college and employment applications” (1995, p. 38). Alexander (1993) stated the same theme by saying, “High standards are necessary and important as replacements for the current defacto, low-level standards implicit in most textbooks and tests.” (p. 9).

Resnick (1992-93) lamented that schooling today is based on a model that was in effect at the beginning of the 20<sup>th</sup> century when the goal was to educate a small elite group for future

leadership positions. Standards-based reform, she argued, engages teachers and parents in the public process of developing high performance standards. Most assessments to evaluate the success of reforms would be embedded in the regular curriculum rather than in annual barrier tests. In this manner, stakeholders would take part in both the development of standards and their related assessments in a very public forum.

Shanker (1994) stated that the development of high performance standards will really just expect the same from American students as is being expected from students from throughout the world. Performance exams are common in countries that normally compete with the United States economically. These expectations will allow and require students to engage in learning at a much higher level. Without stakes, Shanker argued, “nobody has to take education seriously” (p. 16). Shanker believed that the development of a national curriculum would save much time in curriculum development in the 50 states, and it would ensure that students throughout the country would have the opportunity to learn the same, high quality curriculum.

### Purpose of the Study

As a veteran Virginia educator, I had serious concerns about policies enacted by the Virginia State Board of Education that have resulted in state-mandated curricula and high-stakes testing. Numerous researchers (Allington & McGill-Franzen, 1992a, 1992b; Cameron, 1997; Corbett & Wilson, 1991; Fries, 1998; Koretz, Barron, Mitchell & Stecher, 1996; Massell, Kirst & Hoppe, 1997; Natriello & Pallas, 1999; Newman, King & Rigdon, 1997; Smith & Rottenberg, 1991; Wilkins, 1999) have indicated that policies resulting in high-stakes testing and measurement-driven instruction produce harmful consequences for students, schools, and educators. Other researchers (Davison, Schleisman, Koeppen, Wu & Kwak, 2001; New, 1995; Noll, 1999; Schleisman, 2000; Shanker, 1994; Waters, Burger & Burger, 1995; Young, 1996) have identified potentially positive outcomes of high-stakes testing and measurement-driven instruction. Through this research, I proposed to determine the opinions of Virginia teachers about the outcomes of measurement-driven instruction and high-stakes testing. This information will be shared with policy-makers to help influence the development and evaluation of regulatory practices for Virginia students, educators, and schools.

### Research Question

The following research question will guide the study: In the view of Virginia teachers, how do Virginia’s Standards of Accreditation (SOA) and Standards of Learning (SOL) tests and curriculum affect (1) outcomes for students, (2) outcomes for instructional practices, (3) outcomes for schools, (4) outcomes for public confidence in teachers and schools, and (5) outcomes for teachers?

### Theoretical Model

The theoretical model that guided this study was quite simple (see figure 1). First, evidence indicates that there is a strong national movement underway in this country to require states to conceptualize and develop content and performance standards for students and schools.

This movement has caused Virginia's political and educational leaders to conceptualize and implement a set of rigorous academic standards and requirements that are unique to this state. In effect, these standards and requirements became the Virginia Standards Movement. The national and state standards movements have benefits and liabilities for students and teachers, and these benefits and liabilities are likely to affect teacher beliefs about Virginia's academic standards and related assessments. My goal was to help explain teacher beliefs about these topics.

My search of the literature and my own experience led me to conclude that likely outcomes can be found in five domains: (1) outcomes for students, (2) outcomes for instructional programs, (3) outcomes for schools, (4) outcomes related to public confidence, and (5) outcomes for teachers themselves. For purposes of this study, these five domains functioned as *dependent variables*, and information was gathered to describe teacher opinions in each domain.

I hypothesized that five independent variables would affect the way teachers viewed the standards movement in Virginia. These included the following: (1) years of teaching experience, (2) the socio-economic level of teachers' schools, (3) SOL test grade status (whether the teacher taught in an SOL test grade), (4) teaching assignment (assignment to a particular grade range or specialty area), and (5) tenure status (whether or not teachers had reached continuing contract status). My experience led me to believe that these variables would affect the way teachers viewed the outcomes of Virginia's standards movement.

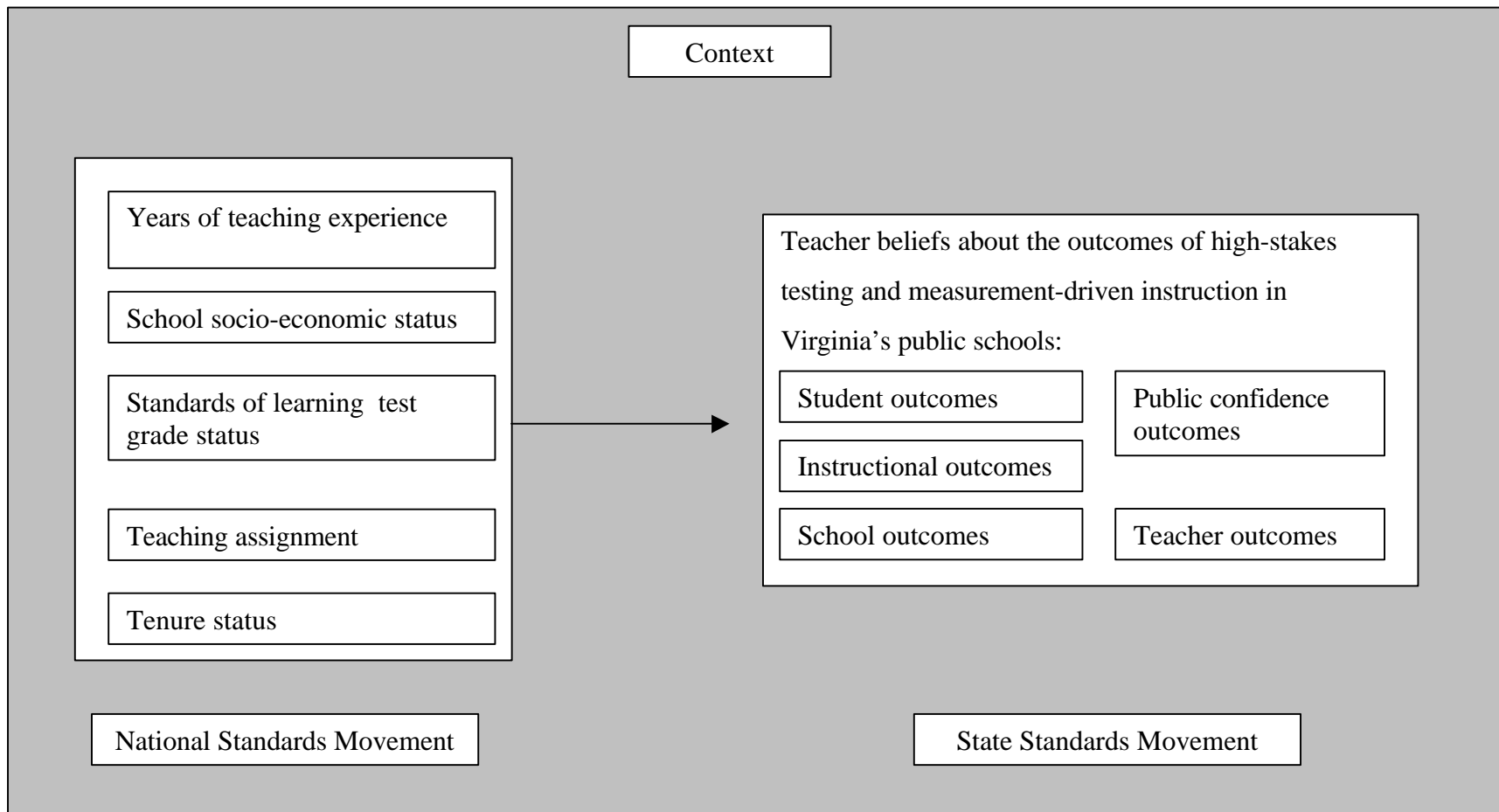
#### Definitions

**High-stakes tests:** Tests that have the effect of threatening punishment or consequences to teachers, students, schools, or school districts as a means of influencing curricular and instructional practices.

**Measurement-driven instruction:** The use of sanctions or contingencies associated with students' performance to influence the instructional program.

**Standards movement:** A philosophy first evident in the 1980s that called for objective assessment of the outputs or products of schools and school divisions rather than on inputs such as dollars spent, class size, or the square footage of facilities.

**Standards of Accreditation (SOA):** Known officially as *Regulations Establishing Standards for Accrediting Public Schools in Virginia*, these are Virginia Board of Education policies that specify the rules for public school accreditation.



**Figure 1. Theoretical model:** Variables associated with teacher beliefs about the outcomes of measurement-driven instruction and high-stakes testing in the context of national and state standards movements.

In a stark departure from previous standards, the 2000 accreditation standards establish state-mandated curricula in English, social studies, math, and science, and they establish benchmark achievement scores that schools must meet to retain their accreditation. The SOA also establish benchmark scores that students must reach on *verified units of credit* tests at the high school level to receive a regular high school diploma.

**Standards of Learning (SOL):** The SOL are a state-mandated K-12 curriculum in English, social studies, math, and science, and technology.

### Personal Bias

Like all researchers, I brought my personal and professional background, experiences, values, and perceptions to this study (Miles & Huberman, 1994). These factors influenced the topic I chose to study and how I conceptualized and carried out the research. My experiences, beliefs, values, and perceptions had the potential to introduce bias into the study. I became increasingly aware of this danger as I proceeded with the study and strove to ensure that the items on the questionnaire and interview protocol were balanced pro and con and that my portrayal of the interviewees' views represented their thoughts and beliefs and not my own.

Several purposeful strategies helped to mitigate the effects of researcher bias. First, the statewide survey was sent to a systematic sample of teachers who were members of the Virginia Education Association at the beginning of the 2000-2001 school year. The use of both quantitative (survey) data and qualitative (interview) data helped to provide a form of triangulation to validate or to reject potential conclusions. The survey and interview instruments were subjected to content validation studies to help ensure that the instruments were, in fact, measuring what they purported to measure. Additionally, participants in telephone interviews were selected randomly from a matrix that categorized teachers according to the independent (demographic) variables in the study.

Having instituted these safeguards, however, there is no question that I was seriously worried about the outcomes of high-stakes testing and measurement-driven instruction when I began this study. My early review of literature only served to intensify these concerns. These factors may have contributed to bias in the development of language in the survey and in the interview questions. However, in the latter stages of my research, I began to find recent scholarly studies that documented some of the positive outcomes of the standards movement nationally. I believe that these reports have been fairly portrayed in the review of literature and that they contributed appropriately to the conclusions I reached.

### Limitations

Several limitations should be considered when reading and evaluating this study. First, the study was limited only to responses from a sample of Virginia teachers who belonged to the Virginia Education Association. The opinions of this group may vary from the general teacher population in the state and from the members of other professional organizations to which Virginia teachers may belong. Though the survey was a random sample of the VEA membership,



those responding may have been teachers who held particularly unfavorable opinions about SOL testing. Approximately one fourth of the sample did not respond to the survey, so I do not know how their answers to the survey might have changed the results.

Concerning the qualitative (interview) part of the study, teachers who completed the mailed survey volunteered to participate in telephone interviews. It is possible that those who volunteered to be interviewed were more negatively disposed to SOL testing than the remainder of the sample, thus the qualitative data may be over-represented with views that portray unfavorable outcomes from SOL testing.

Finally, it is important to consider that Virginia teachers responded to my questions and statements based on only three years of SOL test results. That is, they had the opportunity to read and hear about low scores, but they did not have the benefit of seeing long-term trends in student achievement. Only about two percent of Virginia schools were fully accredited based on the results of the first year of SOL testing in 1998; only about six percent of the schools were fully accredited after the second year of SOL testing in 1999; approximately 22% of Virginia schools met requirements for full accreditation after SOL testing in 2000. Though the number of accredited schools has increased, the vast majority of Virginia schools have not achieved scores that will allow them to become accredited, and teachers who participated in my study had no way of knowing (nor do they yet) what outcomes will result after five or ten years of standards-based reform. The negative opinion that is predominant at the beginning of the initiative may diminish if testing in subsequent years produces measurable, meaningful results and if outcomes for students and teachers are not seen as harmful.

### Summary of Chapter One and Overview of the Dissertation

In chapter one I described the regulatory background and potential problems associated with high-stakes testing and measurement-driven instruction in the context of national and state environments. Potentially important consequences were described for students, teachers, schools, and for instructional programs.

Teacher opinions about likely outcomes of high-stakes testing and measurement-driven instruction were gathered in a questionnaire organized around five domains that may be affected by the state's mandated curriculum and assessment program: (1) outcomes for students, (2) outcomes for instructional practices, (3) outcomes for schools, (4) outcomes for public confidence in teachers and schools, and (5) outcomes for teachers. The survey was administered to a systematic sample of 2000-2001 members of the Virginia Education Association. A subset of questionnaire respondents was then interviewed by telephone to gather additional teacher insights about the effects of Virginia's mandated curricula and assessment program. Personal bias and limitations of the study were discussed.

A review of related literature is presented in chapter two. Population and sample, instrument development, data collection procedures, reliability and validity, and analytical procedures are described in chapter three. Results of the study are described in chapter four, and a

summary of findings, conclusions, policy considerations, and recommendations for further research are presented in chapter five.